

# **Evaluation de la justesse vocale en contexte mélodique**

**Pauline Larrouy-Maestri**

**Thèse présentée en vue de l'obtention  
du grade de Docteur en Sciences  
Psychologiques et de l'Education**

Sous la direction de Dominique Morsomme

Jury de thèse composé d'Irène Deliège,  
Nathalie Henrich, Régine Kolinsky, Steve  
Majerus et Dominique Morsomme



**Faculté de Psychologie et  
des Sciences de l'Education  
Logopédie de la voix**

**Université de Liège**



# REMERCIEMENTS

---

Depuis 1999 et mes premières études, tous les travaux ont commencé par “Merci à tous ceux, qui, de près ou de loin, ont contribué à la réalisation de ce projet”. Pour mener à bien une thèse, la ténacité ne suffit pas et je souhaite remercier les personnes sans qui ce travail n’aurait tout simplement pas pu avoir lieu.

Merci à Dominique Morsomme de m’avoir invitée à rejoindre son équipe, d’avoir cru en ce projet et de l’avoir rendu possible.

Merci à Régine Kolinsky et Steve Majerus d’avoir accepté de me guider.

Merci à Irène Deliège et Nathalie Henrich pour l’attention portée.

Merci à l’Université de Liège de m’avoir si bien accueillie et d’avoir financé ce projet.

Merci également au FNRS et à la Communauté Française de Belgique d’avoir subsidié les séjours de recherche et congrès.

Merci à Guillaume Videlier, au centre Henri Pousseur, à Jean-Marc Sullon, à Jacques Sougné et à Frédéric Pluquet de m’avoir initiée à l’art informatique.

Merci à Yohana Lévêque, Sean Hutchins, David Magis, Daniele Schön et Isabelle Peretz, pour nos instructives collaborations.

Merci aux chercheurs, reviewers et éditeurs qui m’ont aidée à avancer.

Merci aux chanteurs occasionnels et professionnels d’avoir chanté.

Merci aux musiciens et non musiciens d’avoir écouté.

Merci aux étudiantes qui ont sérieusement joué le jeu à les tester.

Merci à Serge Brédart, Isabelle Halleux et Jean-Marie Molle pour leurs conseils avisés.

Merci à mes amis, à mes familles, à mes collègues, de m’avoir soutenue et supportée.

Merci à Mireille, Chanterelle et Guillaume pour le coup de main de dernière minute.

Merci à Chopin, Liszt, Schubert, Szymanowski et Bach d’avoir patienté.

Last but not least, thanks to J. to be there and to believe in me.



# TABLE DES MATIERES

---

<b>PREFACE</b>	<b>11</b>
<b>INTRODUCTION THEORIQUE</b>	<b>15</b>
<b>Chapitre I</b>	
<b>Perception de la justesse</b>	<b>17</b>
Définition de la justesse mélodique	18
Musique occidentale	19
Relation entre les notes : contours mélodiques et tonalités	20
Relation entre les notes : intervalles	21
Erreurs mélodiques	23
Perception de la dimension mélodique	25
Apprentissage des règles musicales	26
Evolution de la perception mélodique	27
Perception de la justesse chez l'auditeur adulte	29
Effet de l'expertise sur la perception mélodique	30
Effet de l'entraînement musical	30
Effet des bagages linguistique et culturel	32
Synthèse	34
<b>Chapitre II</b>	
<b>Evaluation de la justesse vocale</b>	<b>37</b>
Evaluation subjective de la justesse vocale	40
Etat des lieux	40
Facteurs liés aux juges	41
Influence du matériel musical sur la perception mélodique	42
Evaluation objective de la justesse vocale	45
Outils d'analyse	45
Tâches expérimentales	47
Limites de l'évaluation objective	49
Synthèse	51

**Chapitre III****La justesse des chanteurs occasionnels 53**

Développement du chant et de la justesse	53
Les troubles de justesse	56
Trouble perceptif	56
Trouble relatif à la mémoire des hauteurs	58
Trouble d'intégration sensorimotrice	58
Trouble au niveau moteur	59
Variabilité de la justesse vocale	60
Variabilité liée au matériel musical	60
Variabilité liée à la méthode d'évaluation	63
Variabilité liée au sujet	64
Variabilité liée au contexte de l'évaluation	64
Synthèse	65

**Chapitre IV****La justesse des chanteurs lyriques 67**

L'instrument vocal ou la physiologie de la voix chantée	67
Du chanteur occasionnel au chanteur entraîné	71
Composantes acoustiques des voix lyriques	72
Vibrato	73
Timbre ou balance spectrale	74
Autres paramètres acoustiques	76
La justesse des chanteurs lyriques	77
Éléments acoustiques influençant la perception de la justesse vocale	78
Synthèse	81

**PARTIE EXPERIMENTALE 85****OBJECTIFS DES ETUDES 87****Etude 1****Criteria and tools for objectively analysing the vocal accuracy of a popular song 93**

Introduction	93
Methods	97
Participants	97
Material	98
Analytical procedure	98

Calculations	101
Results	103
Discussion	105
<b>Etude 2</b>	
<b>The evaluation of singing voice accuracy: A comparison between subjective and objective methods</b>	<b>109</b>
Introduction	110
Methods	112
Participants	112
Procedure	113
Description of the objective method	114
Results	115
Interval deviation and subjective judgments of vocal accuracy	116
Acoustic parameters predicting the judge rating	119
Discussion	120
<b>Etude 3</b>	
<b>Singing ability is rooted in vocal-motor control of pitch</b>	<b>123</b>
Introduction	124
Methods	127
Participants	127
Stimuli and procedure	128
Analyses	132
Results	134
Slider-voice comparisons	134
Vocal quality	135
Target pitch height	135
Melodic singing ability and its relationship to other tasks	137
Discussion	138
Scaling up to melodies	139
<b>Etude 4</b>	
<b>The effects of stress on singing voice accuracy</b>	<b>143</b>
Introduction	144
Methods	147
Participants	147
Procedure	147
Material	148
Evaluation of sung performances	150

Statistical analyses	151
Results	152
Evaluation of stress level	152
Evaluation of sung performances	153
Relationships between stress level and vocal accuracy	155
Discussion	156
Changes in stress level	156
Changes in singing voice accuracy	157
Implications	159
<b>Etude 5</b>	
<b>The effect of melody and technique on the singing voice accuracy of trained singers</b>	<b>161</b>
Introduction	161
Methods	163
Participants	163
Material and procedure	163
Acoustic analysis	164
Statistical analysis	164
Results	165
Discussion and conclusion	166
<b>Etude 6</b>	
<b>Effects of melody and technique on acoustical and musical features of Western operatic singing voices</b>	<b>169</b>
Introduction	170
Methods	173
Participants	173
Material and procedure	174
Acoustical analysis	174
Statistical analysis	176
Results	177
Effect of melody and technique on the singing parameters	178
Profile of the operatic singing technique	181
Discussion	182
Effects of melody and technique	182
Profile of the operatic singing technique	185



<b>Etude 7</b>	
<b>The evaluation of vocal accuracy: The case of operatic singing voices</b>	<b>187</b>
Introduction	188
Methods	191
Participants	191
Material	191
Acoustical analysis	192
Perceptual task	194
Statistical analyses	194
Results	196
Intra-judge reliability	196
Inter-judge reliability	199
Predictive model of vocal accuracy evaluation	200
Discussion	203

---

## **DISCUSSION GENERALE** **209**

<b>DISCUSSION</b>	<b>211</b>
Synthèse des études	212
Evaluation objective de la justesse mélodique	219
Méthode d'analyse de la justesse vocale en contexte mélodique	220
Avantages et limites des outils d'analyse	221
Avantages et limites des critères d'évaluation	225
Piste à explorer : seuils de justesse	227
Pertinence d'une méthode objective	227
Pertinence de notre méthode d'évaluation	228
Applications en pédagogie musicale	229
Applications en clinique	231
Piste à explorer : facteurs influençant l'évaluation	232
Vers une définition de la justesse vocale	233
Capacité à évaluer la justesse mélodique	234
Notion de justesse mélodique	236
Piste à explorer : expertise musicale	239
<b>CONCLUSIONS</b>	<b>243</b>

---

## **REFERENCES BIBLIOGRAPHIQUES** **247**



# PRÉFACE

---

*“La musique est une science qui doit avoir des règles certaines ;  
ces règles doivent être tirées d'un principe évident,  
et ce principe ne peut guère nous être connu  
sans le secours des mathématiques.”*

*Jean-Philippe Rameau, 1683-1764*  
Traité de l'harmonie réduite à ses principes naturels

Si vous tendez l'oreille le jour d'un anniversaire, vous noterez très certainement que les participants à la fête ne chantent pas toujours juste. Quels sont les critères qui vous conduisent à un tel jugement ? Auriez-vous les mêmes critères s'il s'agissait de la production d'un candidat au concours Reine Elisabeth ou d'un ami qui n'a jamais pris de cours de chant ? Si nous portons facilement un jugement sur la qualité d'une performance, il n'est pas évident d'explicitier les critères sur lesquels notre évaluation se base.

Ce travail de recherche explore l'évaluation de la justesse vocale d'un point de vue méthodologique et théorique. D'une part, l'objectif de ce travail concerne la mise au point et l'application d'outils d'évaluation objectifs pour évaluer la justesse vocale en contexte mélodique. D'autre part, pour que ces outils soient pertinents, il est nécessaire d'observer la perception de la justesse et les éléments saillants pour un auditeur.

L'introduction théorique s'articule autour de quatre chapitres destinés à éclairer le lecteur sur notre démarche. Dans un premier temps, nous définissons la justesse d'un point de vue musical et perceptif. Dans un deuxième temps, nous proposons un état des lieux concernant les différentes méthodes d'évaluation disponibles pour estimer la justesse vocale ainsi que leurs avantages et leurs limites. Par ailleurs, le chant est une activité musicale que la plupart d'entre nous

pratique quotidiennement, depuis l'enfance, à diverses occasions, mais nous ne sommes pas tous des chanteurs professionnels pour autant. Dès lors, l'évaluation de la justesse doit s'adapter au public concerné. Le troisième chapitre est dédié au développement de la justesse vocale, à la justesse chez l'adulte non entraîné et aux différents troubles qui peuvent altérer la justesse d'une performance vocale. Dans un dernier chapitre, nous décrivons les chanteurs pour lesquels l'instrument vocal est travaillé à la manière d'un instrument de musique et plus spécifiquement les chanteurs lyriques.

Dans la partie expérimentale, nous développons les sept études menées au cours de cette thèse. Alors que la première étude expose les outils et critères d'évaluation de la justesse élaborés dans le cadre de nos recherches, la seconde a pour but d'observer leur pertinence auprès d'un jury d'experts en musique. Une méthode d'évaluation n'a de sens que par son application. Les troisième et quatrième études ont donc pour objectif d'utiliser cette méthode dans des contextes expérimentaux. D'une part, nous avons comparé nos outils avec les méthodes existantes. D'autre part, cette méthode nous a permis d'examiner l'effet du contexte d'évaluation sur la justesse d'une performance. Par ailleurs, la justesse des voix lyriques est peu explorée à l'heure actuelle et nous avons souhaité, dans une cinquième étude, observer les effets de la mélodie et de la technique vocale sur la justesse d'une performance en contexte mélodique. Pour finir, afin de mieux comprendre les processus d'évaluation guidant le jugement des voix lyriques, nous avons réalisé deux études explorant les versants de production et perception de ces voix. La sixième étude tente donc d'observer le profil acoustique des voix lyriques et la septième compare l'analyse acoustique de ces voix avec l'évaluation d'experts.

Au travers de ces sept études, nous tenterons donc de développer une méthode permettant d'évaluer objectivement la justesse d'une performance vocale en contexte mélodique, d'observer sa pertinence et de clarifier la notion de justesse. Une synthèse des principaux résultats est présentée dans la discussion générale et des perspectives d'application en pédagogie, en clinique et en recherche fondamentale sont proposées à la lumière des conclusions des recherches menées au cours de cette thèse.





# **Introduction théorique**





# CHAPITRE I

## Perception de la justesse

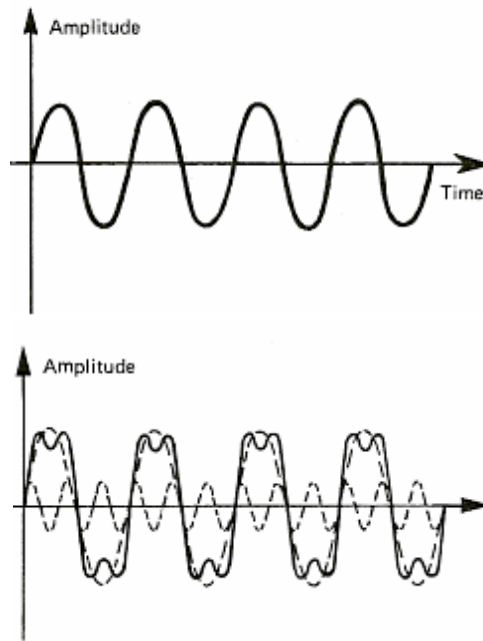
---

*Juste... par rapport à quoi ?*

L'évaluation de la justesse vocale en contexte mélodique ne peut être étudiée sans faire référence aux règles musicales de notre culture. En effet, la notion même de justesse est associée à un modèle à respecter, différent selon les cultures et les époques, et se base donc sur un système musical particulier. Par ailleurs, la musique ne se résume pas à une notation musicale et si nous pouvons observer plusieurs types d'erreur selon les règles musicales de notre système, il est également nécessaire d'examiner la perception mélodique de l'auditeur. Ce chapitre est dédié à la description du système musical occidental dans lequel se situent nos recherches et les erreurs de justesse qui peuvent être observées. Dans un deuxième temps, nous décrivons l'apprentissage implicite de ces règles et plus généralement la perception mélodique et de la justesse chez des auditeurs non experts et experts.

Le chant, tout comme la musique instrumentale, peut être défini par quatre paramètres acoustiques : la hauteur, la durée, l'intensité et le timbre. Dans le cadre de cette thèse, nous nous concentrons sur le paramètre de hauteur et sur l'organisation des sons en contexte mélodique. En effet, une mélodie étant une succession de notes obéissant aux conventions et contraintes culturelles (Ringer, 2002), l'organisation des hauteurs est fondamentale. La hauteur se caractérise, au niveau physique, par la fréquence de vibration sonore exprimée en Hertz. Tout comme la majorité des sons de notre environnement, les sons instrumentaux ou vocaux sont des sons complexes, c'est-à-dire composés d'une fréquence fondamentale ( $F_0$ ) et de ses harmoniques (multiples de cette fréquence fondamentale). Comme l'illustre la Figure

1, le signal acoustique correspond alors à la combinaison de vibrations simples (fréquence fondamentale -  $F_0$ ) et de ses harmoniques.



**Figure 1.** Représentation schématique d'un son pur (vibration simple, en haut) et d'un son complexe (vibration simple et ses harmoniques, en bas).

## Définition de la justesse mélodique

L'activité musicale est présente depuis des millénaires, son importance chez l'humain n'est plus à discuter (Jancke, 2008, 2009 ; Koelsch, 2011 ; Nettl, 2000) et les recherches sur sa perception et sa production sont incontournables dans la littérature scientifique relative à la cognition et à la pédagogie musicale. Afin de situer ce travail de thèse, nous décrivons dans cette partie le contexte musical de nos recherches, c'est-à-dire la musique occidentale, ainsi que les caractéristiques mélodiques (contour, tonalité et intervalles) qui la définissent. Sur la base du système musical occidental, différents types d'erreur de justesse en contexte mélodique sont observées. La littérature relative à l'histoire musicale, à l'harmonie et à l'acoustique étant particulièrement vaste et fournie, nous nous limitons ici à une synthèse personnelle, élaborée à la suite d'une formation et d'une pratique musicale classique. Cette brève synthèse ne se substitue pas

aux ouvrages spécifiques à ces domaines mais a pour objectif de familiariser le lecteur aux éléments musicaux détaillés dans la suite de cette thèse.

## ***Musique occidentale***

Comme il existe différents langages, il existe également différents systèmes musicaux, chacun avec ses gammes, ses règles et ses grammaires dictant l'organisation des sons (Cross, 2001 ; Lerdahl & Jackendoff, 1983). La musique occidentale se rapporte à la musique dite populaire, traditionnelle ou savante influencée par la culture européenne et américaine. Elle apparaît à la suite de la musique grecque ancienne et de la Rome antique et trouve son origine dans le chant chrétien puis dans le chant grégorien. Tout d'abord transmise de manière orale, la musique est ensuite transmise à l'aide de systèmes d'écriture. Cela permet le développement et la complexification des formes musicales, des chants monodiques à la polyphonie. La musique occidentale savante est exclusivement vocale pendant toute la durée du Moyen Âge. La voix, unique instrument de la musique profane, s'intègre ensuite dans les œuvres des époques musicales dites classique puis romantique. A cette période, entre le début du XIX<sup>ème</sup> siècle et le milieu du XX<sup>ème</sup> siècle, l'aspect expressif de la partie mélodique s'accroît (voir par exemple Samson, 2001) et les mélodies foisonnent. Parallèlement à la musique savante, d'autres genres musicaux ont évolué, fortement liés ou associés à un folklore, à une culture nationale ou religieuse, voire à une zone géographique. A partir du XX<sup>ème</sup> siècle, des genres musicaux, qualifiés de populaires, prennent une importance grandissante comme le jazz, le rock, le blues, la musique funk, le métal et bien d'autres.

Malgré la diversité des genres musicaux, la consonance "occidentale" persiste. En effet, dès le XVI<sup>ème</sup> siècle et grâce notamment à l'arrivée de l'imprimerie musicale, les règles musicales et harmoniques se répandent et se perpétuent moyennant quelques modifications. La musique occidentale, au sens large, est hiérarchisée et structurée par un système tonal. Apparue à la Renaissance, ce système est encore majoritaire dans notre environnement musical actuel.

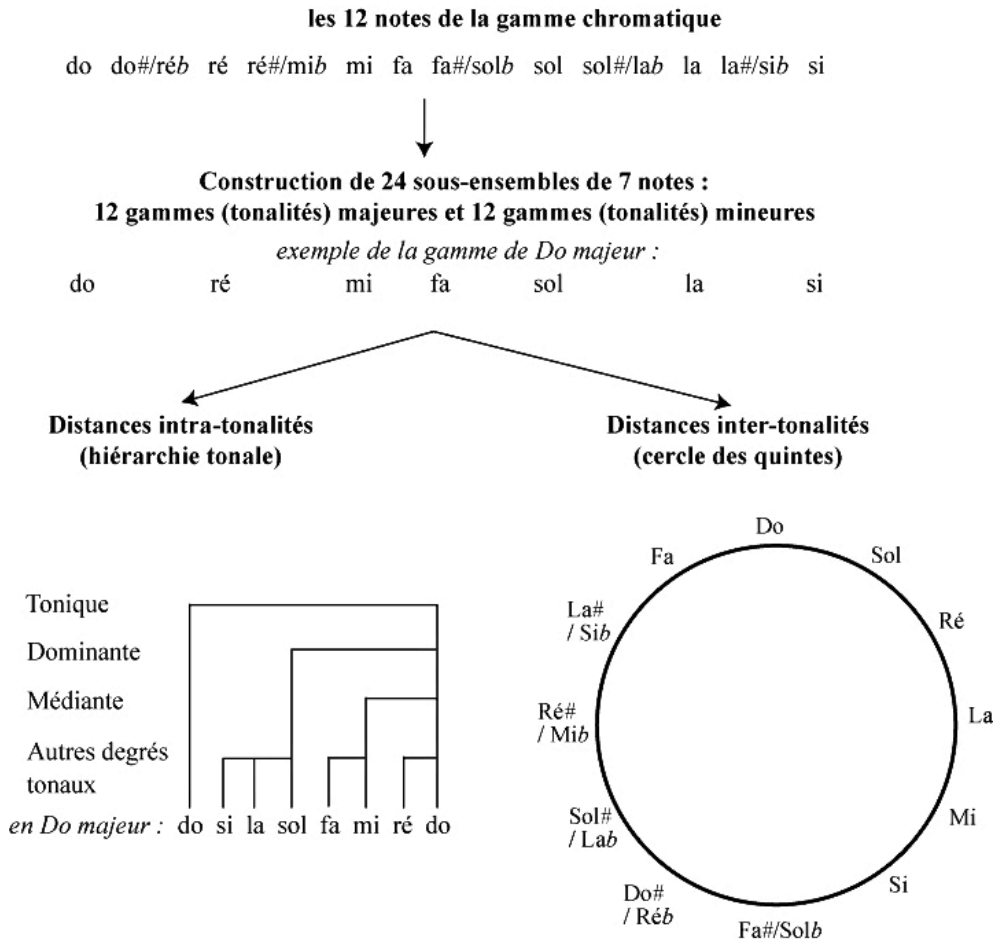
## ***Relation entre les notes : contours mélodiques et tonalités***

La musique occidentale, au sens large, se base sur un ensemble fini d'unités que sont les 12 notes de la gamme chromatique et qui constituent l'octave (voir Figure 2). La succession de notes dans le temps va former des mélodies de contours mélodiques variés. L'organisation des notes au sein de la mélodie va définir la tonalité. En effet, la musique tonale s'organise par sous-ensembles de 7 notes qui constituent une gamme dite diatonique associée à une tonalité particulière. Les intervalles entre les notes de la gamme diatonique, c'est-à-dire la place des tons et des demi-tons, caractérisent également le mode de la tonalité : majeur ou mineur. Autrement dit, à partir des 12 notes de la gamme chromatique, 12 tonalités majeures et 12 tonalités mineures peuvent être construites, comme schématisé sur la Figure 2.

Au sein d'une gamme, les notes sont hiérarchisées et seront donc plus ou moins importantes selon leur degré. Comme illustré au bas (gauche) de la Figure 2, dans une gamme de Do Majeur, le 1<sup>er</sup> degré représente la tonique, le 3<sup>ème</sup> correspond à la médiate et le 5<sup>ème</sup> est la dominante. Ces trois degrés sont les plus stables et constituent, par ailleurs, l'accord parfait de cette tonalité. Les 2<sup>ème</sup> (sus-tonique), 4<sup>ème</sup> (sous-dominante), 6<sup>ème</sup> (sus-dominante) et 7<sup>ème</sup> (sensible) degrés sont moins stables et seront attirés par les notes constituant l'accord. Ces degrés constituent donc principalement des notes de passage entre les degrés stables.

Au-delà de la relation hiérarchique entre les notes, les gammes sont reliées entre elles. En effet, les notes et accords constitutifs d'une tonalité appartiennent en réalité à plusieurs tonalités et ont des fonctions différentes selon la tonalité dans laquelle ils se trouvent. Comme indiqué au bas (droite) de la Figure 2, il est possible de moduler, c'est-à-dire de passer d'une tonalité à l'autre. Cependant, une modulation peut avoir lieu entre tonalités voisines (le cas le plus courant : en suivant le cercle dans un sens ou dans l'autre sur la Figure 2), entre tonalités relatives (passage du mode majeur au mode mineur ou inversement) ou entre tonalités plus éloignées. Dans le cadre de nos recherches, nous observons également des modulations intra-tonales. En effet, l'instrument vocal n'est pas un instrument à clavier où les notes

sont séparées par des demi-tons mais, tout comme les instruments à vent ou les instruments à cordes sans frettes, la voix permet de produire des sons de manière continue. De ce fait, la voix permet une déviation du centre tonal sans pour autant moduler à une tonalité voisine, relative ou éloignée.



**Figure 2.** Représentation schématique du système tonal (d'après Bigand, 1994).

### ***Relation entre les notes : intervalles***

Défini comme la distance entre deux notes de hauteurs différentes, l'intervalle minimal en musique occidentale est le demi-ton. Dans un contexte mélodique, les notes sont jouées ou chantées de manière successive et constituent alors des intervalles mélodiques. Les différents intervalles existant au sein d'une gamme se nomment la seconde, la

tierce, la quarte, la quinte, la sixte, la septième et l'octave. Selon le nombre de tons et demi-tons contenus, ces intervalles peuvent être majeurs, mineurs, justes, augmentés ou diminués.

Si la notion d'intervalle existe depuis des siècles, leur dimension a beaucoup évolué. En effet, différents tempéraments ont été proposés dans le but de répartir les intervalles de la gamme sur les instruments à sons fixes (les claviers par exemple). L'objectif était de trouver un compromis acceptable entre l'inaccessible exactitude acoustique des harmoniques naturelles et le système harmonique. Du tempérament où les quartes et quintes sont justes mais où les tierces sont fausses (tempérament pythagoricien, période médiévale), aux tempéraments où tierces et quintes sont privilégiées (tempéraments zarliniens, périodes de la Renaissance et du Baroque), le choix s'est finalement porté sur un tempérament égal (du XVIII<sup>ème</sup> siècle à nos jours). Ce tempérament égal est une solution purement mathématique et une aberration harmonique, où tous les intervalles, hormis l'octave, sont faux par rapport à la réalité acoustique des harmoniques naturelles.

Note	Tempérament égal	Tempérament pythagoricien		Tempérament zarlinien	
	cents	ratio	cents	ratio	cents
Do	0	1 / 1	0,00	1 / 1	0,00
Do# / Réb	100	256 / 243	90,22	16 / 15	111,73
Ré	200	9 / 8	203,91	9 / 8	203,91
Ré# / Mib	300	32 / 27	294,13	6 / 5	315,64
Mi	400	81 / 64	407,82	5 / 4	386,31
Fa	500	4 / 3	498,04	4 / 3	498,04
Fa# / Solb	600	1024 / 729	588,27	45 / 32	590,22
Sol	700	3 / 2	701,96	3 / 2	701,96
Sol# / Lab	800	128 / 81	792,18	8 / 5	813,69
La	900	27 / 16	905,87	5 / 3	884,36
La# / Sib	1000	16 / 9	996,09	9 / 5	1017,60
Si	1100	243 / 128	1109,78	15 / 8	1088,27
Do	1200	2 / 1	1200,00	2 / 1	1200,00

**Figure 3.** Tableau comparatif des tempéraments égal, pythagoricien et zarlinien.

Comme l'illustre la Figure 3, le tempérament égal consiste à diviser l'octave en douze demi-tons strictement égaux. Bien que les intervalles soient faux relativement aux harmoniques naturelles, ce tempérament est facilement supporté par l'oreille. Cela provient probablement du fait qu'ils sont tous faux de la même façon et que ce système a été adopté pour l'accord instrumental depuis quelques siècles déjà (depuis 1750 environ). Notons que l'utilisation du tempérament égal simplifie les transpositions et facilite également la mesure des intervalles. En effet, la gamme chromatique tempérée étant composée de 12 demi-tons et chaque demi-ton contenant 100 cents, les intervalles peuvent donc être calculés en multipliant le nombre de demi-tons (par exemple, 200 cents pour un ton, 400 cents pour une tierce majeure). Comme indiqué au sujet du centre tonal, la voix permet la production de sons de manière continue et donc de chanter des intervalles inférieurs au demi-ton. Dans le cadre de nos recherches, nous quantifions ces micro-variations en cents.

### ***Erreurs mélodiques***

L'évaluation de performances vocales se base sur des aspects esthétiques relatifs à une culture particulière. Dans le cadre de notre culture occidentale et au vu des éléments musicaux décrits précédemment, la justesse mélodique correspondrait au respect de la relation entre les notes, appelé contour mélodique, au respect de la taille des intervalles et au respect de la tonalité.

**Version correcte****Erreur de contour à la note 2 (cette note descend au lieu de monter)****Erreur d'intervalle de 200 cents (un ton) entre les notes 1-2 et 2-3 (le contour de la mélodie est respecté mais l'intervalle est trop grand)****Erreur de centre tonal à la note 5 (erreur d'intervalle entre les notes 4-5 non compensée par la suite)****Figure 4.** Erreurs de justesse observables en contexte mélodique.

Comme illustré dans la Figure 4, une erreur de contour mélodique correspond à une direction de l'intervalle contraire à la direction de cet intervalle dans la mélodie de référence. Ici, la seconde note de l'intervalle est plus basse que la première alors qu'elle devrait être plus haute, changeant ainsi le contour mélodique attendu. Une erreur d'intervalle est définie comme une différence entre deux notes qui ne correspond pas à celle attendue, alors que la direction de l'intervalle est respectée. Notons que la figure représente une erreur d'un ton (ré devient mi) mais que les déviations peuvent être plus petites, de l'ordre de quelques cents. Le dernier type d'erreur repose sur le non respect du centre tonal. Dans ce cas, une erreur d'intervalle n'est pas compensée par une déviation similaire et mène à une modulation. Ici encore, l'erreur peut être suffisamment grande pour mener à une tonalité voisine ou relative à la tonalité de départ mais peut également se limiter à un déplacement plus subtil.



Plusieurs de ces erreurs ont été observées dans les études sur l'évaluation vocale (*cf.* Chapitre II) mais à notre connaissance, aucune étude n'a permis de confronter ces erreurs afin d'en observer l'importance relative dans le jugement d'une performance. De manière confondue, les erreurs mélodiques que nous déduisons de la théorie musicale font référence à l'intonation (dans le sens variation de la fréquence fondamentale). Avant d'aborder plus précisément la perception mélodique, notons que Watts, Barnes-Burroughs, Andrianopoulos et Carr (2003) soulignent que l'intonation est primordiale pour juger la qualité d'une performance vocale chez des chanteurs occasionnels. En effet, ils ont mené une étude auprès de 1009 professeurs de chant du "National Association of Teachers of Singing" (NATS). Les participants devaient citer les critères qu'ils prenaient en compte pour évaluer la qualité d'une performance vocale. Parmi les critères cités, l'intonation apparaissait avant le timbre et la musicalité. Cependant, les critères d'intonation, c'est-à-dire les critères de justesse, n'ont pas été détaillés. En effet, une intonation adéquate nécessite le respect des contours, de la taille des intervalles et de la tonalité mais le questionnaire proposé aux enseignants ne permettait pas de distinguer ces caractéristiques mélodiques.

## **Perception de la dimension mélodique**

L'évaluation de la justesse sous-entend le bon fonctionnement de l'appareil auditif et une audition suffisante pour percevoir les variations de la fréquence fondamentale. Cependant, percevoir ne se résume pas à recevoir des stimuli externes. En effet, le cerveau utilise en permanence des connaissances liées au contexte de présentation et à l'expérience du sujet. Il s'agit d'influences "top-down", terme qui désigne la façon dont nos connaissances influencent notre perception (pour une revue, voir Kveraga, Boshyan, & Bar, 2007). Au niveau musical, chaque culture est régie par un système où les gammes, les structures mélodiques et rythmiques sont spécifiques (Cross, 2001 ; Lerdahl & Jackendoff, 1983). Au sein d'une mélodie, les sons sont organisés dans une syntaxe qui varie selon les époques et les contextes culturels. La perception de la justesse serait donc dépendante de l'évolution des capacités de perception mais également d'un apprentissage implicite de règles

musicales propres à la culture environnante. Cette partie vise à décrire les processus d'apprentissage implicite qui permettent aux auditeurs de devenir des experts de leur culture. Dans un second temps, nous détaillons l'évolution de la perception mélodique puis la perception de la justesse chez l'auditeur adulte.

## ***Apprentissage des règles musicales***

Que ce soit au sujet de la musique ou du langage, de nombreuses études ont observé l'acquisition de connaissances relatives aux régularités structurelles par apprentissage implicite, c'est-à-dire par une exposition passive à leur environnement sonore (pour une revue, voir Patel, 2008). Un exemple d'apprentissage implicite est l'acquisition de la langue maternelle. Le jeune enfant acquiert le langage simplement par exposition et cette langue se développe alors qu'il ne peut verbaliser les règles qui en régissent sa structure. Saffran et ses collaborateurs ont étudié l'influence des régularités linguistiques dans l'acquisition du langage chez des enfants et des adultes (Saffran, Aslin, & Newport, 1996 ; Saffran, Newport, & Aslin, 1996). De manière synthétique, ces travaux montrent que des auditeurs deviennent sensibles aux régularités statistiques, c'est-à-dire à la probabilité qu'une unité linguistique soit suivie d'une autre. Ces régularités permettent d'extraire les mots d'une séquence verbale entendue. Saffran, Johnson, Aslin et Newport (1999) ont montré que l'apprentissage de régularités statistiques n'est pas limité au langage mais est également observable avec des séquences de sons musicaux. Notons que cet effet est particulièrement visible lorsque les modalités musicales et linguistiques sont superposées, c'est-à-dire quand le matériel linguistique est chanté (Schön, Boyer, Moreno, Besson, Peretz, & Kolinski, 2008).

A la suite de l'apprentissage des probabilités transitionnelles d'un nouveau système musical, des auditeurs reconnaissent les séquences musicales proposées (Jonaitis & Saffran, 2009). L'auditeur est donc capable d'intégrer les probabilités de combinaisons de séquences mélodiques et d'intégrer les règles de nouveaux systèmes musicaux par imprégnation (Loui, Wessel, & Kam, 2010). Notons que l'étendue des séquences présentées aura des conséquences sur le type d'apprentissage. En effet, l'exposition à un large ensemble permet

l'extraction des régularités statistiques formant les mélodies, tandis qu'une exposition à un ensemble plus restreint favoriserait la reconnaissance des mélodies en tant que telles (Loui & Wessel, 2008). L'existence d'une sensibilité aux probabilités transitionnelles inhérentes à la musique occidentale a également été mise en évidence sur le plan neurophysiologique. En effet, Kim, Kim et Chung (2011) observent une corrélation entre le niveau de probabilité des séquences présentées et l'activation physiologique.

Stalinski et Schellenberg (2012) estiment que le processus d'enculturation – processus par lequel le groupe transmet à l'enfant, dès sa naissance, des éléments culturels, des normes et des valeurs – de la culture musicale est terminé dès l'âge de 5 ans. Etant donné que les enfants sont exposés à la musique très tôt (Miyamoto, 2007), ils acquièrent des connaissances musicales de manière implicite sans pour autant être capables de verbaliser les régularités structurelles apprises implicitement. Ces connaissances permettent une bonne compréhension de la structure musicale occidentale (Trainor, 2005 ; Trainor, Marie, Gerry, Whiskin, & Unrau, 2012) et vont également conditionner la perception de la justesse qui va s'affiner avec l'âge et l'évolution de la perception mélodique.

### ***Evolution de la perception mélodique***

La perception des sons débute in utero et avec elle la perception mélodique. Déjà dans le ventre de sa mère, à 25-28 semaines de gestation, le fœtus répond aux sons ou vibrations, comme le montre l'augmentation de son rythme cardiaque et de ses mouvements (notamment observée par Kisilevsky & Low, 1998). Le fœtus répond donc à la musique environnante et à la voix de la mère (pour une revue, voir Tan, Pfordresher, & Harré, 2010). Dans le cas du chant, Satt (1984) (cité par Deliège & Sloboda, 2003 et Tan et al., 2010) a observé que le nouveau-né préfère les mélodies chantées par la mère pendant le dernier trimestre de sa grossesse plutôt qu'une mélodie jamais entendue. Ces mélodies seraient donc perçues lors de la gestation et intégrées de manière à être reconnues après la naissance. Le traitement mélodique va cependant évoluer et s'affiner progressivement.

En 1977, Chang et Trehub ont montré que des enfants de 5 à 6 mois ont compris qu'une mélodie transposée n'est pas une nouvelle mélodie. En plus d'être reconnus, les contours mélodiques semblent pouvoir être catégorisés par les enfants dès l'âge de 10 mois (Ferland & Mendelson, 1989). De plus, la position du centre tonal importe peu tant que la relation harmonique entre les notes est respectée. En effet, Plantinga et Trainor (2005) ont observé que les enfants n'avaient pas de préférence entre une mélodie dans sa tonalité de référence et la même mélodie transposée. Ces études confirment donc que très tôt, les enfants ont une représentation précise des contours mélodiques et de la relation globale entre les notes. Notons que la capacité d'extraire les contours mélodiques serait une propriété générale du système auditif (McDermott, Lehr, & Oxenham, 2008) tandis que la discrimination des intervalles musicaux ne serait pas spécifique à ce système (McDermott, Keebler, Micheyl, & Oxenham, 2010). Ceci pourrait expliquer la moindre importance de la taille des intervalles observée par Trehub et ses collaborateurs (1984, 1987). En effet, ils remarquent qu'une mélodie est reconnue malgré des modifications au niveau de la taille des intervalles (Trehub, Bull, & Thorpe, 1984 ; Trehub, Thorpe, & Morrongiello, 1987). Ce phénomène est visible tant que la modification des intervalles n'a pas d'incidence sur le contour général de la mélodie. Progressivement, la taille des intervalles au sein des contours mélodiques est de mieux en mieux perçue. En effet, des déviations de 4, 2, 1, 0.5 et 0.3 demi-tons sont perçues par des enfants de 5 ans et, à partir de 8 ans, les enfants repèrent des déviations de 0.1 demi-ton, comme les adultes (Stalinski, Schellenberg, & Trehub, 2008). Les intervalles deviennent également de plus en plus saillants dans le jugement mélodique. En effet, Stalinski et Schellenberg (2010) ont observé des enfants de 5 à 12 ans ainsi que des adultes dans une tâche de comparaison de mélodies. Les stimuli étaient soit similaires, soit transposés, soit modifiés, soit transposés et modifiés. Ils remarquent que tous les participants utilisent les informations relatives aux intervalles et que ces informations sont de plus en plus utilisées avec l'âge. La notion de tonalité s'affine également à partir de l'âge de 7 ans et semble acquise entre 9 et 12 ans (Hannon & Trainor, 2007 ; Miyamoto, 2007). À l'âge de 12 ans, la perception de la justesse en contexte mélodique n'évoluerait plus que moyennant une pratique musicale (Gooding & Standley, 2001).

De manière synthétique, les études réalisées chez l'enfant montrent qu'ils sont d'abord sensibles aux contours mélodiques et à l'organisation globale des notes au sein de la mélodie avant d'utiliser les informations relatives à la taille des intervalles ou à la tonalité.

### ***Perception de la justesse chez l'auditeur adulte***

Chez l'adulte, des travaux en neuropsychologie et notamment l'observation de patients présentant des lésions cérébrales, suggèrent que le contour et les intervalles seraient traités de manière distincte (Liégeois-Chauvel, Peretz, Babai, Laguitton, & Chauvel, 1998 ; Schuppert, Munte, Wieringa, & Altenmuller, 2000). En effet, le traitement des intervalles peut être affecté sans que celui des contours ne le soit. Cependant, il se pourrait que le traitement des contours soit plus robuste et donc mieux préservé que le traitement des intervalles. Ces conclusions pourraient également être interprétées comme une séparation des processus dédiés aux contours et à la tonalité, étant donné que les intervalles sont largement utilisés pour le traitement de la tonalité (Peretz, 1993). Quelles que soient les étapes dans la perception mélodique et la séparabilité/distinction des traitements, les études effectuées chez l'adulte confirment que les informations relatives au contour mélodique (Dowling & Fujitani, 1970 ; Edworthy, 1985), à la tonalité (Trainor & Trehub, 1992) et aux intervalles (Stalinski et al., 2008) sont intégrées et utilisées sans apprentissage explicite.

Si l'auditeur est capable de discriminer des intervalles musicaux, les nombreux auteurs cherchant à identifier la déviation nécessaire pour qu'une erreur d'intervalle soit détectée ne parviennent pas à un consensus. En effet, cette déviation serait aux alentours de 10 à 45 cents (Moran & Pratt, 1926 ; Rakowski, 1990 ; Rakowski & Miskiewicz, 1985 ; Vurma & Ross, 2006 ; Ward, 1954), entre 50 et 70 cents (Lindgren & Sundberg, 1972, cité par Sundberg, 1979, 1982), au-delà d'un demi-ton (Burns & Ward, 1978 ; Halpern & Zatorre, 1979 ; Rakowski, 1976 ; Schellenberg, 2001 ; Siegel & Siegel, 1977 ; Sundberg, Prame, & Iwarsson, 1996) ou plus récemment entre 20 et 50 cents (Hutchins & Peretz, 2012 ; Hutchins, Roquet, & Peretz, 2012).

Par ailleurs, des différences individuelles apparaissent en ce qui concerne le type d'écoute. Selon Schneider et Wengenroth (2009), les

auditeurs adultes adopteraient un mode d'écoute holistique, spectral ou mixte. Dans le cas d'une écoute spectrale, l'auditeur serait particulièrement sensible à la perception décomposée des harmoniques d'un son, tandis qu'une écoute holistique reflèterait une écoute où harmoniques et fréquence fondamentale sont perçues de manière globale. Notons cependant que, quel que soit le type d'écoute, la perception de la hauteur des sons et donc de la justesse en contexte mélodique ne devrait pas être affectée.

## **Effet de l'expertise sur la perception mélodique**

L'apprentissage implicite ou l'enculturation permet à l'individu de développer des connaissances relatives aux régularités de la structure tonale de sa propre culture et rendent l'auditeur expert de sa culture (Bigand & Delbé, 2010). Cependant, des facteurs tels que l'entraînement musical ou le bagage linguistique et culturel ont une incidence sur la perception mélodique.

### ***Effet de l'entraînement musical***

L'apprentissage explicite des règles musicales, à l'occasion de cours de solfège, d'analyse, d'écoute, de chant choral, d'écriture, d'harmonie mais aussi grâce à la pratique instrumentale, en soliste ou en musique de chambre, entraîne une expertise particulière dont les effets sont régulièrement observés. Par ailleurs, les bénéfices d'une expertise musicale ne se limitent pas à la connaissance des règles musicales. En effet, les conséquences de la pratique intensive d'un instrument de musique sont visibles, tant au niveau anatomique, cognitif, qu'au niveau de la perception musicale et de la justesse.

Les musiciens entraînés ont généralement commencé leur formation musicale dans l'enfance et ont passé un temps considérable à pratiquer et à écouter de la musique (Krampe & Ericsson, 1996 ; Sloboda, 2005). Par exemple, Ericsson, Krampe et Tesch-Römer (1993) rapportent que des musiciens entraînés (des violonistes reconnus) ont accumulé environ 10 000 heures de pratique avant l'âge de 20 ans. Etant donné cette expérience, il n'est pas surprenant d'observer des

différences neuroanatomiques et fonctionnelles entre des non musiciens et des musiciens (pour des revues, voir Trainor & Hannon, 2012 ; Schlaug, 2011 ; Zatorre, Chen, & Penhune, 2007). Si l'écoute ou la pratique musicale influencent la réactivité et l'organisation des zones cérébrales, la pratique musicale va également stimuler et faciliter des compétences plus générales (Patel, 2010). Ces effets de transfert ont été largement étudiés (pour des revues, voir Besson, Chobert, & Marie, 2011 ; Hannon & Trainor, 2007 ; McAuley, Henry, & Tuft, 2011 ; Moreno, Marques, Santos, Santos, Castro, & Besson, 2008). Notons que parmi les compétences non musicales, certaines auront néanmoins une influence indirecte sur le traitement de la musique, comme la mémoire (par exemple, Franklin, Sledge Moore, Yip, Jonides, Rattray, & Moher, 2008 ; George & Coch, 2011 ; Jakobson, Lewycky, Kilgour, & Stoesz, 2008), les fonctions exécutives (Moreno, Bialystok, Barac, Schellenberg, Cepeda, & Chau, 2011) ou encore l'apprentissage implicite de probabilités transitionnelles musicales (François & Schön, 2011 ; François, Tillmann, & Schön, 2012). La littérature est vaste et nous nous limitons ici aux bénéfices de l'expertise musicale sur les compétences musicales requises pour l'évaluation de la justesse.

En ce qui concerne la discrimination de sons purs, Moore (1973) rapporte que des musiciens entraînés peuvent distinguer des sons de 1000 et 1002 Hz. Comparés à des non musiciens, des musiciens entraînés montrent de meilleures capacités de discrimination de sons purs et complexes (Micheyl, Delhommeau, Perrot, & Oxenham, 2006 ; Tervaniemi, Just, Koelsch, Widmann, & Schröger, 2005). En effet, dans l'étude de Micheyl et al. (2006), les seuils de discrimination fréquentielle des non musiciens étaient 6 fois plus grands que ceux des musiciens (formation musicale classique). Ils observent également qu'il suffit de 4 à 8 heures d'entraînement pour que les seuils des non musiciens rejoignent ceux des musiciens alors que ces derniers n'ont pas amélioré leurs performances avec la pratique. Les auteurs confirment donc le bénéfice de l'expertise musicale sur la discrimination de sons mais concluent également que la plupart des individus seraient capables de discriminer des différences de hauteurs minimales grâce à un apprentissage spécifique. La supériorité des musiciens se retrouve également dans les tâches de perception de hauteur de sons isolés (Hutchins & Peretz, 2012), lorsqu'il s'agit d'estimer la taille d'un

intervalle (Russo & Thompson, 2005), de comparer des sons complexes (vocaux ou instrumentaux) ou de timbres différents en contexte isolé (Hutchins, Roquet, & Peretz, 2012 ; Vurma, Raju, & Kuuda, 2010) ou d'intervalles (Zarate, Ritson, & Poeppel, 2012). La supériorité des musiciens ne se limite pas aux tâches relatives aux sons isolés et aux intervalles. En effet, la reconnaissance de mélodies est meilleure chez des musiciens (Orsmond & Miller, 1999). En contexte mélodique, les déviations sont mieux détectées (Hutchins et al., 2012 ; Warrier & Zatorre, 2002), les modifications de contours et d'intervalles sont mieux repérées (Fujiroka, Trainor, Ross, Kakigi, & Pantev, 2004), le traitement de la hauteur tonale est plus efficace (Schön, Magne, & Besson, 2004), l'intégration de la structure tonale est meilleure (par exemple, Koelsch, Jentschke, Sammler, & Mietchen, 2007) et le traitement mélodique est plus rapide (Micheyl et al., 2006 ; Schellenberg & Moreno, 2010 ; Strait, Kraus, Parbery-Clark, & Ashley, 2010). Notons également que les performances de musiciens et non musiciens peuvent être identiques dans des tâches simples mais qu'une différence apparaît avec la complexité de la tâche. Par exemple, Besson et Faïta (1995) ont demandé à des participants de décider si les mélodies présentées étaient familières ou non et de catégoriser la dernière note comme juste ou fausse. Les auteurs remarquent que les performances des musiciens et non musiciens sont similaires lorsque la détection des incongruités est simple (incongruités non diatoniques dans un contexte de mélodies familières) tandis que les musiciens sont plus performants lorsque la tâche nécessite des connaissances musicales (incongruités non diatoniques dans un contexte de mélodies non familières ou incongruités diatoniques).

### ***Effet des bagages linguistique et culturel***

Les variations de la fréquence fondamentale sont présentes en musique mais également dans le langage. Au niveau linguistique, les informations données par ces variations dépendent de la culture. En effet, les locuteurs de langues à tons (par exemple le cantonais ou le mandarin) font appel à des différences de hauteur tonale pour signifier des termes distincts alors que des langues comme le français ou l'anglais utilisent des changements de hauteur pour former des contours intonatifs mais ces variations n'ont pas d'incidence sur la signification



des termes utilisés (Yip, 2002). L'utilisation des hauteurs dans un contexte linguistique faciliterait l'utilisation des hauteurs dans des contextes non linguistiques. Autrement dit, le bagage linguistique de l'auditeur aurait une influence sur la perception du son. Cette hypothèse a été testée par Pfordresher et Brown en 2009. Lors d'une tâche de perception de hauteur de sons isolés, un tel avantage n'apparaît pas mais les locuteurs de langues à tons montrent une meilleure perception lorsqu'il s'agit d'intervalles musicaux. En 2012, Wong, Ciocca, Chan, Ha, Tan et Peretz ont observé les performances au test en ligne de détection d'amusie (Peretz, Gosselin, Tillmann et al., 2008) de deux populations : Canadiens et Hongkongais. Les résultats confirment la supériorité des locuteurs d'une langue à tons pour la tâche relative à la perception mélodique alors qu'aucune différence n'apparaît pour les autres tâches proposées (Wong et al., 2012). L'exposition prolongée aux tons lexicaux permettrait donc le développement d'une habileté spécifique à percevoir les hauteurs, indépendamment des critères d'âge, de niveau d'éducation ou de formation musicale.

Au sein de chaque culture, les auditeurs développent des connaissances implicites par exposition à la musique environnante. Comme indiqué dans la première partie de ce chapitre, la musique occidentale est régie par des règles et l'auditeur occidental sera sensible aux régularités de la structure tonale propre à cette culture. Ce système abstrait est constitué par les fréquences d'occurrence des relations entre les notes. Par exemple, la probabilité que la dernière note d'une mélodie soit la tonique d'une tonalité donnée (1<sup>er</sup> degré) est plus grande que la probabilité que ce soit la sous-dominante (4<sup>ème</sup> degré). La tonique sera donc plus attendue qu'une sous-dominante en fin de mélodie. Ces fonctions tonales et donc les attentes tonales de l'auditeur ont une influence sur la perception de la hauteur (Marmel, Tillmann, & Dowling, 2008). Ces auteurs ont proposé à des non musiciens d'évaluer la justesse de la note finale de mélodies présentées par paires. La note modifiée de 0, 17 ou 35 cents était soit la tonique, soit la sous-dominante. Les résultats indiquent un effet des attentes tonales sur la perception de la justesse. En effet, la discrimination juste/faux était meilleure et plus rapide pour les toniques que pour les sous dominantes (moins reliées à la tonalité de la mélodie). Ce pattern de résultats a été retrouvé pour des modulations plus subtiles (9 cents) avec de

meilleures performances lorsque la modification était effectuée sur la tonique.

Les différentes études présentées ici mettent en évidence l'importance de spécifier l'expertise de l'auditeur. Une récente étude comparant des logopèdes, des musiciens et des non experts confirme la spécificité de l'expertise musicale par rapport à l'expertise langagière dans une tâche utilisant l'apprentissage implicite de régularités musicales (Larrouy-Maestri, Leybaert, & Kolinsky, 2013). En ce qui concerne l'expertise musicale, Seppänen, Brattico et Tervaniemi (2007) ont observé des différences au niveau de la perception selon la pratique musicale. En effet, les musiciens qui pratiquent leur instrument en utilisant une partition ont un meilleur niveau de performance dans une tâche de discrimination de variations de contour ou d'intervalle au sein de séquences mélodiques que les musiciens qui privilégient l'improvisation. Ces études soulignent donc l'intérêt de tenir compte du type d'expertise.

## Synthèse

La justesse vocale en contexte mélodique est régie par les règles musicales d'une culture et se base donc sur un système musical particulier (Cross, 2001). Dans le cadre du système musical occidental, les erreurs mélodiques observables ont trait au contour mélodique, au respect de la taille des intervalles et de la tonalité. La perception de ces erreurs apparaît très tôt dans le développement. Elle serait dépendante de l'évolution des capacités de perception mais également de l'apprentissage implicite des règles musicales propres à la culture environnante dans laquelle l'enfant évolue (Trainor, 2005 ; Trainor, Marie, Gerry, Whiskin, & Unrau, 2012). Concrètement, les enfants sont d'abord sensibles aux contours mélodiques et à l'organisation globale des notes au sein de la mélodie puis tiennent compte de la taille des intervalles. Les informations relatives au contour mélodique (Dowling & Fujitani, 1970 ; Edworthy, 1985), à la tonalité (Trainor & Trehub, 1992) et aux intervalles (Stalinski et al., 2008) sont finalement intégrées par l'auditeur adulte. Ces informations lui permettent d'être sensible aux régularités de la structure tonale de sa propre culture (Bigand & Delbé, 2010) et de développer des attentes mélodiques (Marmel et al., 2008).

Par ailleurs, un entraînement musical a une incidence sur la perception mélodique, tant au niveau des mécanismes sous-jacents à la perception qu'au niveau de la discrimination mélodique. Ce premier chapitre met donc en évidence que les règles musicales régissant l'évaluation de la justesse mélodique sont d'autant plus perçues par des auditeurs qu'ils ont développé une expertise relative au domaine musical. Cependant, l'évaluation d'une performance musicale ne se limite pas à la perception des erreurs mélodiques. Le chapitre suivant expose la complexité des processus d'évaluation et les méthodes actuellement disponibles pour évaluer la justesse vocale en contexte mélodique.



## CHAPITRE II

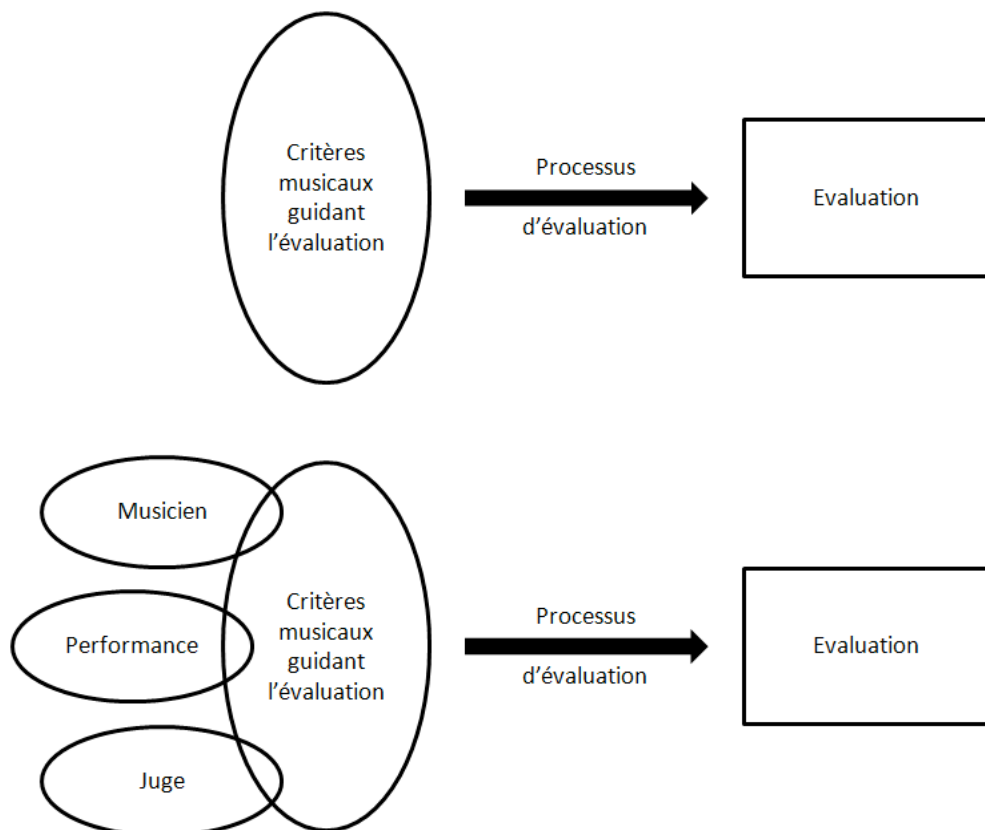
# Evaluation de la justesse vocale

---

*Evaluer... mais comment ?*

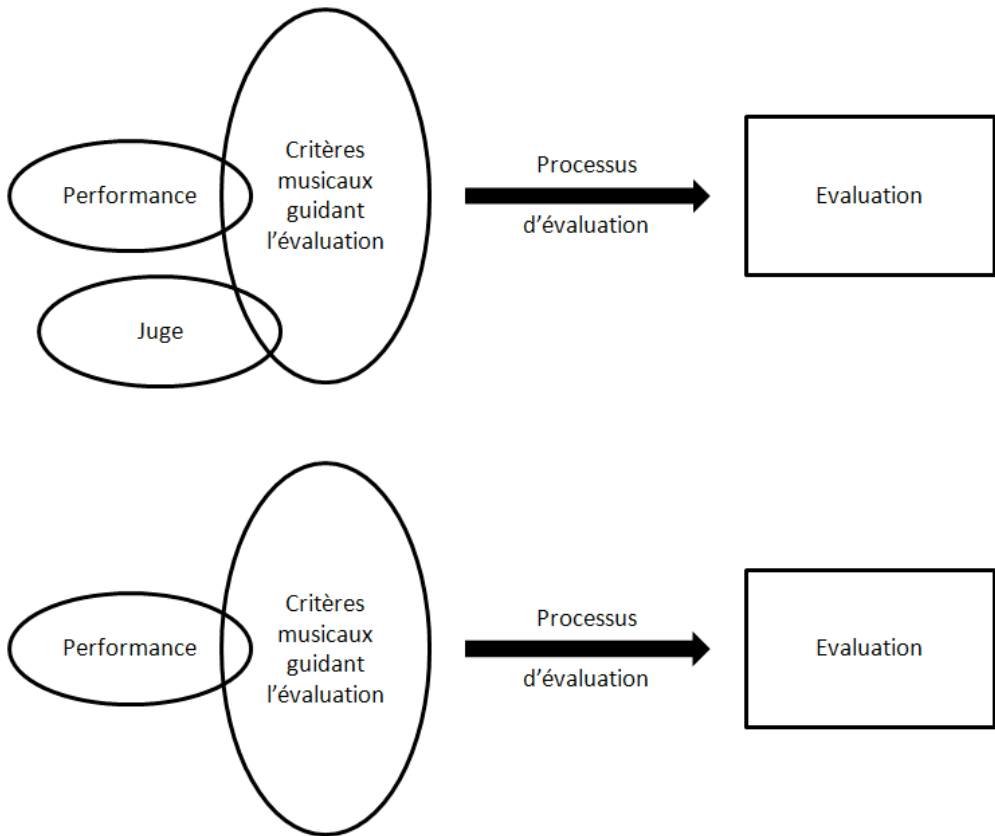
L'évaluation d'une performance musicale peut avoir un objectif formatif, certificatif ou ne pas avoir d'autre objectif que de catégoriser une performance en "bonne" ou "mauvaise" dans des contextes moins formels. McPherson et Schubert (2004) proposent de schématiser le processus idéal d'évaluation par deux éléments : les critères musicaux et l'évaluation de la performance (Figure 5, en haut), une flèche représentant l'utilisation des critères musicaux pour expliquer le jugement.

Cependant, une performance est un événement qui fait intervenir de multiples éléments relatifs au musicien, à la performance et au juge (Godlovitch, 1998). Le modèle de Landy et Farr (1980) ou la revue de littérature de McPherson et Thompson (1998) ont marqué les recherches en pédagogie musicale. Ils mettaient déjà en évidence la complexité de l'évaluation d'une performance musicale (synthétisée au bas de la Figure 5). A la suite de ces premiers travaux, une série d'études a décrit l'influence de facteurs liés aux caractéristiques du musicien (Behne & Wöllner, 2011 ; Davidson & Edgar, 2003 ; Elliott, 1996), à son comportement sur scène (Howard, 2012 ; Juchniewicz, 2008 ; Kurosawa & Davidson, 2005 ; Wapnick, Mazza, & Darrow, 1998 ; Wapnick, Mazza, & Darrow, 2000), à ses expressions faciales (Livingstone, Thompson, & Russo, 2009), à son aspect attractif (Ryan, & Costa-Giomi, 2004 ; Wapnick, Darrow, Kovacs, & Dalrymple, 1997 ; Wapnick et al., 1998, 2000) ou à sa tenue vestimentaire (Griffiths, 2008, 2010 ; Wapnick et al., 2000).



**Figure 5.** Schéma "idéal" (en haut) proposé par McPherson et Schubert (2004). Facteurs en jeu lors de l'évaluation d'une performance musicale (en bas).

D'autres facteurs, liés au mode de présentation, c'est-à-dire visuelle et/ou auditive (Connell, Gai, & Holler, 2013 ; Howard, 2012 ; Thompson, Graham, & Russo, 2005 ; Thompson & Russo, 2007), ou encore relatifs au contexte de l'évaluation (Hash, in press ; Sheldon, 1994), vont influencer l'évaluation d'une performance musicale. Lorsque la performance est évaluée à partir d'un enregistrement, les facteurs liés au musicien sont alors limités (Figure 6).



**Figure 6.** Schéma de l'évaluation subjective (en haut) et objective (en bas) de la justesse vocale.

En ce qui concerne l'évaluation de la justesse vocale, deux méthodes sont utilisées. Nous décrivons ici l'évaluation "subjective" lorsqu'elle est effectuée par des juges et "objective" lorsqu'elle est effectuée sur la base d'analyses acoustiques. Dans le premier cas, les facteurs influençant l'évaluation d'une performance sont limités au juge et à la performance (Figure 6, haut). Les facteurs liés au juge disparaissent lorsqu'il s'agit d'appliquer des programmes informatiques (Figure 6, bas).

# Evaluation subjective de la justesse vocale

## *Etat des lieux*

Afin de catégoriser les performances d'un chanteur, d'observer un trouble musical spécifique, de comparer des performances avec ou sans paroles, les performances mélodiques sont régulièrement évaluées par des juges (par exemple, Alcock, Passingham, Watkins, & Vargha-Khadem, 2000a ; Alcock, Wade, Anslow, & Passingham, 2000b ; Hébert, Racette, Gagnon, & Peretz, 2003 ; Kinsella, Prior, & Murray, 1988 ; Lévêque, Giovanni, & Schön, 2012 ; Prior, Kinsella, & Giese, 1990 ; Racette, Bard, & Peretz, 2006 ; Schön, Lorber, Spacal, & Semenza, 2004 ; Wise & Sloboda, 2008). Dans ces études, des juges qualifiés de musiciens notent des performances enregistrées. Le jugement peut porter sur la réussite globale de la tâche (Alcock et al., 2000a, 2000b ; Kinsella et al., 1988 ; Lévêque et al., 2012) ou être guidé par une grille d'évaluation ou des consignes précises concernant les types d'erreur à détecter. Par exemple, Wise et Sloboda (2008) ont demandé à deux juges, ayant l'habitude de travailler avec des chanteurs occasionnels (professeur ou directeur de chœur), de remplir une grille d'évaluation. Ces auteurs ont proposé une échelle inspirée des travaux de Rutkowski (1990) et Welch, Rush et Howard (1991). Leur grille d'évaluation contient huit catégories. La première correspond à "peu de variations dans le chant, proche de la voix parlée" alors que la dernière catégorie est attribuée aux chanteurs qui produisent la mélodie de manière "juste avec un bon maintien de la tonalité". Ils observent une forte convergence des réponses entre les deux juges mais les critères tels que les variations de hauteur, les contours et le respect de la tonalité sont confondus. Par ailleurs, certains critères ne sont pas spécifiques aux règles musicales relatives à la justesse, comme la présence des paroles de la chanson dans la performance. Le jugement peut être également guidé explicitement en demandant aux juges de repérer des erreurs particulières. Par exemple, Schön et al. (2004) procèdent note par note lorsqu'ils évaluent des performances mélodiques et observent le nombre d'erreurs de contours et d'intervalles. Dans l'étude de Racette et al. (2006), deux juges, l'expérimentateur et un étudiant en musique, ont transcrit manuellement les performances chantées et ont donné un point chaque



fois que la taille des intervalles et leur direction étaient correctes. Dans cette étude, seules les notes pour lesquelles les deux juges donnent un point sont comptées dans le score global, c'est-à-dire qu'un désaccord entre les juges mène au retrait de la note en question.

Plusieurs arguments sont en faveur d'une évaluation effectuée par des juges. Racette et al. (2006) estiment qu'il s'agit d'une méthode valide étant donnée la forte convergence entre les juges (de 79 à 95 % selon l'expérience). Par ailleurs, Schön et al. (2004) rapportent qu'il est difficile de quantifier, à l'aide d'outils informatiques, la justesse de performances vocales étant donné qu'une erreur en entraîne une autre. Finalement, l'utilisation d'une méthode quantitative pour évaluer des mélodies paraît inadéquate à Wise et Sloboda (2008) qui défendent l'aspect écologique d'un jugement humain pour ce type d'évaluation. Au niveau pratique, notons également que ce type de procédure d'évaluation n'est pas particulièrement contraignant. En effet, les résultats (nombre d'erreurs observées, utilisation d'une grille ou score global) nécessitent uniquement un matériel d'enregistrement, de diffusion et la participation de quelques musiciens. Cependant, l'évaluation effectuée par des juges a également des limites. En effet, malgré l'utilisation d'enregistrements et donc le contrôle de nombreux facteurs non musicaux liés au chanteur ou à la modalité de présentation, des facteurs relatifs aux juges et à leur perception de la performance vont influencer leur jugement.

### ***Facteurs liés aux juges***

Des études réalisées dans le domaine de la pédagogie musicale mettent en évidence de nombreux effets liés aux caractéristiques des juges. Par exemple, Wapnick et al. (1997) ont observé que des performances musicales étaient notées différemment selon le genre du juge, avec plus de sévérité de la part des femmes. Les juges vont également tenir compte de leurs préférences musicales et favoriser les pièces qu'ils apprécient particulièrement (Glejser & Heyndel, 2001). Un autre facteur est l'attente du juge concernant le niveau des performances (Cavitt, 1997 ; Duerksen, 1972). Cavitt (1997) a proposé à des juges des enregistrements de performances musicales en variant le label de l'enregistrement (défini comme provenant de musiciens

confirmés ou débutants). Chaque performance était présentée aux juges avec les deux labels et les résultats montrent que les attentes du jury augmentaient avec le niveau supposé des musiciens. En d'autres termes, une performance était jugée plus sévèrement lorsqu'elle était attribuée à des musiciens de haut niveau par rapport aux performances attribuées aux débutants.

Par ailleurs, les juges ne sont pas toujours concordants dans leurs jugements (Thompson & Williamon, 2003). La fiabilité intra et inter-juges observée par Racette et al. (2006) ou Wise et Sloboda (2008) n'est pas toujours présente (Kinsella et al, 1988 ; Prior et al., 1990) et peut varier d'une grille d'évaluation à l'autre (Norris & Borst, 2007). Plusieurs études ont montré que la fiabilité des juges augmente avec leur expertise (Kinney, 2009 ; Morrison, Montemayer, & Wiltshire, 2004) mais cet effet n'a pas toujours été observé (Fiske, 1975, 1977, 1978 ; Wapnick, Flowers, Alegant, & Jasinskis, 1993). Ces résultats contradictoires pourraient s'expliquer par un manque de contrôle au niveau du type d'expertise, comme l'ont remarqué Sofranko et Prosek (2012) dans le domaine de la voix. Ces auteurs ont examiné la concordance de logopèdes, de professeurs de chant et de non experts dans une tâche de jugement de la qualité vocale. Ils observent des différences d'évaluations entre les juges. En effet, les experts en voix dysphoniques (logopèdes) montraient une forte concordance inter-juges, alors que cette concordance était modérée pour les professeurs de chant mais toutefois plus élevée que celle des non experts. En ce qui concerne l'évaluation de la justesse vocale, aucune étude, à notre connaissance, n'a observé l'effet du type et du niveau d'expertise sur la concordance des juges mais, étant donnés les effets de l'expertise musicale sur la perception mélodique (*cf.* Chapitre I), il se pourrait qu'un tel effet se retrouve à ce niveau.

### ***Influence du matériel musical sur la perception mélodique***

La familiarité du matériel musical (Kinney, 2009) ainsi que le tempo et la durée de la performance (Wapnick, Ryan, Campbell, Deek, Lemire, & Darrow, 2005) ou l'ordre de présentation des performances (Flores & Ginsburgh, 1996) ont une incidence sur le jugement d'une

performance musicale. En ce qui concerne l'évaluation d'une performance vocale, le contexte musical (présentation de sons isolés versus en contexte) et la taille des intervalles contenus dans les séquences musicales vont intervenir sur la perception mélodique et donc influencer le jugement de la justesse (par exemple, Hutchins et al., 2012 ; Russo & Thompson, 2005 ; Warrier & Zatorre, 2002).

Warrier et Zatorre (2002) ont observé l'effet du contexte de présentation sur la détection d'erreurs chez des non musiciens en utilisant des sons isolés et des séquences mélodiques (notes répétées ou mélodies familières) où les variations (17, 35 ou 52 cents) étaient placées à la fin des séquences. Chaque note indiquée comme "fausse" devait ensuite être évaluée sur une échelle de 1 (peu) à 3 (très fausse). Ils observent que le contexte mélodique joue un rôle facilitateur dans la détection des variations. Ils en déduisent que le contexte mélodique donne des informations au niveau de la structure tonale qui aident l'auditeur à percevoir des différences qu'il ne perçoit pas dans la tâche de sons isolés. Par ailleurs, les résultats montrent que des variations comprises entre 35 et 52 cents sont jugées comme plus grandes en contexte mélodique qu'en contexte isolé. Par ailleurs, Hutchins et al. (2012) ont observé la détection d'erreurs en contexte mélodique et remarquent que la tolérance est meilleure lorsque la déviation va dans la même direction que le dernier intervalle de la séquence. En d'autres termes, si la mélodie est montante, une note trop haute sera mieux acceptée qu'une note trop basse et inversement. Notons que les études relatives à la perception d'erreurs en contexte mélodique (Hutchins et al., 2012 ; Warrier & Zatorre, 2002) se sont uniquement focalisées sur les notes finales et non sur des modifications dans le décours de la mélodie. Outre l'effet du contexte musical (notes isolées, intervalles ou mélodies), la longueur des séquences musicales ainsi que leur vitesse de présentation ont probablement une influence sur la perception de la justesse étant donné que ces deux éléments interviennent dans la reconnaissance de notes (Akiva-Kabiri, Vecchi, Granot, Basso, & Schön, 2009 ; London, 2012).

Les notes sont liées par différents intervalles au sein d'une mélodie (Chapitre I) et l'estimation de la taille d'un intervalle est elle-même affectée par différents facteurs. Par exemple, Russo et Thompson (2005) ont évalué l'effet de la taille des intervalles (du quart de ton à l'octave),

de leur place dans le registre (aigu ou grave) et de leur direction (ascendant ou descendant) sur l'estimation de leur taille. Les résultats montrent que les intervalles présentés dans le registre aigu sont considérés comme plus grands que ceux présentés dans le registre grave. Par ailleurs, les intervalles sont estimés comme plus grands lorsqu'ils sont descendants que lorsqu'ils sont ascendants. Une interaction est également visible entre les deux modalités. En effet, les intervalles ascendants sont perçus comme plus grands que les intervalles descendants lorsqu'ils sont présentés dans le registre aigu, tandis que les intervalles descendants sont perçus comme plus grands que les intervalles ascendants lorsqu'ils sont présentés dans le registre grave. L'estimation de la taille d'un intervalle est donc affectée par le registre et la direction. Se focalisant sur la justesse, Vurma et Ross (2006) ont demandé à des musiciens d'estimer la justesse de différents intervalles mélodiques (seconde mineure, quarte augmentée, quinte juste) produits par des chanteurs classiques, de manière ascendante et descendante. Tous les participants jugent des intervalles déviés de 20 à 25 cents comme justes mais une différence apparaît selon la taille des intervalles, les grands intervalles étant plus souvent jugés faux que les secondes mineures, quelle que soit leur direction (ascendante ou descendante).

Ces études mettent donc en évidence que les caractéristiques musicales d'une performance (contexte musical ou taille des intervalles) influencent la perception mélodique et donc l'évaluation de la justesse. Par ailleurs, une performance vocale en contexte mélodique sous-entend la présence d'éléments non mélodiques tels que le rythme, le tempo, la qualité de la voix ou encore le timbre. Prince (2011) rapporte que différents éléments musicaux vont interagir dans la perception musicale. Dans les études (*cf. supra*, Etat des lieux) utilisant une méthode subjective d'évaluation de la justesse, les juges avaient la consigne de repérer des erreurs (Racette et al., 2006 ; Schön et al., 2004) ou d'estimer la performance vocale dans sa globalité (Alcock et al., 2000a, 2000b ; Kinsella et al., 1988 ; Lévêque et al., 2012). Etant donnée l'interaction entre les éléments mélodiques et non mélodiques, il n'est pas évident que des juges puissent ne tenir compte que de l'aspect mélodique. Finalement, le fait de repérer des erreurs (Racette et al., 2006 ; Schön et al., 2004) permet de guider l'évaluation de la justesse

mais ne donne pas d'indication quant à la taille de l'erreur, d'autant plus que l'oreille humaine aurait tendance à catégoriser une hauteur en fonction de la note la plus proche (Hutchins et al., 2012).

## **Evaluation objective de la justesse vocale**

Les méthodes d'évaluation basées sur l'analyse acoustique des performances vocales permettent de contourner les facteurs liés aux juges et à leur perception (voir Figure 6, bas). Depuis le SINGAD (SINGing Assessment and Development) de Howard et Welch (1989) ou la méthode d'Elmer et Elmer (2000), les méthodes dites objectives ont été développées pour être majoritairement utilisées dans les recherches actuelles (Dalla Bella, Berkowska, & Sowinski, 2011). Les objectifs sont, par exemple, d'examiner les causes d'un trouble de justesse (pour une revue, voir Hutchins & Peretz, 2012) d'estimer la capacité d'une population à chanter juste ou d'observer des profils de chanteurs (Dalla Bella & Berkowska, 2009 ; Dalla Bella, Giguère, & Peretz, 2007 ; Lévêque, Amy de la Bretèque, Giovanni, & Schön, 2009 ; Pfordresher & Brown, 2007 ; Pfordresher, Brown, Meier, Belyk, & Liotti, 2010). Quelles que soient les tâches proposées aux participants, la procédure est d'enregistrer une performance, d'effectuer des analyses acoustiques du signal pour quantifier les variations de la fréquence fondamentale et d'en estimer la similitude avec un modèle musical.

### ***Outils d'analyse***

Pour l'analyse acoustique d'un son, les chercheurs observent de façon quasi consensuelle la partie stable des notes afin d'en extraire la fréquence fondamentale. En effet, ce sont les voyelles qui donnent l'essentiel des informations acoustiques relatives à la hauteur (Murayama, Kashiwagi, Kashiwagi, & Mimura, 2004) et qui marquent le début d'un son musical (Sundberg & Bauer-Huppmann, 2007). La partie stable constitue donc le segment de choix pour quantifier la fréquence fondamentale d'une note. Notons que Pfordresher et Brown (2007) ont mené des analyses comparatives de différentes stratégies de segmentation (avec ou sans les attaques et liaisons entre les notes) et ont observé une forte corrélation ( $r > .99$ ) entre les hauteurs évaluées sur les segments stables ou l'entièreté du signal. La segmentation peut

se faire manuellement mais également automatiquement, à l'aide de programmes informatiques. Cependant, si l'automatisation est possible lorsque le signal est de bonne qualité, elle n'est pas forcément adaptée à des séquences chantées par des chanteurs occasionnels. En effet, un silence ou l'altération du son vont conduire à segmenter le signal, sans pour autant qu'il s'agisse de deux notes différentes. De même, selon la sensibilité des mesures de hauteur, une note dont la fréquence fondamentale est instable pourrait être considérée comme deux éléments distincts. Le logiciel Praat (Boersma, 2001 ; Boersma & Weenink, 2008) est majoritairement cité pour l'inspection visuelle du signal et l'identification des parties stables de chaque note. Ce logiciel, largement utilisé en phonétique et en clinique logopédique, a été créé pour l'analyse de la qualité vocale.

Une fois que les notes sont sélectionnées, plusieurs possibilités s'offrent au chercheur pour extraire l'information relative à la fréquence fondamentale (pour des revues, voir Gomez, Klapuri, & Meudic, 2003 ; Hess, 1983). L'extraction est effectuée à l'aide d'algorithmes qui peuvent être classés en trois groupes principaux, selon qu'ils privilégient l'information temporelle, spectrale ou les deux. En ce qui concerne l'analyse des sons vocaux, la méthode dite d'auto-corrélation semble préférable (Boersma, 1993) et correspond à une analyse où l'aspect temporel est privilégié. Cette technique a l'avantage d'être rapide mais a l'inconvénient de manquer de précision. Etant donné que le logiciel Praat (Boersma, 2001 ; Boersma & Weenink, 2008) est majoritairement utilisé dans les études relatives à la justesse vocale et que ce logiciel fonctionne sur la base d'auto-corrélations, il n'est pas surprenant que les auteurs mentionnent la présence d'erreurs de détection de hauteur. En effet, il est précisé dans plusieurs articles que des erreurs de type "saut d'octave" sont corrigées manuellement par les chercheurs (par exemple, Dalla Bella et al., 2007 ; Granot, Israel-Kolatt, Gilboa, & Kolatt, in press ; Pfordresher et al., 2010). Une alternative est l'utilisation de l'algorithme de Yin (De Cheveigné & Kamahara, 2002). Cependant ces analyses ont jusqu'à présent été limitées à des sons isolés (Hutchins & Peretz, 2012) et la programmation pour l'analyse de séquences de notes n'a pas été menée par ces auteurs. Il s'agit également d'une méthode basée sur l'information temporelle mais elle est considérée par ses concepteurs comme une version améliorée de l'auto-corrélation. Afin de limiter les

erreurs d'estimation de la fréquence fondamentale, une autre possibilité consiste à régler manuellement les paramètres d'analyse, selon le signal acoustique à analyser, comme le permettent les outils développés par l'Institut de Recherche et Coordination Acoustique/Musique (Ircam). Ces outils ont été conçus dans un but de recherche musicale et avec l'objectif d'aider les compositeurs dans l'analyse et la synthèse d'éléments acoustiques. A notre connaissance, ils n'ont pas été expérimentés dans le cadre de recherches sur la justesse vocale.

## ***Tâches expérimentales***

Afin d'examiner la justesse d'une performance vocale, plusieurs tâches de production vocale peuvent être proposées. Il s'agit soit de répéter des sons isolés, des intervalles ou des séquences mélodiques, soit des mélodies connues du participant. Si les analyses acoustiques sont similaires, le calcul des erreurs va diverger selon la tâche proposée.

### **Tâches de répétition**

Les tâches de production de sons isolés utilisent soit des sons purs (Watts, Murphy, & Barnes-Burroughs, 2002), soit des sons complexes (Amir, Amir, & Kishon-Rabin, 2003 ; Hutchins & Peretz, 2012 ; Moore, Estis, Gordon-Hickey, & Watts, 2008 ; Moore, Keaton, & Watts, 2007 ; Nikjeh, Lister, & Frisch, 2009 ; Pfordresher & Brown, 2007, 2009 ; Pfordresher et al., 2010). Notons que les sons complexes à répéter sont soit vocaux, soit synthétiques. La voix du participant peut également servir à la création des stimuli (Hutchins & Peretz, 2012 ; Moore et al., 2008). La consigne est alors de reproduire le son entendu, la performance est comparée à la cible et la déviation peut ainsi être calculée par rapport au tempérament égal.

Lorsque le sujet doit reproduire des intervalles, deux mesures sont possibles. Soit chaque note est comparée à la note cible, soit la distance entre les deux notes est calculée et comparée à l'intervalle cible. Certaines études tiennent compte de ces deux mesures (par exemple Granot et al., in press ; Pfordresher & Brown, 2007, 2009, Pfordresher et al., 2010). La répétition de séquences mélodiques a été proposée par Pfordresher et al. (2007, 2009, 2010). Le principe est le même que pour les intervalles. En effet, chaque note peut être comparée à la note cible

ou les notes sont comparées entre elles. Outre la précision des intervalles, cette tâche permet d'examiner la précision vocale lorsque des notes identiques se suivent.

Ces différentes tâches donnent des indications sur la possibilité de répéter avec précision une note donnée, des intervalles ou encore sur la précision vocale. Cependant, ces tâches sont peu écologiques étant donné que des chanteurs occasionnels sont amenés à chanter des mélodies plus qu'à répéter des sons isolés. L'analyse de performances vocales en contexte mélodique semble donc plus adéquate afin d'évaluer la justesse de chanteurs occasionnels.

### **Tâche de production en contexte mélodique**

Dalla Bella et al. (2007, 2009) ainsi que Pfordresher et al. (2010) ont demandé à des chanteurs occasionnels de produire de mémoire une chanson familière et ont mesuré objectivement les performances. Notons que les représentations en mémoire de ces mélodies peuvent varier d'un individu à l'autre et que les mélodies sont donc chantées dans différentes tonalités. De ce fait, les mesures utilisées pour analyser les tâches de répétition (sons isolés, intervalles ou séquences mélodiques) ne sont donc plus adaptées et sont alors basées sur la relation entre les notes.

Comme exposé au chapitre I, plusieurs erreurs sont observables : sur le contour mélodique, c'est-à-dire un changement de direction de l'intervalle par rapport à l'intervalle inscrit sur la partition, et sur la distance entre les notes. En effet, une mélodie étant une suite d'intervalles, il est alors possible d'observer la direction des intervalles (ascendants versus descendants) et d'estimer la moyenne des déviations des différents intervalles constituant la mélodie. De plus, au sein même d'une mélodie, la même séquence mélodique peut apparaître, que ce soit lors d'une répétition ou d'un refrain. Dans ce cas, il est possible de comparer une note avec une cible, c'est-à-dire avec la note produite la première fois au cours de la mélodie (Dalla Bella & Berkowska, 2009 ; Dalla Bella et al., 2007). Cette mesure informe le chercheur sur d'éventuelles modulations au cours de la mélodie.

L'aspect écologique des performances en contexte mélodique est indéniable mais la quantification des erreurs est complexe. Si les tâches



de répétition ont un intérêt certain et que leur évaluation est particulièrement aisée, il n'est pour l'instant pas prouvé qu'elles puissent remplacer les informations obtenues à l'aide d'une performance vocale en contexte mélodique. Il serait donc intéressant de comparer les résultats aux différentes tâches afin de choisir la tâche expérimentale la plus adéquate.

### ***Limites de l'évaluation objective***

Si l'évaluation objective des performances vocales permet d'éviter les biais liés aux juges, des limites persistent. En effet, les outils informatiques sont programmés par des individus et les conclusions dépendront des choix effectués par le chercheur qui programme et interprète les données.

#### **Choix du seuil de justesse**

En ce qui concerne la déviation des intervalles, le seuil à partir duquel une performance est considérée comme fautive peut être calculé sur la base du seuil statistique de déviation par rapport au score d'un groupe contrôle (par exemple +3 écarts-types). Cependant, comme le soulignent Pfordresher et Brown (2009), le seuil statistique dépend de l'échantillon de la population qui sert de contrôle. Une alternative est de se baser sur la notation musicale elle-même. Dans ce cas, une note (ou un intervalle) est considérée comme fautive lorsque sa hauteur est déviée de plus de 100 cents par rapport à la cible, c'est-à-dire une erreur de plus d'un demi-ton. Pfordresher et Brown (2009) utilisent ce seuil car ils estiment que le demi-ton est prédominant dans la musique occidentale. En effet, ce seuil serait intégré par simple exposition à la musique environnante et constitue donc une limite pour ces auteurs. Cependant, Zatore et al. (2012) remarquent que cette limite est apprise et intériorisée à la suite d'un entraînement musical et non par simple exposition.

D'autres auteurs utilisent un seuil de 50 cents (voir Brattico, Tervaniemi, Näätänen, & Peretz, 2006). Par exemple, Hutchins et Peretz (2012) estiment que la plupart des notes qui présentent une erreur de 50 à 100 cents sont perçues comme fautes, tant en comparaison de notes qu'en contexte mélodique. En 2012, Hutchins et al. observent

qu'une note située plus qu'à mi-chemin d'une note voisine sera perçue comme étant cette dernière. Cependant, si de nombreuses études sur la perception d'intervalles appuient leur choix (par exemple Moran & Pratt, 1926 ; Rakowski 1990 ; Rakowski & Miskiewicz 1985 ; Vurma & Ross, 2006 ; Ward, 1954) d'autres études observent des seuils de tolérance supérieurs à 50 cents (Burns & Ward, 1978 ; Halpern & Zatorre, 1979 ; Lindgren & Sundberg, 1972, cité par Sundberg 1979, 1982 ; Rakowski, 1976 ; Schellenberg, 2001 ; Siegel & Siegel, 1977 ; Sundberg, Prame, & Iwarsson, 1996). Outre ces informations contradictoires au sujet de la discrimination d'intervalles, aucune étude n'a validé qu'en contexte mélodique, une déviation de 50 cents représente une erreur. En effet, les travaux utilisant des contextes mélodiques (Hutchins et al., 2012 ; Warrier & Zatorre, 2002) ne modifient que la dernière note de la mélodie et l'effet du contexte n'a pas été observé.

Une alternative à cette problématique de seuil est donc de considérer la déviation en tant que telle. En effet, si cette mesure ne permet pas de classer des performances vocales en deux catégories (juste versus fausse), elle permet néanmoins d'observer précisément une performance sur un continuum allant du "parfaitement juste" à "beaucoup moins juste".

### **Choix des critères**

Les analyses précédemment décrites permettent d'examiner deux règles musicales (respect des contours et des intervalles) mais l'aspect global de la performance, c'est-à-dire le respect du centre tonal, n'est pas observé. En effet, les transpositions qui ont été évoquées par Dalla Bella et al. (2007, 2009) ne permettent pas d'observer des modulations multiples au cours de la mélodie ou une déviation progressive du centre tonal. Dans leurs études, ils comparent deux séquences mélodiques similaires au sein d'une même mélodie mais une absence d'erreur ne permet pas d'écartier que le participant ait modulé plusieurs fois avant de retrouver la bonne tonalité.

Comme décrit dans le chapitre I de cette thèse, les règles musicales du système tonal occidental sont dictées par la relation entre les notes (contours mélodiques et intervalles musicaux) et par l'organisation tonale. Par ailleurs, le développement de la perception mélodique

confirme l'importance de ces règles et leur intégration par simple exposition à la musique environnante. Pour une évaluation objective et pertinente de la justesse en contexte mélodique, il semble donc approprié d'observer le respect de ces différentes règles ainsi que leur importance pour l'auditeur.

## Synthèse

Deux méthodes sont utilisées pour évaluer la justesse d'une performance vocale en contexte mélodique : l'évaluation par des juges et l'évaluation à l'aide d'outils informatiques. Dans le cadre de cette thèse, nous nommons "subjective" la méthode faisant appel à des juges experts dans le domaine musical (par exemple, Alcock et al., 2000a, 2000b ; Hébert et al., 2003 ; Kinsella et al., 1988 ; Lévêque et al., 2012 ; Prior et al., 1990 ; Racette et al., 2006 ; Schön et al., 2004 ; Wise & Sloboda, 2008) et "objective", la méthode utilisant des analyses acoustiques (par exemple, Dalla Bella et al., 2007, 2009 ; Elmer & Elmer, 2000 ; Lévêque et al., 2009 ; Pfordresher et al., 2007, 2010). Cette dernière a été développée depuis une décennie et semble actuellement préférée par les équipes de recherches souhaitant observer la justesse d'une performance (Dalla Bella et al., 2011). En effet, la méthode subjective est discutable étant donnée l'influence des facteurs liés aux juges. De plus, les caractéristiques du matériel musical à évaluer influencent la perception humaine et donc l'évaluation de la justesse (voir notamment Hutchins et al., 2012 ; Russo & Thompson, 2005 ; Vurma & Ross, 2006 ; Warrier & Zatorre, 2002). A l'inverse, l'utilisation d'outils informatiques permet de contourner les facteurs liés aux juges et à leur perception. Cependant, cette méthode a également des limites étant donné la complexité d'une production vocale et le choix des critères d'analyse (erreurs mélodiques à observer) dont la pertinence n'a à ce jour pas été confirmée.

Les deux méthodes décrites dans ce chapitre sont donc discutables et l'objectif principal de cette thèse est de développer une méthode objective mais pertinente d'évaluation de la justesse vocale en contexte mélodique.

Après avoir vu les types d'évaluation existants, les chapitres suivants visent à décrire deux populations concernées par l'évaluation

de leurs performances, à savoir les chanteurs occasionnels (Chapitre III) et les chanteurs lyriques (Chapitre IV).

## CHAPITRE III

# La justesse des chanteurs occasionnels

---

### *Evaluer la justesse... de qui ?*

Comme l'indiquent les deux premiers chapitres, la musique occidentale est régie par des règles qui définissent la justesse vocale et deux méthodes sont actuellement utilisées pour définir la qualité d'une performance vocale. Nous détaillons ici la population principale concernée par l'évaluation : les chanteurs occasionnels. En effet, l'activité de chanter commence très tôt (Papoušek, 1996), est omniprésente (Welch, 2005) et appréciée (Chong, 2010). De plus, le chant est une activité musicale particulièrement accessible (Casals, Ayats, & Vilar, 2010). En effet, contrairement aux instruments musicaux, les schémas moteurs sont déjà développés étant donné que les mécanismes vocaux sont exploités en voix parlée. Dans un premier temps, nous nous penchons sur le développement de la justesse et la description des chanteurs occasionnels. Par ailleurs, certains chanteurs occasionnels éprouvent des difficultés et nous décrivons, dans un second temps, les hypothèses avancées pour expliquer un trouble de justesse. Dans un troisième temps, nous nous penchons sur la variabilité de la justesse vocale et sur l'incidence du type de tâche proposé, de la méthode d'évaluation, des caractéristiques du sujet ou encore de la situation d'évaluation sur la qualité d'une performance vocale.

## **Développement du chant et de la justesse**

Tout comme le langage, les habiletés musicales se développent spontanément et de manière implicite dès la naissance (Curtis & Bharucha, 2009 ; Perruchet, 2008 ; Tillmann, Bharucha, & Bigand, 2000). Au niveau des premières productions vocales, la mélodie des cris

de nouveau-nés semble influencée par la forme prosodique de la langue dans laquelle il évolue (Mampe, Friederici, Christophe, & Wermke, 2009). Pendant les premiers mois de vie, les enfants vocalisent (Papoušek, 1996) et produisent des éléments mélodiques tels que des glissandi, pouvant être interprétés comme des précurseurs de l'intonation musicale (Adachi, 2011). A la fin de la première année, des productions vocales plus structurées et plus systématiques émergent, avec des voyelles chantées sur des notes relativement stables. Progressivement, les mélodies chantées deviennent reconnaissables vers l'âge de 18 mois et prennent la forme de courtes phrases musicales répétées de nombreuses fois (Dowling, 1999 ; Ostwald, 1973 ; Welch, 2006) pour ensuite se diversifier.

Hannon et Trainor (2007) proposent que l'exposition quotidienne à une culture spécifique permettrait de développer de manière systématique des structures cérébrales spécifiques à la musique. Un entraînement formel, par des cours de musique par exemple, mènerait au développement des structures en place et de processus plus généraux tels que l'attention et les fonctions exécutives. Notons qu'une pratique musicale (formelle ou informelle) a de nombreuses répercussions (Bond, 2012 ; Seddon, 2011), que ce soit au niveau physique (Clift, Hancox, Morrison, Hess, Kreutz, & Stewart, 2009), psychologique (Clift & Hancox, 2001 ; Clift, Hancox, Morrison, Hess, Kreutz, & Stewart, 2007 ; Davidson, 2011 ; Kreutz, Bongard, Rohrmann, Hodapp, & Grebe, 2004 ; MacDonald, Kreutz, & Mitchell, 2012 ; Sandgren, 2009 ; Welch, 2005), au niveau du développement d'habiletés sociales (Bailey & Davidson, 2005 ; Davidson, 2011 ; Louhivuori, Salminen, & Lebaka, 2005 ; Odena, 2007 ; Portowitz, Lichtenstien, Egorov, & Brand, 2009), des habiletés cognitives (Schellenberg, 2009 ; Schellenberg & Moreno, 2010 ; Schlaug, Norton, Overy, & Winner, 2005) ou encore langagières (pour une revue, voir Larrouy-Maestri, Leybaert, & Kolinsky, 2013). Il n'est donc pas étonnant que ces activités soient encouragées dans le contexte scolaire.

En ce qui concerne le chant, le développement des habiletés vocales est lié à l'étendue vocale (Welch, 1979) et plus généralement à l'âge de l'enfant (pour une revue, voir Hedden, 2012). L'étude décrite par Welch (2009), menée sur 8000 enfants de 150 écoles primaires du Royaume-Uni, confirme l'importance de l'âge dans l'évolution vocale

mais également le bénéfice d'un programme d'entraînement sur le développement du chant. Par ailleurs, d'autres facteurs vont intervenir dans le développement de la justesse, tels que le genre, la langue maternelle ou encore le milieu socio-culturel (Demorest & Clements, 2007 ; Leighton & Lamont, 2006 ; Welch, 2006, 2009).

L'acquisition de la justesse vocale chez l'enfant a suscité de nombreuses études depuis les années 1970 (Apfelstadt, 1984 ; Davidson, 1994 ; Dowling, 1999 ; Flowers & Dunne-Sousa, 1990 ; Geringer, 1983 ; Greenberg, 1979 ; Hargreaves, 1996 ; Moog, 1976 ; Rutkowski, 1990, 1997 ; Rutkowski & Miller, 2003 ; Welch, 1979, 1986, 1994, 1998, 2002, 2006 ; Welch et al., 1991). De ces études, ressortent trois critères : le respect des contours mélodiques, des intervalles et de la tonalité.

Concernant les **contours mélodiques**, l'enfant calque ses productions vocales sur les schémas mélodiques proposés par l'adulte mais il faudra attendre le contrôle conscient de ses productions vocales et l'augmentation de l'étendue vocale pour que les contours se précisent. Observant les performances vocales de 93 enfants âgés de 3 à 5 ans dans une tâche d'imitation de contours mélodiques, Flowers et Dunne-Sousa (1990) remarquent que les enfants de 5 ans sont capables de reproduire fidèlement le contour mélodique de 84 % des séquences de notes proposées. Le **respect des intervalles** progresse fortement entre l'âge de 7 et 11 ans. Welch (2006) observe que 30 % des enfants de 7 ans ne seraient pas capables de restituer vocalement les intervalles entre les notes d'une mélodie populaire alors qu'à l'âge de 11 ans, seulement 4 % des enfants éprouveraient encore cette difficulté. L'étape relative au **respect de la tonalité** est décrite comme la plus tardive. Alors qu'une certaine stabilité est observée à l'âge de 5 ans (Dowling, 1999), 68 % des enfants de trois ans observés par Flowers et Dunne-Sousa (1990) changent une à plusieurs fois de tonalité au cours d'une chanson. Ce pourcentage décroît avec l'âge et l'élargissement de l'étendue vocale de l'enfant.

L'ensemble des résultats suggère que la capacité à chanter "juste" se met en place progressivement pour devenir proche des performances observées chez l'adulte. En effet, seulement 7 % des enfants de 11 ans chanteraient faux (Welch, 1994). Notons cependant que l'évolution de la

justesse est d'autant plus évidente que les tâches sont simples (Welch, Sergeant, & White, 1996, 1997, 1998). En effet, ces auteurs ont conduit une étude longitudinale de 1996 à 1998 et montrent que des enfants de niveau primaire progressent davantage sur des tâches simples (comme reproduire un schéma mélodique général, un court fragment mélodique ou des notes isolées) par rapport à des tâches plus complexes (production de chants "construits").

Comparé à la littérature abondante dans le domaine de l'éducation musicale, relativement peu d'études se sont intéressées à la justesse vocale chez l'adulte. En 2007, Dalla Bella et al. (2007) ont demandé à 62 chanteurs occasionnels de chanter "Gens du pays", un chant d'anniversaire typique du Québec. A la suite de l'analyse acoustique des productions, ils ont observé que la majorité de ces participants chantait relativement juste. A cette même époque, Pfordresher et Brown (2007) trouvent des résultats similaires. En effet, en analysant l'imitation de séquences de quatre sons, ils ont observé que 87 % des 78 participants non musiciens étaient capables de chanter juste.

## **Les troubles de justesse**

Les conclusions des études relatives au développement de la justesse et aux performances de chanteurs occasionnels témoignent de la capacité à chanter juste pour une majorité de la population. Cependant, une partie des chanteurs occasionnels, estimée à 10-20 % de la population (Dalla Bella & Berkowska, 2009 ; Pfordresher et al., 2010), présente des difficultés. Les causes d'un problème de justesse ont intéressé de nombreux chercheurs et quatre types de troubles ont été documentés : perceptif, mémoriel, sensorimoteur et moteur.

### ***Trouble perceptif***

Comme l'indique le nom désignant les personnes qui chantent faux, "tone deaf", la difficulté à chanter juste a longtemps été associée à un déficit de discrimination des hauteurs de sons musicaux. On comprend aisément qu'un trouble au niveau de la perception des hauteurs engendre des difficultés à respecter la relation entre les notes. En effet, un tel trouble peut affecter l'encodage de la mélodie à répéter,



ou encore l'intégration du feed-back sonore nécessaire au contrôle de la justesse. Cette hypothèse a été vérifiée au travers d'études menées dans le domaine de la surdité. Par exemple, Nakata, Trehub, Mitani et Kanda (2006) ont comparé les performances vocales d'enfants sourds munis d'implants cochléaires et d'enfants normo-entendants. Ils ont remarqué que le rythme était correctement produit tandis que les contours mélodiques ne correspondaient pas à la mélodie cible pour les enfants sourds. Un trouble perceptif peut également être observé chez des individus définis comme amusiques. En effet, l'amusie se traduit par des difficultés dans le domaine musical, plus précisément au niveau du traitement des hauteurs (Foxton, Dean, Gee, Peretz, & Griffiths, 2004 ; Hyde & Peretz, 2004), qui ne sont pas causées par un déficit auditif, cognitif, neurologique, ou par une sous-exposition à un environnement musical (Ayotte, Peretz, & Hyde, 2002 ; Peretz, Ayotte, Zatorre, Mehler, Ahad, Penhure, & Jutras, 2002 ; Peretz, 2008). Concernant les productions vocales, le profil des personnes dites amusiques se traduit par un chant ponctué d'erreurs mélodiques (Ayotte et al., 2002 ; Dalla Bella, et al., 2009 ; Tremblay-Champoux, Dalla Bella, Phillips-Silver, Lebrun, & Peretz, 2010). Notons cependant que de récentes études montrent qu'il existe aussi des amusiques "perceptifs" chez qui la production est préservée (Dalla Bella et al., 2009 ; Loui, Guenther, Mathys, & Schlaug, 2008). L'amusie n'entraîne donc pas systématiquement des troubles au niveau du chant, notamment lorsque les personnes amusiques chantent immédiatement après la présentation d'un modèle cible (Tremblay-Champoux et al., 2010). De nombreuses recherches sont en cours pour mieux comprendre les corrélats cérébraux de ce trouble (Hyde, Zatorre, Griffiths, Lerch, & Peretz, 2006 ; Hyde, Zatorre, & Peretz, 2011). Quoi qu'il en soit, la majeure partie des chanteurs ayant des difficultés à chanter juste n'a pas de trouble perceptif lié à une surdité ou à une amusie. Par ailleurs, une série d'études n'est pas parvenue à observer une corrélation significative entre les performances à des tâches de production et de perception (Bradshaw & McHenry, 2005 ; Dalla Bella et al., 2007 ; Lévêque et al., 2009 ; Pfordresher & Brown, 2007 ; Wise & Sloboda, 2008). Alors que l'hypothèse d'un déficit perceptif peut être soutenue pour une partie des mauvais chanteurs présentant un problème de justesse, les causes d'un trouble de justesse ne peuvent pas se résumer à cette hypothèse.

## ***Trouble relatif à la mémoire des hauteurs***

Chanter une mélodie implique le stockage en mémoire à long terme d'une trace suffisamment solide du schéma mélodique. L'hypothèse d'un trouble relatif à la mémoire des hauteurs est soutenue par des études menées sur des sujets dits amusiques. En effet, l'étude de Berkowska et Dalla Bella (2009) indique que la mémoire des hauteurs est clairement déficitaire chez une partie des individus amusiques. Par ailleurs, leurs performances sont particulièrement fausses lorsqu'ils chantent une mélodie populaire sans les paroles. De plus, les études sur les effets d'interférence lors de tâches de mémoire musicale montrent une sensibilité aux interférences auditives chez les participants non musiciens (Estis, Coblenz, & Moore, 2009 ; Estis, Dean-Claytor, Moore, & Rowell, 2011) et donc l'importance de l'aspect mémoriel dans la production mélodique. Les résultats de Wise et Sloboda (2008) fournissent également des arguments en faveur d'un déficit de mémoire chez des chanteurs ayant des difficultés à chanter juste. Ils ont proposé aux participants de répéter des séquences courtes et longues. En comparant deux groupes ("bons" et "mauvais" chanteurs), ils observent que la différence de performances entre les groupes se creuse pour les séquences longues, avec des difficultés importantes pour les "mauvais" chanteurs. Ils observent également que l'accompagnement musical, qui peut partiellement alléger la charge en mémoire de travail, contribue à une meilleure justesse. Au vu de ces études, la relation entre la mémoire et la justesse est évidente mais il reste néanmoins à préciser cette relation étant donné qu'un trouble au niveau de la mémoire des hauteurs peut être secondaire à des difficultés de perception ou de production.

## ***Trouble d'intégration sensorimotrice***

Un trouble du lien existant entre perception et production reste peu exploré. Les travaux de Pfordresher et al. (2007, 2009) apportent quelques arguments en faveur de cette hypothèse. En effet, ces auteurs ont observé des chanteurs occasionnels dans des tâches d'imitation et notent une compression des intervalles qui témoignerait de déformations dans le passage de l'intention à l'action (Pfordresher et al. 2009). Dans ce cas, la représentation sonore de la mélodie serait

correcte mais subirait des déformations lors de la projection vers les représentations motrices en vue de la production. De plus, ils observent que la présence d'un feedback n'est pas toujours bénéfique lorsqu'il est proposé aux chanteurs éprouvant des difficultés à chanter juste (Pfordresher et al., 2007). Un déficit au niveau sensorimoteur pourrait affecter à la fois la programmation motrice d'une mélodie stockée en mémoire, la répétition d'une mélodie et l'intégration du feedback pendant la production du chant.

Un autre argument en faveur de cette hypothèse est l'effet du timbre observé lors de tâches d'imitation de sons isolés ou de séquences mélodiques. En effet, les tâches d'imitation de notes isolées sont mieux réussies quand le modèle est la propre voix du sujet, que ce soit chez les enfants (pour une revue, voir Goetze, Cooper, & Brown, 1990) ou chez les adultes (Watts & Hall, 2008). La traduction du timbre entre le modèle et la propre voix du participant constituerait donc une étape importante. Il est alors tentant d'imaginer que certains chanteurs occasionnels éprouveraient des difficultés à associer le son de la cible et leur propre performance vocale, ce qui expliquerait des difficultés à chanter juste.

### ***Trouble au niveau moteur***

L'activité de chanter nécessite le contrôle des muscles laryngés et une coordination pneumo-phonatoire adéquate. De ce fait, l'hypothèse d'un trouble à ce niveau pourrait aussi expliquer un problème de justesse (Joyner, 1969). Cleall (1979), cité par Goetze, Cooper et Brown, (1990) spécifie que les défauts de justesse auraient pour origine un développement insuffisant de l'étendue vocale. Plus récemment, Pfordresher et al. (2007) opposent un argument à cette hypothèse. En effet, ils ont remarqué que les chanteurs ayant des difficultés à chanter juste ont une étendue vocale identique aux chanteurs occasionnels n'ayant pas de difficulté. Leurs possibilités vocales seraient donc similaires et leur permettraient de produire des sons suffisamment aigus ou graves pour chanter la mélodie cible. Cela n'écarte cependant pas l'hypothèse d'un déficit moteur. En effet, comme le relèvent Hutchins et Peretz (2011), une bonne étendue vocale ne permet pas de confirmer que le contrôle des muscles laryngés est efficient. Par ailleurs,

les données fournies par Pfordresher et al. (2010) montrent la variabilité des erreurs des chanteurs dits “imprécis” et soutiennent donc l’hypothèse d’un trouble du contrôle moteur. En 2012, Hutchins et Peretz ont cherché à observer la présence d’un trouble vocal d’origine motrice en comparant la justesse de performances exécutées à la voix avec celles d’un instrument appelé “slider”. Cet instrument permet de reproduire une note sans l’utilisation de la voix à proprement parler mais en pressant un doigt sur un support ressemblant à une règle non millimétrée de 50 cm. Notons que, comme pour une imitation vocale, le participant reçoit un feed-back et peut ainsi ajuster avec précision la hauteur de ses productions. Alors que seule la modalité de réponse est modifiée (voix versus “slider”), les auteurs examinent ainsi l’aspect moteur du chant. Leurs résultats montrent qu’un cinquième du groupe de chanteurs occasionnels est significativement plus performant quand il utilise le “slider” que lorsqu’il vocalise, confirmant ainsi la possibilité d’un trouble de justesse lié au contrôle moteur.

## **Variabilité de la justesse vocale**

Les troubles décrits, qu’ils soient relatifs à la perception, à la mémoire, à l’intégration sensorimotrice ou à la production, vont affecter la justesse d’une performance vocale. Sur la base de ces différentes hypothèses, il serait alors possible de cibler les difficultés des chanteurs occasionnels ne parvenant pas à chanter juste et de leur proposer des pistes rééducatives (Weill-Chounlamountry, Soyez-Gayout, Tessier, & Pradat-Diehl, 2008). Cependant, la catégorisation en “bon” ou “mauvais” chanteur doit être faite avec précaution. En effet, selon le matériel utilisé, la méthode d’évaluation appliquée, les sujets observés, ou encore le contexte de l’évaluation, la qualité des performances vocales varie.

### ***Variabilité liée au matériel musical***

Plusieurs tâches peuvent être proposées pour juger de la capacité d’un chanteur occasionnel à chanter juste (*cf.* Chapitre II). Si les matériaux musicaux sur lesquels porte l’évaluation de la justesse n’ont jamais été clairement comparés, une série d’études souligne que la qualité des performances varie selon le type de tâches proposées, le contexte de la performance ou encore le modèle à imiter.

En ce qui concerne le type de tâche, des données récoltées auprès d'enfants de 2 à 3 ans indiquent que l'imitation est meilleure lorsque le modèle est la mélodie entière plutôt que des phrases individuelles de la même chanson (Tafari, 2009). Chez l'adulte, la qualité d'une performance va également varier selon la tâche proposée, avec une plus grande facilité dans la production de mélodies, par rapport aux tâches d'imitation de sons isolés (Pfordresher & Brown, 2007 ; Pfordresher & Mantell, 2009). La capacité à répéter des sons isolés (Amir et al., 2003 ; Mürbe, Pabst, Hofmann, & Sundberg, 2002 ; Ternström, Sundberg, & Colldén, 1988) semble également inférieure à celle de retenir la hauteur ou le tempo d'une mélodie populaire chez les chanteurs occasionnels (Bergeson & Trehub, 2002 ; Halpern, 1988 ; Levitin, 1994 ; Levitin & Cook, 1996). Par ailleurs, Lévêque et al. (2009) ainsi que Dalla Bella et Berkowska (2009) remarquent que des chanteurs occasionnels présentant des difficultés au niveau de la précision des intervalles lors d'une performance mélodique (chant spontané) ne sont pas pour autant déficitaires en ce qui concerne la répétition.

Au sein d'une même tâche, les consignes vont également influencer la qualité de la performance. Par exemple, Dalla Bella et al. (2007, 2009) observent qu'en ralentissant le tempo de la production à l'aide d'un métronome, la justesse des participants est meilleure. De plus, chanter une mélodie sans les paroles mais en articulant la syllabe neutre /la/ est facilitateur chez l'enfant (par exemple, Levinowitz, 1989). Chez l'adulte, les résultats sont controversés. En effet, Berkowska et Dalla Bella (2009) observent une meilleure justesse lorsque la mélodie est chantée sans les paroles tandis que Racette et al. (2006) ou Mantell et Pfordresher (2013) rapportent le contraire.

Outre le type de tâche et les consignes, le contexte de la performance a un effet sur la qualité des performances vocales. En effet, Goetze et Horii (1989) remarquent que les enfants chantent mieux lorsqu'ils se produisent individuellement par rapport à une performance de groupe. Les recommandations en éducation musicale vont d'ailleurs dans ce sens, c'est-à-dire encouragent les pratiques en solo ou petit groupe pour améliorer la précision des performances (Rutkowski, 1996 ; Rutkowski & Snell Miller, 2003). Notons également que Petzold (1969) a trouvé que des enfants chantent mieux sans accompagnement plutôt qu'avec. Chez l'adulte, Wise et Sloboda (2008)

observent le résultat inverse, avec une nette progression des performances lorsque les participants bénéficient d'un accompagnement instrumental. Chanter en même temps que le modèle semble aider les chanteurs éprouvant des difficultés (Tremblay-Champoux et al., 2010). Cependant, Hutchins, Zarate, Zatorre et Peretz (2010) n'observent pas de différence entre des performances réalisées avec ou sans modèle et Pfordresher et Brown (2007) précisent que l'effet bénéfique ou pas de la présence d'un modèle varie selon l'individu.

Que ce soit chez l'enfant ou l'adulte, le type de modèle à imiter va également affecter la qualité de la performance. En effet, le jeune chanteur occasionnel imite mieux une voix d'enfant qu'une voix d'adulte, surtout s'il s'agit d'un modèle masculin (Green, 1989 ; Sims, Moore, & Kuhn, 1982 ; Small & McCachern, 1983). De plus, la présence de vibrato dans le modèle dégrade la performance de l'enfant (Yarbrough, Bowers, & Benson, 1992). Récemment, des résultats similaires ont été trouvés chez l'adulte. En effet, des chanteurs occasionnels sont capables d'imiter (avec une déviation inférieure à un demi-ton de la cible) des sons de voix synthétiques ou naturelles (Pfordresher & Brown, 2007 ; Wise & Sloboda, 2008). Par ailleurs, il leur est plus facile d'imiter leur propre voix par rapport à des sons complexes non vocaux (Moore, Estis et al., 2008). Lévêque et al. (2012) ont également observé l'avantage d'un modèle humain sur un son synthétique. Ils expliquent cet avantage par l'activation des représentations motrices qu'induit la perception d'un modèle vocal humain et qui faciliterait l'imitation du geste vocal adéquat. Les résultats de Hutchins et Peretz (2012) vont dans le même sens, avec une meilleure précision lorsque le modèle est la voix du participant par rapport à une voix synthétique. Notons également que les bénéfices des modèles vocaux humains sont d'autant plus importants lorsque les sons sont présentés en direct et non à l'aide d'un enregistrement (Granot et al., in press).

Etant donnée l'influence, sur les performances de chanteurs occasionnels, du type de tâche proposé, du contexte de la performance et du modèle à imiter, le choix du matériel musical prend toute son importance. En contexte écologique, nous répétons rarement des sons isolés mais nous chantons généralement des mélodies, l'évaluation de

performances vocales en contexte mélodique semble donc à privilégier. Par ailleurs, la comparaison des résultats obtenus à différentes tâches permettrait d'observer la pertinence des différents matériels musicaux et leurs relations pour évaluer la justesse vocale.

### ***Variabilité liée à la méthode d'évaluation***

Les méthodes d'évaluation de la justesse vocale (par des juges versus à l'aide d'analyses acoustiques), décrites au chapitre II, n'ont jamais été directement comparées. Cependant, quelques études ont permis de constater que l'évaluation par le chanteur occasionnel lui-même ne correspond pas aux résultats obtenus par une méthode objective. En effet, l'utilisation d'analyses acoustiques met en évidence que la majorité de la population est capable de chanter juste mais lorsque les individus estiment leur capacité à chanter juste, les résultats sont différents. Par exemple, Pfordresher et Brown (2007) ont observé que 60 % des 1000 étudiants interrogés rapportaient qu'ils étaient incapables d'imiter une simple mélodie alors que seulement 15 % montraient des difficultés à reproduire correctement des séquences mélodiques. De manière générale, les gens semblent éprouver des difficultés à s'auto-évaluer (Cuddy, Balkwill, Peretz, & Holden, 2005 ; Sloboda, Wise, & Peretz, 2005 ; Wise & Sloboda, 2008 ; Wise, Sloboda, & Peretz, 2007). En ce qui concerne les capacités musicales générales, Cuddy et al. (2005) observent que la proportion d'étudiants pensant qu'ils ont un trouble relatif à la musique atteint 17 % des étudiants interrogés alors que la passation de tests réduit considérablement cette proportion.

Par ailleurs, les méthodes objectives utilisées par les différents auteurs pour évaluer la justesse de performances vocales n'ont pas été comparées. Etant donnée la variabilité liée au choix des critères et des seuils de justesse (*cf.* Chapitre II), il est fort probable que la proportion de mauvais chanteurs va aussi dépendre des critères utilisés. L'étude de Pfordresher et Brown (2009) soutient cette hypothèse. En effet, ils rapportent que lorsque la précision vocale (capacité à répéter une même note plusieurs fois) est prise en compte, la proportion de chanteurs occasionnels manifestant des difficultés de justesse augmente par rapport à l'estimation réalisée sur la base d'autres mesures

objectives (comparaison d'une note avec la cible ou mesure de la taille des intervalles).

Lorsque l'évaluation est effectuée par des juges, le facteur humain entre en jeu et va influencer l'évaluation (*cf.* Chapitre II). Néanmoins, lorsque l'évaluation est effectuée à l'aide de mesures objectives, le choix des tâches, des critères et du seuil de justesse aura également une influence sur les résultats. En d'autres termes, les méthodes subjectives et objectives montrent des limites pour l'évaluation de la justesse de performances vocales en contexte mélodique et les conclusions sur la justesse des chanteurs occasionnels sont alors discutables.

### ***Variabilité liée au sujet***

La description de l'évolution de la justesse et du profil général des chanteurs occasionnels adultes montre que les performances vocales respectent globalement les règles musicales décrites au chapitre I. Cependant, selon la culture du chanteur occasionnel, des différences apparaissent. En effet, Pfordresher et Brown (2009) ont observé que des locuteurs dont la langue maternelle est une langue à tons sont plus performants lorsqu'ils imitent de courtes phrases musicales par rapport à des locuteurs américains. Les auteurs estiment que la maîtrise d'une langue à tons permet de développer la perception/production de micro-tonalités et donc la précision des performances vocales lors de tâches d'imitation vocale. La justesse va également être améliorée avec l'entraînement musical (Amir et al., 2003 ; Dalla Bella et al., 2007 ; Murry, 1990 ; Pfordresher, 2012 ; Watts et al., 2003). Les bénéfices d'un entraînement sont multiples et seront développés dans le chapitre consacré aux chanteurs lyriques (*cf.* Chapitre IV).

### ***Variabilité liée au contexte de l'évaluation***

Comme détaillé au chapitre II, l'évaluation de performances vocales nécessite soit la présence de juges, soit l'utilisation d'un matériel d'enregistrement. De ce fait, les conditions expérimentales sont peu écologiques et le caractère "naturel" du chant est peu respecté. Alors qu'une situation d'examen ou de performance en public est rapportée comme stressante pour des musiciens (Kenny, 2011), il est aisé



d'imaginer qu'il en est de même pour des chanteurs occasionnels. Une situation stressante induit une augmentation des manifestations physiologiques comme la fréquence cardiaque (Abel & Larkin, 1990 ; Brotons, 1994 ; Craske & Craig, 1984 ; Fredrikson & Gunnarsson, 1992 ; Hamann & Sobaje, 1983 ; LeBlanc, Chang Jin, Obert, & Siivola, 1997 ; Yoshie, Kudo, Ohtsuki, 2008 ; Yoshie, Shigemasu, Kudo, & Ohtsuki, 2008). Les conséquences du stress sur une performance musicale ont été examinées depuis les travaux de Yerkes et Dodson (1908). Yoshie, Kudo, Murakoshi et Ohtsuki (2009) montrent, par exemple, qu'une situation stressante altère les performances de pianistes. Cependant, leurs résultats ne corroborent pas les précédentes études qui montrent quant à elles que la performance n'est pas altérée par le stress (Craske & Craig, 1984 ; Yoshie et al., 2008) voire même que celui-ci a un effet positif sur la performance (Hamann & Sobaje, 1983). Les conclusions de ces études sont discutables et les différences observées pourraient être causées par un manque d'analyse du niveau de stress des participants. Dans le contexte d'une performance vocale, nous savons qu'un état anxieux influence le contrôle de la fréquence fondamentale de la voix parlée (Hagenaars & van Minnen, 2005 ; Ruiz, Absil, Harmegnies, Legros, & Poch, 1996 ; Scherer, 1977). La relation entre la fréquence cardiaque et le contrôle de la fréquence fondamentale a été confirmée pour la voix parlée (Giddens, Barron, Byrd-Craven, Clark, & Winter, 2013) et chantée (Bermudez de Alvear, Baron-Lopez, Alguacil, & Dawid-Milner, in press). Cependant, à ce jour, l'influence positive ou négative du stress sur la justesse vocale n'a pas encore été étudiée et la variabilité liée au contexte de l'évaluation reste à explorer.

## Synthèse

Chanter est une activité musicale accessible (Casals et al., 2010), omniprésente (Welch, 2005) et particulièrement appréciée (Chong, 2010). Les conclusions des études relatives au développement de la justesse et aux performances de chanteurs occasionnels adultes témoignent de la capacité à chanter juste pour une majorité de la population (Dalla Bella & Berkowska, 2009 ; Dalla Bella et al., 2007 ; Pfordresher et al., 2010). Pour les 10 à 20 % de la population présentant des difficultés, des hypothèses ont été avancées et de multiples études

ont mis en évidence qu'un trouble pourrait être lié à un déficit perceptif, mémoriel, sensorimoteur ou encore moteur. Cependant, ces recherches se basent sur une catégorisation de la population en "bon" ou "mauvais" chanteur. Au vu du chapitre dédié aux méthodes d'évaluation de la justesse vocale, cette catégorisation doit être faite avec précaution. De plus, selon le type de tâche proposé ou encore le contexte de l'évaluation, la qualité des performances vocales varie. Ces différents éléments devront donc être pris en compte pour une évaluation pertinente de la justesse au sein de la population générale ainsi que pour la compréhension des troubles de la justesse vocale chez des chanteurs occasionnels.

Alors que les chanteurs occasionnels constituent la population principale concernée par l'évaluation, l'activité de chanter peut également être développée de manière professionnelle. Le chapitre suivant concerne les chanteurs lyriques dont la voix est travaillée à la manière d'un instrument de musique.

# CHAPITRE IV

## La justesse des chanteurs lyriques

---

*Evaluer la justesse... de qui ?*

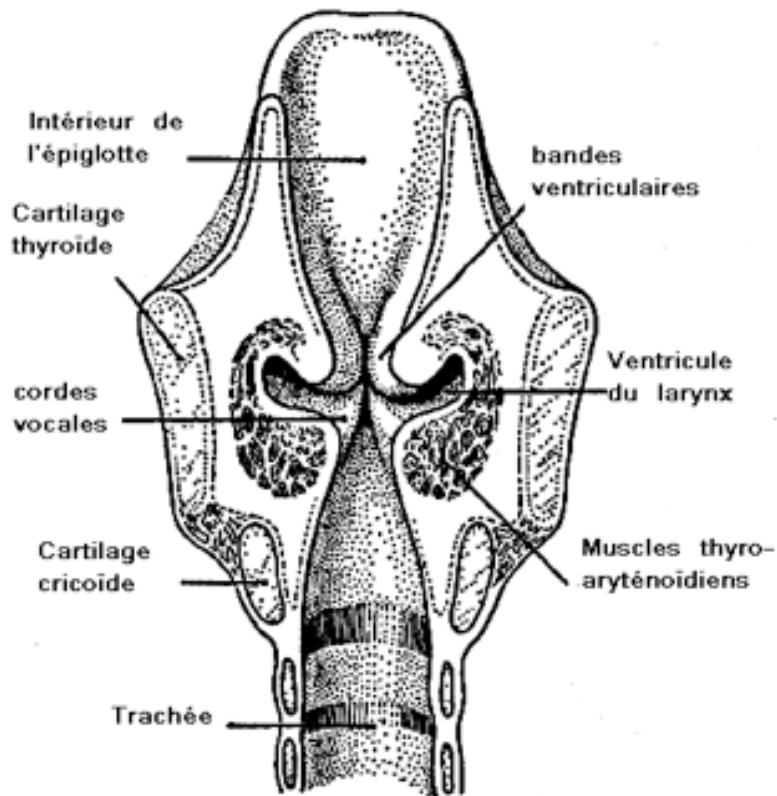
Le fait d'avoir dix doigts ne fait pas de nous des virtuoses du piano pour autant. Il en est de même pour l'instrument vocal. S'il est vrai que tout le monde possède un instrument vocal et l'utilise régulièrement pour chanter (Chong, 2010 ; Welch, 2005), nous ne sommes pas tous à même de nous produire sur scène. Par ailleurs, la voix peut être développée à la manière d'un instrument de musique et travaillée à des fins artistiques, comme le font les chanteurs lyriques. Ce chapitre décrit comment l'instrument vocal peut être développé et les conséquences acoustiques que l'entraînement vocal engendre. Par ailleurs, la justesse est une qualité essentielle, d'autant plus que les performances vocales sont régulièrement accompagnées d'un orchestre, d'un ensemble ou de choristes. Paradoxalement, la justesse n'est pas toujours observée chez des chanteurs professionnels (Sundberg, Prame, & Iwarsson, 1996 ; Vurma & Ross, 2006). Dans une dernière partie, nous décrivons en quoi les paramètres acoustiques développés par l'entraînement vocal peuvent influencer la perception mélodique et donc l'évaluation de la justesse des voix lyriques.

### **L'instrument vocal ou la physiologie de la voix chantée**

De manière schématique et anecdotique, Sundberg (1987) explique que pour parler ou chanter, il suffit de bouger les lèvres, la langue, la mâchoire et le larynx alors qu'un flux d'air passe, à partir des poumons, par les plis vocaux. L'instrument vocal est un instrument à vent classiquement divisé en trois niveaux : respiratoire, phonatoire et

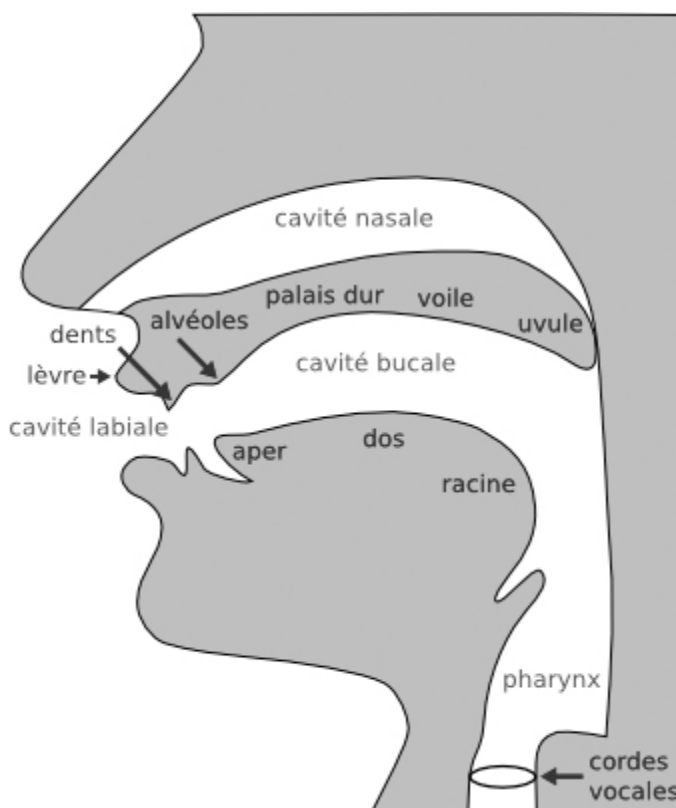
articulatoire/résonantiel. Le fonctionnement de l'instrument vocal a été largement décrit, dans le cadre de la production vocale parlée ou chantée (Fant, 1960 ; Henrich, 2012 ; Sundberg, 1999, 2013 ; Titze, 1994, 2000 ; Welch & Sundberg, 2002).

Le système respiratoire fournit une énergie aérodynamique, constituée par la pression et le débit d'air sous-glottique (à l'intérieur de la trachée, sous le larynx). Lors de son passage au niveau laryngé (Figure 7), l'énergie aérodynamique se transforme en énergie mécanique et acoustique et détermine ainsi la fréquence fondamentale. En effet, la force élastique des plis vocaux, couplée à un effet aérodynamique de Bernouilli, ou effet de rétro-aspiration, définit le cycle d'ouverture et de fermeture glottique entraînant une modulation périodique du débit d'air glottique. Notons que l'élévation de la pression sous-glottique et du débit d'air glottique entraîne une augmentation à la fois du niveau de pression sonore et de la hauteur tonale. Celle-ci est également dépendante de la tension thyro-aryténoïdienne (en se contractant, le muscle vocal augmente la raideur du système et donc la fréquence de vibration des plis vocaux) et de la bascule crico-thyroïdienne (la contraction des muscles crico-thyroïdiens entraîne une bascule du cartilage thyroïde par rapport au cartilage cricoïde et donc un allongement des plis vocaux, un raidissement et une réduction de la masse vibrante).



**Figure 7.** Illustration de la structure laryngée.

L'onde sonore passe ensuite par le conduit vocal. Comme pour la parole, le chanteur a la possibilité de modifier la forme de ce conduit, constitué du pharynx, des cavités buccale, nasale et labiale (Figure 8). Le rôle joué par les positions linguales et vélaires (voile et uvule sur la Figure 8) ainsi que le degré d'aperture buccale est primordial dans la production des différentes voyelles. Sur le plan articulatoire, les composantes fréquentielles des ondes générées dans le conduit vocal sont donc amplifiées ou atténuées selon l'espace des cavités de résonance et créent un timbre particulier (voir le schéma du son complexe, Figure 1, page 18).



**Figure 8.** Illustration des principaux résonateurs vocaux.

Un équilibre entre les différentes forces physiques (aérodynamique, biomécanique et acoustique) est nécessaire pour chanter. Ces paramètres auront un impact sur la hauteur, le niveau de pression sonore et le timbre du son. Alors que le niveau de pression sonore et la fréquence fondamentale du son sont corrélés, une élévation du niveau de pression sonore entraînant une élévation de la hauteur tonale (Titze, 1989 ; Titze & Sundberg, 1992), le chanteur – occasionnel ou professionnel – est capable de maintenir la fréquence fondamentale du son alors qu’il modifie le niveau de pression sonore et inversement. De même, le timbre est lié au niveau de pression sonore et à la fréquence fondamentale mais peut être modifié en adaptant la géométrie du conduit vocal par un contrôle précis des articulateurs (Bournes & Garnier, 2012 ; Garnier, Henrich, Smith, & Wolfe, 2010 ; Henrich, Kiek, Smith, & Wolfe, 2007 ; Henrich, Smith, & Wolfe, 2011).

Jouant avec les possibilités et les contraintes physiologiques de son instrument, le chanteur module donc la hauteur de sa voix, son niveau de pression sonore et son timbre.

## **Du chanteur occasionnel au chanteur entraîné**

Les chanteurs lyriques ont le même instrument vocal que les chanteurs occasionnels. Cependant, outre les bénéfices de l'expertise présentés dans le chapitre I, l'entraînement musical permet de développer une expertise spécifique à l'instrument pratiqué (Zatorre et al., 2007). Dans le cas des chanteurs, une pratique quotidienne aura un effet au niveau de la perception (Kraus & Chandrasekaran, 2010) et de la production (Amir et al., 2003 ; Dalla Bella et al., 2007 ; Murry, 1990 ; Watts et al., 2003). Les chanteurs acquièrent également des connaissances qui leur permettent d'améliorer leurs performances, comme par exemple l'adoption d'une posture appropriée (Wilsonarboleda & Frederick, 2008) ou encore la pratique d'un échauffement vocal adéquat (Amir, Amir, & Michaeli, 2005).

Dans le cas des chanteurs lyriques, le répertoire classique nécessite une maîtrise de l'instrument vocal, que ce soit au niveau respiratoire (Collyer, Kenny, & Archer, 2011 ; Leanderson, Sundberg, & Von Euler, 1987 ; Sundberg, Leanderson, & Von Euler, 1989 ; Thorpe, Cala, Chapman, & Davis, 2001), laryngé (Thalén & Sundberg, 2001 ; Zangger Borch, Sundberg, Lindestad, & Thalén, 2003) ou articuloire (Sundberg, 2009). En général, les chanteurs occasionnels utilisent les mêmes mécanismes qu'en voix parlée, ce qui peut avoir pour conséquence une forte pression sous-glottique et des tensions musculaires pour atteindre des notes de haute fréquence (Titze, 2000). Avec un entraînement adapté, les différents niveaux physiologiques de l'instrument vocal sont de mieux en mieux maîtrisés. Etant donné que la qualité du son produit dépend de la coordination de ces différents niveaux, la maîtrise de l'instrument vocal permet une combinaison complexe de paramètres acoustiques qui différencie ces voix entraînées des voix de chanteurs occasionnels. En effet, en comparant des chanteurs occasionnels et entraînés ou en observant l'effet d'un entraînement vocal, des différences apparaissent au niveau du timbre

(Brown, Rothman, & Sapienza, 2000 ; Lundy, Roy, Casiano, Xue, & Evans, 2000 ; Omori, Kacker, Carroll, Riley, & Blaugrund, 1996), du vibrato (Mürbe, Zahnert, Kuhlisch, & Sundberg, 2007), de la tessiture (Awan, 1991 ; Hunter, Svec, & Titze, 2006 ; LeBorgne & Weinrich, 2002 ; Mendes, Rothman, Sapienza, & Brown, 2003 ; Mürbe et al., 1999 ; Roubeau, Castellengo, Bodin, & Ragot, 2004 ; Sulter, Schutte, & Miller, 1995) et du niveau de pression sonore (Ferguson, Kenny, & Cabrera, 2010 ; Mendes et al., 2003). La pratique vocale permet donc un élargissement de l'étendue vocale, donnant la possibilité de chanter sans effort des notes de hauteurs variées, un enrichissement de la bande fréquentielle contenant le formant du chanteur et une homogénéité du vibrato. En effet, l'étude longitudinale de Mürbe et al. (2007) permet d'observer qu'un vibrato rapide deviendra plus lent après quelques années d'entraînement alors qu'un vibrato particulièrement lent deviendra plus rapide en fin de formation. Par ailleurs, Mitchell et Kenny (2008) observent que le niveau de pression sonore des performances est plus élevé chez les étudiantes de 3<sup>ème</sup> cycle qui ont réussi leur formation avec succès par rapport à celles qui ont davantage de difficultés.

Notons que ces paramètres acoustiques diffèrent également selon le style musical que le chanteur choisit (Cleveland, Sundberg, & Stone, 2001 ; de Almeida Bezerra, Cukier-Blaj, Duprat, Camargo, & Granato, 2009 ; Ferguson et al., 2010 ; Gable, 1992 ; Levin & Edgerton 1999 ; Stone, Cleveland, Sundberg, & Prokop, 2002 ; Sundberg, Gramming, & Lovetri, 1993 ; Sundberg, Gu, Huang, & Huang, 2012). Nous nous focalisons dans cette thèse sur la technique lyrique développée dans la musique classique occidentale.

## **Composantes acoustiques des voix lyriques**

Depuis des décennies, les chercheurs observent différents paramètres acoustiques visant à décrire la technique vocale lyrique (pour des revues, voir Sundberg, 1987, 1999, 2013) mais la complexe combinaison de ces paramètres reste à clarifier. Le concept même de "belle voix" n'est pas précis étant donné qu'il dépend de nombreux paramètres. Parmi ces paramètres acoustiques, le vibrato et la balance spectrale (timbre) ont été particulièrement décrits. Par ailleurs, ces



deux paramètres acoustiques semblent particulièrement appréciés et trouvent une réalité perceptive chez les auditeurs. En effet, Garnier, Henrich, Castellengo, Sotiropoulos et Dubois (2007) ont relevé les nombreuses expressions pouvant décrire les qualités d'une voix lyrique. Ils ont observé que les juges sont consensuels quant à l'utilisation des termes vibrato et timbre pour décrire une "bonne voix".

## ***Vibrato***

Seashore (1938) a été l'un des premiers à s'intéresser au vibrato et le définit comme une ondulation quasi périodique de la fréquence fondamentale, principalement caractérisée par sa fréquence et son amplitude. Au niveau esthétique, des études perceptives ont permis d'observer une association entre ce paramètre et la définition d'une "belle voix" (Ekholm, Papagiannis, & Chagnon, 1998 ; Rothman, Rullman, & Arroyo, 1990).

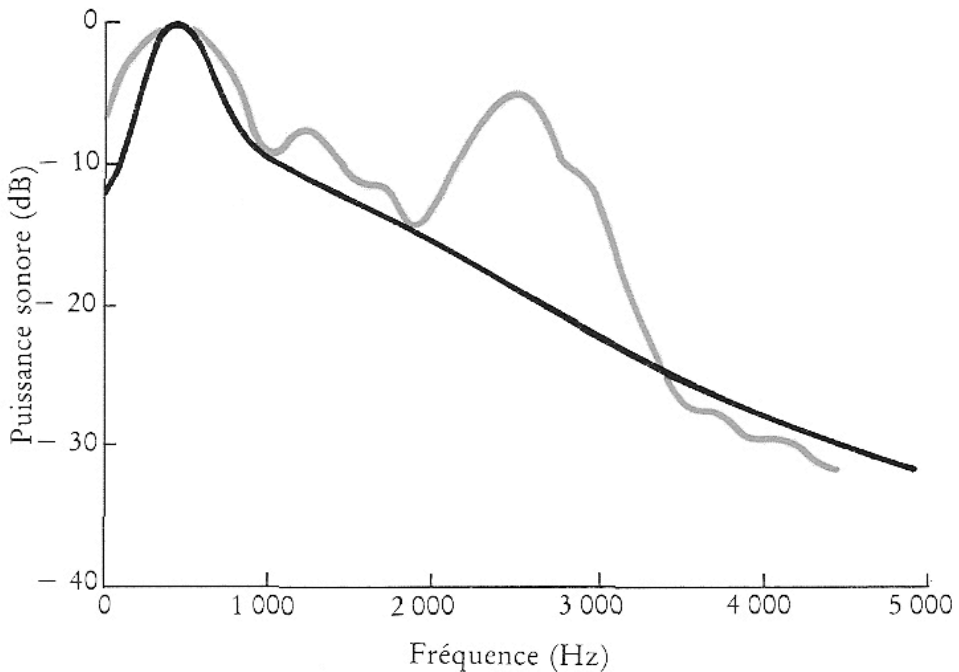
La fréquence du vibrato correspond au nombre de fluctuations par seconde et varie, selon le chanteur, entre 5 et 7 Hz (Prame, 1994). Cet auteur a analysé les performances de 10 chanteurs classiques reconnus et rapporte une fréquence moyenne de 6 Hz, avec une variation d'environ 8 % autour de la moyenne du chanteur. Plus récemment, de Almeida Bezerra et al. (2009) rapportent une fréquence de vibrato chez des chanteurs classiques comprise entre 4.55 et 6.25 Hz. En ce qui concerne les chanteurs lyriques, Howes, Callaghan, Davis, Kenny et Thorpe (2004) notent une fréquence comprise entre 6.28 et 7.14 Hz tandis que Stone et al. (2002) observent une fréquence moyenne de 5.7 Hz. Chez un chanteur lyrique, la fréquence du vibrato est considérée comme constante (Shipp, Leanderson, & Sundberg, 1980) quel que ce soit le contexte dans lequel la performance est réalisée, c'est-à-dire en tant que soliste ou que choriste (Reid, Davis, Oates, & al., 2007). Cependant, la fréquence du vibrato peut être délibérément modifiée (King & Horii, 1993), grâce à un échauffement vocal (Moorcroft & Kenny, 2012) ou, comme précisé précédemment, à l'aide d'un entraînement spécifique (Mürbe et al., 2007).

L'amplitude du vibrato correspond à l'amplitude des fluctuations autour de la fréquence fondamentale moyenne (Hirano, Hibi, & Hagino, 1995 ; Shonle & Horan, 1980 ; Sundberg, 1994). Contrairement à la

fréquence du vibrato, l'amplitude varie considérablement chez un même chanteur (Prame, 1997) et entre les chanteurs (Bretos & Sundberg, 2003). La variabilité de l'amplitude du vibrato, déjà pointée par Seashore (1938), a été confirmée depuis, avec une moyenne de 71 cents selon Prame (1997), une moyenne de 98 cents selon Stone et al. (2002), une variation entre 0.38 et 3.26 demi-tons selon Hirano, Hibi et Hagino (1995) ou encore une variabilité comprise entre 0.54 et 1.66 demi-tons selon de Almeida et al. (2009).

### ***Timbre ou balance spectrale***

Les chanteurs lyriques se produisent généralement dans des salles de concert, chantent sans amplification, parfois accompagnés d'un orchestre, d'un piano, d'un chœur ou d'un ensemble instrumental. Afin d'être entendu par le public dans ces conditions, l'énergie de la bande fréquentielle comprise entre 2 et 4 kHz est particulièrement intense (Barnes, Davis, Oates, & Chapman, 2004 ; Omori et al., 1996 ; Sundberg, 1995, 2001, 2013 ; Thorpe et al., 2001). Cette bande fréquentielle a été identifiée comme primordiale pour une projection vocale efficace des chanteurs solistes (Bloothoof & Plomp, 1988 ; Sundberg, 2001) et contient le "formant du chanteur" (Sundberg, 1995, 2001, 2013). Le renforcement de l'énergie de la bande fréquentielle autour des 3 kHz est d'ailleurs associé avec une bonne qualité vocale étant donné qu'elle permet au chanteur d'être audible malgré la puissance sonore de l'orchestre ou de l'accompagnement (Figure 9).



**Figure 9.** Illustration de la distribution en fréquence de la puissance sonore pour un orchestre (courbe noire) et pour un ténor (courbe grise) (d'après Pierce, 1984, p.126). Alors que la puissance de l'orchestre décroît régulièrement, la voix du ténor présente un pic de puissance entre 2 et 3 kHz, permettant son audibilité malgré la présence d'un orchestre.

Mesurer l'amplitude du formant du chanteur n'est pas aisé car il est influencé par le niveau de pression sonore (Fergusson et al., 2010), la hauteur de la note (Collyer, David, Thorpe, & Callaghan, 2009) et la voyelle produite (Bloothoof & Plomp, 1985 ; Seidner, Schutte, Wendler, & Rauhut, 1983). En 2001, Sundberg propose de mesurer la différence d'amplitude du troisième formant par rapport à celle du premier. Il compare cette différence à ce qu'elle serait dans le cas de la parole, en prédisant l'amplitude qu'aurait le troisième formant à partir des deux premiers. Une autre approche consiste à estimer l'amplitude du formant du chanteur en observant une bande de fréquence déterminée (entre 2 et 4 kHz pour la majorité des études). Il s'agit alors de comparer l'énergie de deux bandes fréquentielles (2-4 kHz / 0-2 kHz) ou de comparer l'énergie d'une bande (2-4 kHz) par rapport à l'énergie totale. Les appellations sont différentes, "singing power ratio" (Omori et al., 1996), "energy ratio" (Thorpe et al., 2001), " $\alpha$  coefficient" (Sundberg &

Nordenberg, 2006) mais le principe reste similaire. A partir du spectre global, également nommé “Long Time Average Spectra”, ces mesures permettent d'estimer la distribution de l'énergie spectrale.

Parmi les chanteurs occasionnels, Watts, Barnes-Burroughs, Estis et Blanton (2006) observent que le spectre fréquentiel est particulièrement riche dans cette bande lorsque les participants sont jugés comme talentueux par rapport à ceux qui sont jugés comme non talentueux. Comme décrit précédemment, des différences apparaissent également à la suite d'un entraînement vocal.

### ***Autres paramètres acoustiques***

D'autres paramètres que le vibrato et la balance spectrale peuvent être observés. En 2009, Butte, Zhang, Song, & Jiang (2009) observent que des mesures relatives à la perturbation du signal acoustique sont pertinentes pour la description de voix lyriques. En effet, les auteurs comparent 26 enregistrements de différents chanteurs et de différents styles musicaux. Ils observent que les voix lyriques sont plus chaotiques, c'est-à-dire de nature apériodiques, que les autres styles (country, jazz, pop, soul, “Musical Theater”<sup>1</sup>), mais que les valeurs restent dans ce que les auteurs considèrent comme la “norme”. Ils ajoutent que malgré la présence d'un vibrato, ces mesures restent interprétables et permettent de définir les caractéristiques d'une technique vocale particulière. Des mesures de stabilité du signal acoustique telles que le jitter (perturbation au niveau de la fréquence) et le shimmer (perturbation au niveau de la pression sonore), couramment utilisées pour l'évaluation de voix parlées pathologiques (Sataloff, 2005), semblent donc pertinentes pour des voix non périodiques (Lieberman, 1961, 1963 ; Zhang, Jiang, & Wallace, 2005) et donc pour des voix lyriques (Butte et al., 2009).

---

<sup>1</sup> Ce terme fait référence aux comédies musicales mais correspond aujourd'hui à une technique vocale utilisée dans le chant de variété, caractérisée par l'utilisation de plusieurs registres vocaux, de vibrato et d'une pression sonore importante.

## La justesse des chanteurs lyriques

Le fait d'être un chanteur adulte entraîné sous-tend que les différentes habiletés (contour, intervalle, tonalité) sont maîtrisées. Par ailleurs, les chanteurs sont des musiciens et, comme exposé dans le chapitre I, ils développent une perception mélodique fine. De plus, l'utilisation d'une technique lyrique nécessite indéniablement un entraînement et donc un développement des capacités physiologiques liées à l'instrument vocal et à sa maîtrise. En conséquence, un chanteur lyrique est censé pouvoir chanter juste.

Cependant, Vurma et Ross (2006) observent que des chanteurs professionnels peuvent chanter des intervalles avec une déviation de 20 à 25 cents par rapport à la cible à reproduire. Dans un contexte mélodique, Sundberg et al. (1996) ont analysé la fréquence fondamentale de 10 enregistrements de l'Ave Maria de Schubert et rapportent des différences entre ce qui est produit et ce qui est noté sur la partition.

Ces résultats étonnants pourraient être expliqués par le style musical adopté par ces chanteurs. En effet, le terme lyrique recouvre, de nos jours, la technique employée par des chanteurs interprétant des œuvres de type "opéra" et plus généralement du répertoire vocal "romantique". Du début du XIX<sup>ème</sup> siècle au milieu du XX<sup>ème</sup> siècle environ, la période romantique est caractérisée par une attention portée à l'expressivité de la ligne mélodique (Samson, 2001). Dès cette époque, de nombreuses modulations apparaissent, la tonalité devient ambiguë avec des tensions harmoniques et des cadences surprenantes. Les règles musicales présentées dans le chapitre I sont toujours utilisées par les compositeurs mais l'écriture s'enrichit par rapport à l'époque classique ou à la musique populaire. De plus, la ligne mélodique se complexifie avec des ornements du type "appoggiature, trille, mordent, arpège" (Hanning, 2011). L'interprétation de ces œuvres complexes va donc permettre de travailler la précision de la justesse mais également de développer un "style" lyrique. En effet, l'interprétation d'œuvres plus simples ou d'un autre style (par exemple populaire) avec une technique lyrique (c'est-à-dire en appliquant des caractéristiques musicales romantiques à d'autres mélodies) engendrera logiquement l'addition d'ornementations afin de "sonner" lyrique et donc des erreurs de

justesse. Outre cette explication “musicale”, les résultats de Sundberg et al. (1996) et Vurma et Ross (2006) peuvent également être expliqués d’un point de vue acoustique. En effet, étant donnée la complexité acoustique du signal de ces voix particulières, des erreurs de justesse ne seraient peut être pas perçues ou, en tout cas, évaluées comme telles.

## **Éléments acoustiques influençant la perception de la justesse vocale**

La hauteur d’un son est principalement déterminée par sa fréquence fondamentale. Cela explique que la relation entre les notes est reconnue, que la mélodie soit jouée par un violoncelle, une flûte, fredonnée ou chantée par un chanteur d’opéra, interprétée à faible ou forte intensité. Cependant, les signaux acoustiques de ces instruments sont différents. Les informations perçues par l’auditeur vont lui donner des indices relatifs à l’instrument et à la qualité des performances mais peuvent également influencer la perception de la hauteur des sons.

En ce qui concerne la voix chantée, le présent chapitre témoigne de la complexité du signal acoustique lorsqu’une technique vocale lyrique est utilisée. Or, les caractéristiques recherchées par des chanteurs lyriques, telles que le vibrato, le timbre ou la pression sonore vont avoir une incidence sur la forme du signal acoustique et donc sur sa perception.

Le vibrato est largement apprécié dans la culture occidentale classique (LeBorgne, Lee, Stemple, & Bush, 2010 ; Sundberg, 1994). Lors des premières études acoustiques, les chercheurs ont conclu qu’il n’influence pas la perception de la hauteur d’une note (Shonle & Horan, 1980 ; Sundberg, 1978). En effet, ces études ont montré que la hauteur perçue correspond à la moyenne de la fréquence fondamentale, quelle que soit l’amplitude du vibrato. Cependant, van Besouw, Brereton et Howard (2008) ont remarqué que la perception de la hauteur d’un son vibré peut être influencée par sa phase, c’est-à-dire le moment du cycle où le vibrato débute et sa durée. Dans une étude sur la tolérance de la justesse, van Besouw et Howard (2008) ont proposé à des musiciens d’écouter des arpèges de trois sons synthétiques dont ils ont modifié la hauteur de la note centrale. Les résultats montrent que la tolérance pour

les sons vibrés est de 34 cents alors qu'elle est de 24 cents pour les sons sans vibrato. Les auteurs en concluent que le vibrato influence la perception de la hauteur lorsqu'il est présent dans une séquence mélodique. Au niveau vocal, la présence d'un vibrato va également augmenter la valeur du seuil de perception d'erreur. En effet, Hutchins et al. (2012) ont proposé une tâche de comparaison de hauteurs de notes chantées à des non musiciens et ont remarqué que le seuil de 50 cents passe à 90 voire 100 cents lorsqu'un vibrato est ajouté à la voix.

Contrairement au vibrato qui affecte la fréquence fondamentale, le timbre est une composante acoustique qui intervient indirectement sur la perception de la hauteur d'un son. Le timbre permet à un auditeur de juger deux sons de même hauteur et de même niveau de pression sonore, comme différents (ANSI, 1994). Le timbre varie principalement selon la richesse du spectre fréquentiel composant un son complexe (voir Howard & Angus, 2009, pour une définition exhaustive). Comme l'illustre parfaitement la gamme de Shepard (1964), les caractéristiques spectrales ont une influence sur la perception de la hauteur. En effet, en manipulant la fréquence des harmoniques et leurs amplitudes respectives – mais en conservant la même fréquence fondamentale –, Shepard a créé une séquence de notes que l'auditeur perçoit comme une suite de notes dont la hauteur s'élève indéfiniment alors qu'il ne s'agit que de la même séquence de notes présentée en boucle (voir également Krumhansl & Iverson, 1992 ; Melara & Marks, 1990a, 1990b, 1990c ; Pitt, 1994).

L'effet du timbre a été confirmé dans des tâches d'estimation de la taille d'intervalles (Russo & Thompson, 2005) mais également dans des tâches d'estimation de la justesse. En effet, les seuils de détection d'erreurs varient selon le timbre des stimuli utilisés. Dans des tâches de comparaison de notes dont une est légèrement déviée, Vurma et al. (2010) ont observé que les notes produites avec un timbre brillant (ténor, trompette), c'est-à-dire riche en harmoniques aiguës, sont perçues 15 à 20 cents plus haut que des sons de même fréquence fondamentale mais avec un timbre plus sombre (violon alto), c'est-à-dire pauvre en harmoniques aiguës. Les sons brillants seraient donc perçus comme plus aigus que les sons définis comme sombres. Ces conclusions sont en accord avec les résultats de Warrier et Zatorre (2002) qui avaient déjà remarqué l'effet du timbre sur la perception,

que ce soit en contexte isolé ou mélodique. Les tâches consistaient à détecter des variations (17, 35 ou 52 cents) entre 2 sons (un son pur et un son complexe) ou entre deux séquences (dont la dernière note, déviée, était congruente ou pas par rapport au timbre du début de la séquence). Dans une expérience similaire, Hutchins et Peretz (2012) se sont intéressés à la voix et ont observé que des non musiciens perçoivent une différence à partir de 50 cents lorsqu'il s'agit de son vocaux alors qu'il suffit d'une variation de 30 cents pour percevoir une différence entre deux stimuli synthétiques. Alors que Moore et al. (2008) ne reportaient pas de différence de discrimination selon le type de stimuli proposé (son complexe non vocal, voix de femme, voix du sujet), l'effet du timbre (son vocal versus son instrumental) a été confirmé lors d'une tâche de comparaison de mélodies, avec une perception à 90-100 cents pour la voix chantée et 50 cents pour les stimuli joués au violon (Hutchins et al., 2012).

Comme pour le timbre, le niveau de pression sonore a une incidence sur la perception de la relation entre les notes, c'est-à-dire sur la taille des intervalles. Thompson, Peter, Olsen et Stevens (2012) ont proposé à des non musiciens d'évaluer la taille d'intervalles glissés (son continu entre les deux notes) sur une échelle de 1 (petit intervalle) à 5 (grand intervalle). Les stimuli étaient soit ascendants soit descendants et quatre conditions étaient proposées : niveau de pression sonore faible ou fort pour les deux sons, niveau de pression sonore croissant ou décroissant entre les deux sons. Les résultats montrent que les intervalles avec un niveau constant de pression sonore "forte" étaient perçus comme plus grands que les intervalles avec un niveau constant de pression sonore "faible". Lorsque le niveau de pression sonore augmente, les intervalles sont perçus comme plus grands que lorsqu'il diminue entre les deux notes. Ce même pattern de résultats est retrouvé lorsque les intervalles ne sont pas glissés mais constitués de deux notes séparées, mais les variations de niveau de pression au sein de l'intervalle n'ont une incidence sur le jugement de leur taille que pour les intervalles ascendants.

Notons que si l'influence des caractéristiques du son sur la perception de la fréquence fondamentale est maintenant bien connue, elle reste peu explorée en contexte mélodique. En effet, les études ont principalement porté sur la discrimination entre deux sons, dans un



contexte d'intervalle et quand il s'agissait de mélodies, la modification était limitée à la dernière note. Par ailleurs, la majorité de ces études porte sur des paramètres acoustiques isolés. Les différentes conclusions apportent des informations concernant la perception de la justesse de voix lyriques mais des études combinant les différents paramètres acoustiques seraient nécessaires afin de mieux comprendre les conclusions de Vurma et Ross (2006) et de Sundberg et al. (1996).

## Synthèse

Chanter juste est une qualité attendue chez les chanteurs professionnels classiques. En effet, ils sont censés respecter l'écriture musicale de ce qu'ils interprètent. Cependant, il a été montré que ce n'est pas toujours le cas (Sundberg et al., 1996 ; Vurma & Ross, 2006). Les conclusions de ces études sont étonnantes étant donné qu'un entraînement musical a des conséquences positives sur la qualité de performances vocales (Amir et al., 2003 ; Dalla Bella et al., 2007 ; Murry, 1990 ; Watts et al., 2003). Par ailleurs, l'appréciation de ces voix est également dépendante de caractéristiques acoustiques telles que le vibrato ou le timbre (Garnier et al., 2007). Or, ces caractéristiques ont une incidence sur la perception de la hauteur des sons (Hutchins et al., 2012 ; Shepard, 1964 ; van Besouw & Howard, 2008) ou de la taille des intervalles mélodiques (Russo & Thompson, 2005 ; Vurma et al., 2010) et donc sur la perception de la justesse vocale. Ce chapitre met donc en évidence les particularités acoustiques des voix lyriques et également que l'évaluation de la justesse de voix lyriques devrait tenir compte de ces particularités.







# **Partie expérimentale**



# OBJECTIFS DES ÉTUDES

---

L'objectif général de ce travail de thèse porte sur l'évaluation de la justesse vocale en contexte mélodique, d'un point de vue méthodologique – outils et critères à utiliser pour évaluer la justesse – et théorique – définition de la justesse vocale en contexte mélodique. En effet, la justesse vocale est un critère primordial pour apprécier les productions mélodiques d'un chanteur occasionnel (Watts et al., 2003) mais son évaluation n'est pour l'instant pas optimale. Le chapitre I propose une définition de la justesse d'un point de vue musical et perceptif. Dans le système tonal occidental, la justesse peut être schématisée à l'aide de trois éléments : le respect des contours mélodiques, de la taille des intervalles et de la tonalité. Ces éléments sont perçus très tôt par l'enfant (Trainor, 2005 ; Trainor, Marie, Gerry, Whiskin, & Unrau, 2012) et sont intégrés à l'âge adulte (Dowling & Fujitani, 1970 ; Edworthy, 1985 ; Stalinski et al., 2008 ; Trainor & Trehub, 1992). Par ailleurs, un entraînement musical permet d'affiner la perception mélodique (Hutchins & Peretz, 2012 ; Hutchins et al., 2012 ; Orsmond & Miller, 1999 ; Russo & Thompson, 2005 ; Vurma et al., 2010 ; Zatare et al., 2012). Il n'est donc pas étonnant que les recherches visant à évaluer la justesse d'une performance vocale emploient des juges experts, ayant suivi un entraînement musical (par exemple, Alcock et al., 2000a, 2000b ; Hébert et al., 2003 ; Kinsella et al., 1988 ; Lévêque et al., 2012 ; Prior et al., 1990 ; Racette et al., 2006 ; Schön et al., 2004 ; Wise & Sloboda, 2008). Cependant, le chapitre II met en évidence la complexité à évaluer la justesse d'une performance musicale. En effet, de nombreux facteurs influencent l'évaluation effectuée par un juge, qu'ils soient liés au juge (Cavitt, 1997 ; Duerksen, 1972 ; Kinney, 2009 ; Morrison et al., 2004) ou au matériel musical à évaluer (Hutchins et al., 2012 ; Russo & Thompson, 2005 ; Vurma & Ross, 2006 ; Warrier & Zatorre, 2002). Dans le cadre de cette thèse, nous qualifions cette méthode de "**subjective**" du fait de la présence de facteurs influençant l'évaluation et du manque

de précision des critères utilisés pour évaluer la justesse d'une performance vocale par un être humain. Afin de pallier la subjectivité des juges, des outils informatiques ont été développés pour évaluer « objectivement » la justesse de performances vocales en contexte mélodique (Dalla Bella et al., 2007, 2009 ; Elmer & Elmer, 2000 ; Howard & Welch, 1989 ; Lévêque et al., 2009 ; Pfordresher et al., 2007, 2010). Cette méthode dite "**objective**" semble maintenant privilégiée (Dalla Bella et al., 2011) mais montre également des limites. En effet, différents auteurs rapportent des erreurs d'extraction de la fréquence fondamentale et donc la correction manuelle de ces erreurs (Dalla Bella et al., 2007 ; Granot et al., in press ; Pfordresher et al., 2010). De plus, les critères musicaux sélectionnés permettent une quantification de la précision des intervalles mais, comme exposé au chapitre I, la précision des intervalles ne représente pas, à elle seule, les différentes règles musicales du système tonal occidental ni ce qu'un auditeur est capable de percevoir.

Les méthodes "subjective" et "objective", destinées à évaluer la justesse de chanteurs occasionnels, coexistent mais restent discutables. Afin d'élaborer une méthode objective et pertinente d'évaluation de la justesse vocale en contexte mélodique, il est alors nécessaire d'explorer les aspects méthodologiques et théoriques. Plus spécifiquement, nous avons focalisé notre attention sur trois points : l'évaluation objective de la justesse d'une performance vocale en contexte mélodique, la pertinence d'une telle méthode et la clarification de la notion de justesse.

Dans un premier temps, nous avons élaboré une méthode d'évaluation qui tient compte des règles musicales dictées par la culture occidentale. Une **première étude** présente le fonctionnement de ces outils et leur application à des performances vocales de chanteurs occasionnels et de chanteurs entraînés. Afin de valider notre démarche, nous avons observé l'effet de l'expertise vocale sur la justesse de performances mélodiques. En effet, les enfants chantent, sans entraînement spécifique, de plus en plus juste en grandissant (Apfelstadt, 1984 ; Davidson, 1994 ; Dowling, 1999 ; Flowers & Dunne-Sousa, 1990 ; Geringer, 1983 ; Greenberg, 1979 ; Hargreaves, 1996 ; Moog, 1976 ; Rutkowski, 1990 ; Rutkowski, 1997 ; Rutkowski & Miller, 2003 ; Welch, 1979, 1986, 1991, 1994, 1998, 2002, 2006). Cependant,



cette qualité est améliorée à l'aide d'un entraînement musical ou vocal (Amir et al., 2003 ; Dalla Bella et al., 2007 ; Murry, 1990 ; Watts et al., 2003) qui permettra au chanteur professionnel de maîtriser les différentes règles musicales dictées par sa culture. Par ailleurs, si les outils actuels ne permettent pas de préciser les règles musicales non respectées par le chanteur occasionnel, Dalla Bella et al. (2007, 2009) ont néanmoins décrit un profil général chez les chanteurs occasionnels. Ils observent que leurs performances sont meilleures lorsque le tempo de la production est ralenti. Nous avons donc vérifié ces effets (expertise et relation tempo-justesse) afin de nous assurer de la pertinence globale de nos outils. En outre, nous avons souhaité observer sur quels critères musicaux se base l'évaluation de juges experts. En effet, si des auditeurs sont capables de percevoir différents types d'erreurs musicales (*cf.* Chapitre I), l'évaluation globale d'une performance ne permet pas de déduire les critères sur lesquels se basent les jugements (*cf.* Chapitre II). Dans une **deuxième étude**, nous avons donc cherché à observer les critères utilisés par des juges afin d'identifier les règles musicales à prendre en considération pour une évaluation objective de la justesse vocale en contexte mélodique. Dans ce but, nous avons comparé les résultats obtenus à l'aide de notre méthode - en quantifiant les scores des trois règles musicales (respect du contour mélodique, de la taille des intervalles et de la tonalité) - à l'évaluation globale effectuée par des juges experts. Autrement dit, en nous basant sur le schéma de l'évaluation objective (bas de la Figure 6 - page 39), nous avons ajouté la variable "juge" et observé leur comportement à l'aide des deux autres variables alors connues (performance et critères). Etant donné que le type d'expertise a des répercussions sur la perception mélodique (*cf.* Chapitre I), nous distinguons les "experts en voix", c'est-à-dire les logopèdes spécialisés dans la rééducation vocale (connaissances explicites sur la voix mais pas de formation musicale), et les "experts en musique". Ceux-ci sont soit des étudiants de Conservatoire, soit des "professionnels", c'est-à-dire des participants ayant suivi une formation de troisième cycle et travaillant dans le domaine musical, en tant que concertiste ou professeur et se produisant régulièrement en public. Les "experts en musique" ont des connaissances musicales mais pas de connaissances explicites dans le domaine de la voix. Pour le recrutement de nos sujets, nous avons tenu compte de leur niveau, c'est-à-dire s'ils

sont étudiants ou professionnels. Chez les professionnels, nous différencions également deux catégories : les instrumentistes et les chanteurs qui eux, ont une double formation permettant des connaissances musicales et vocales explicites.

Ces deux premières études ont donc pour but de valider une démarche objective et pertinente pour évaluer la justesse vocale de chanteurs occasionnels en contexte mélodique.

Dans un second temps, nous avons appliqué cette méthode d'évaluation à des situations expérimentales. En effet, afin d'estimer le développement de la justesse, la prévalence d'un trouble ou encore les causes d'une difficulté à chanter juste, l'évaluation des performances vocales est nécessaire. Différentes tâches peuvent être proposées pour quantifier la justesse de performances vocales (répétition de sons isolés, d'intervalles, de séquences de notes ou production de mélodies complètes) mais la qualité des performances est influencée par le type de tâche proposée au participant (*cf.* Chapitre III). La pertinence même d'une tâche de production d'une mélodie complète reste donc à valider. Dans une **troisième étude**, nous avons comparé les résultats d'une performance en contexte mélodique (analysée à l'aide de nos outils) à ceux d'une tâche de répétition de sons isolés, tâche qui est régulièrement proposée par les laboratoires de recherche en cognition musicale. Par ailleurs, le chapitre III décrit différentes hypothèses explicatives d'un trouble de justesse. Dans ce chapitre, nous avons discuté de l'effet du contexte de l'évaluation et de l'éventuel effet du stress sur la justesse vocale. En effet, une performance vocale en public engendre un état de stress (Kenny, 2011) qui, à son tour, influe sur la qualité de la performance (Yoshie et al., 2009). Nous avons donc souhaité observer, à l'aide de notre méthode d'évaluation, l'effet du stress sur différentes erreurs de justesse dans une **quatrième étude**.

Finalement, l'évaluation objective de performances vocales n'a été que peu explorée pour des voix travaillées, telles que celles des chanteurs lyriques. Or, l'évaluation s'adresse également à des chanteurs qui ont développé leur instrument vocal à la manière d'un instrument de musique (*cf.* Chapitre IV). Deux études ont mis en évidence que les performances vocales de chanteurs professionnels classiques ne sont pas toujours justes lorsqu'elles sont comparées à la partition interprétée

(Sundberg et al., 1996 ; Vurma & Ross, 2006). Deux hypothèses peuvent être formulées pour expliquer les conclusions de Sundberg et al. (1996) et Vurma et Ross (2006). D'une part, les chanteurs lyriques, régulièrement amenés à interpréter des mélodies du répertoire romantiques dont l'ornementation est particulièrement expressive (voir Samson, 2001), appliqueraient ce type d'ornementation à tout support mélodique et dévieraient alors consciemment de la partition musicale dans un but expressif. D'autre part, comme exposé dans le chapitre IV, les voix lyriques se différencient des voix de chanteurs occasionnels par leur complexité acoustique. Etant donnée l'influence des paramètres acoustiques du son sur la perception mélodique d'un juge, la justesse des voix lyriques ne se baserait peut-être pas sur les mêmes critères que pour des voix de chanteurs occasionnels. Afin de tester l'hypothèse relative à l'expressivité de l'interprétation, nous avons souhaité, dans une **cinquième étude**, quantifier la justesse de différentes mélodies (mélodie populaire versus mélodie du répertoire romantique) produites par un large corpus de chanteurs professionnels. Dans le but de tester l'hypothèse relative à la complexité acoustique, nous avons tenté, dans une **sixième étude**, de mieux définir les voix lyriques en observant les effets de la technique vocale et de la mélodie sur les paramètres musicaux et acoustiques développés à la suite d'une formation vocale. Grâce à cette définition acoustique, nous avons souhaité mieux comprendre les critères d'évaluation des juges experts. Dans une **septième étude**, nous avons donc proposé à des juges experts de classer les performances vocales de chanteurs lyriques et comparé ces données avec l'analyse objective de paramètres musicaux et acoustiques définissant ces voix particulières.

La partie expérimentale de cette thèse, constituée de sept études, publiées ou en révision dans des revues internationales, se penche sur les aspects méthodologiques et théoriques de l'évaluation de performances vocales. Elle vise ainsi à apporter des réponses aux trois questions suivantes :

- Comment évaluer objectivement la justesse de performances vocales en contexte mélodique ?
- Quelle est la pertinence d'une telle méthode ?
- Quelles clarifications cette méthode objective apporte-t-elle à la notion de justesse mélodique ?



# ETUDE 1

## Criteria and tools for objectively analysing the vocal accuracy of a popular song

---

**Pauline Larrouy-Maestri & Dominique Morsomme**

<sup>1</sup> Logopédie de la Voix, Department of Psychology: cognition and behaviour, University of Liège, Belgium

In press in Logopedics Phoniatrics Vocology

### **Abstract**

This study aims to validate our method for measuring accuracy in a melodic context. We analysed the popular song “Happy Birthday” sung by 63 occasional and 14 professional singers thanks to AudioSculpt and OpenMusic (Ircam, Paris, France). In terms of evaluation of the pitch interval deviation, we replicated the profile of occasional singers described in the literature (the slower the performance, the more accurate it is). Our results also confirm that the professional singers sing more accurately than occasional singers but not when a Western operatic singing technique is involved. These results support the relevance of our method for analysing vocal accuracy of occasional and professional singers and led us to discuss adaptations to be implemented for analysing the accuracy of operatic voices.

### **Introduction**

The singing voice is an instrument which everyone possesses and can develop (Welch, 2005). Judgement of the quality of sung vocal performance is based on aesthetical criteria related to a specific culture.

A survey of 1009 singing teachers from the “National Association of Teachers of Singing” (NATS) demonstrated that intonation is a criterion of paramount importance, like the tone quality and musicality, in defining vocal quality (Watts, Barnes-Burroughs, Andrianopoulos, & Carr, 2003). The definition of accuracy varies depending on the author but its measurement is always based on variations in fundamental frequency (Dalla Bella, Berkowska, & Sowinski, 2011).

This parameter is regularly observed in estimating the prevalence of accuracy problems called “poor-pitch singing” (Welch, 1979), detecting amusia (Henry & McAuley, 2010), categorising singers as “good” or “bad” (Dalla Bella & Berkowska, 2009; Dalla Bella, Giguère, & Peretz, 2007; Pfordresher & Brown, 2007; Pfordresher, Brown, Meier, Belyk, & Liotti, 2010) and investigating the causes of accuracy troubles (for review, see Hutchins & Peretz, 2012).

Different tasks can be proposed to observe the vocal accuracy. Some involve reproducing sounds (isolated or in short tunes): single pitch-matching (Amir, Amir, & Kishon-Rabin, 2003; Watts, Moore, & McCaghren, 2005), imitation of single pitches (Pfordresher & Brown, 2007; Wise & Sloboda, 2008) or intervals (Pfordresher & Brown, 2007), imitation of short pitch patterns (Wise & Sloboda, 2008), of short novel melodies (Pfordresher & Brown, 2007) or unfamiliar pitch sequences (Pfordresher, Brown, Meier, Belyk, & Liotti, 2010). Others consist in imitation (Berkowska & Dalla Bella, 2009; Dalla Bella & Berkowska, 2009) or singing from memory familiar melodies (Berkowska & Dalla Bella, 2009; Dalla Bella & Berkowska, 2009; Dalla Bella et al., 2007; Pfordresher et al., 2010).

In a melodic context, acoustical analysis consists first in the segmentation of the auditory signal and the fundamental frequency ( $F_0$ ) estimation. Praat software (Boersma, 2001; Boersma & Weenink, 2008) is commonly used (e.g. Dalla Bella et al., 2007; Pfordresher et al., 2010) for the visual inspection of the waveform and the identification of the steady-state phase of each sung note. In these studies, the mean  $F_0$  was then estimated and the pitch detection errors (e.g. octave jumps) were manually corrected. Then, the  $F_0$  of each note produced can be compared with the target but also between the notes of the performance itself. Several errors can thus be observed (see Figure 1): errors in the

melodic contour (the produced interval direction deviated from the musical notation, e.g., the second note of the interval is higher instead of lower than the first), pitch interval errors (the direction of the interval is respected but the distance between the notes is altered) and tonal centre errors (change in tonality during a melodic phrase).

**Accurate version**



**Contour error** at note 2 (this note goes up instead of going down)



**Interval error** of 200 cents (one tone) between notes 1-2 and 2-3 (the contour of the melody is respected but the interval is too big)



**Tonal deviation** at note 5 (interval error between notes 4-5 involving a modulation of the following of the tune)



**Figure 1.** Illustration of errors observed in a melodic context.

In order to validate our criteria and tools for analysing vocal accuracy, we wished to reproduce an experiment conducted on a general population but also to verify the effect of vocal expertise on the accuracy of a popular tune.

The accuracy of singing by occasional singers was analysed by Dalla Bella et al. (2007). These authors recruited 62 occasional singers to sing the popular song “Gens du Pays”, a song sung to celebrate birthdays in Quebec. They measured the pitch of the 32 notes in this song. Their results indicate that the majority of the population can sing relatively accurately. They also investigated the impact of tempo on the quality of the singing and concluded that by slowing down the tempo, the accuracy of the singing was closer to that of the professionals. We

aimed to reproduce the observations of Dalla Bella et al. (2007) about the link between tempo and accuracy using our criteria and tools for analysis in order to validate them.

Professional singers have the same vocal organs as occasional singers. However, musical training enables the development of general musical expertise as well as expertise specific to the instrument used (Zatorre, Chen, & Penhune, 2007). In the case of the singers, their daily practice has an effect on the voice, in terms of perception (Kraus & Chandrasekaran, 2010) and production (Amir, Amir, & Kishon-Rabin, 2003; Dalla Bella et al., 2007; Murry, 1990; Watts, Murphy, & Barnes-Burroughs, 2003), with the accuracy of intonation significantly increasing thanks to musical training. The singers also acquire knowledge concerning the suitable conditions required for good singing, such as the adoption of appropriate posture (Wilsonarboleda & Frederick, 2008) or the performance of a vocal warm-up (Amir, Amir, & Michaeli, 2005).

By comparing the singing of a population of occasional singers who have not received any musical training with that of professional singers well-versed in singing techniques, we aimed to confirm the effect of vocal expertise on vocal accuracy and thus support the choice of our criteria and tools of analysis for observing it.

In the modern sense of the word, operatic singing designates the technique widely employed in opera and (by extension) more generally in Western classical music (e.g. *lieder*, *cantatas*, *oratorios*, etc.). The acoustic qualities of professional singers are documented by various studies (for reviews, see Sundberg, 1987, 1999), in a soloist or choir context (Reid, Davis, Oates, & al., 2007). The intensity of the singer's formant, the presence of vibrato, the maximum phonational frequency range and the intensity increase with musical experience (e.g. Brown, Rothman, Sapienza, 2000; Hunter, Svec, & Titze, 2006; Mendes, Rothman, Sapienza, & Brown, 2003). Few studies have examined accuracy in professional singers. Vurma and Ross (2006) remarked that when professional singers have to produce intervals, they may be false by 20 to 25 cents but the singing is still considered as accurate by expert judges. In this study, the focus was on producing isolated intervals but vocal accuracy in a melodic context was not investigated. The precision



of intonation remains necessary in a melodic context in order to satisfy the expectations of our Western culture, to respect musical composition and to sing with accompanying instruments or in a vocal ensemble. Analysis of vocal performance with an operatic technique, used by classically trained singers allows us to put to the test the criteria and tools developed for the voices of occasional singers. Furthermore, through vocal training, professional singers are led to alter their singing in accordance with specified instructions. Even if the vocal technique developed by a singer cannot be taken from him or her, nonetheless we can observe different types of singing depending on the type of musical genre requested. Comparison between the performances of professional singers using an operatic vocal technique or not will enable highlighting of any possible acoustic characteristics to be taken into account for the study of accuracy when an operatic technique is used.

## Methods

### *Participants*

We analysed the singing of a control group made up of 63 female occasional singers (OS) and an experimental group of 14 female operatic singers.

The voices of the OS come from the database available on: <http://sldr.org/sldr000774/en>. These 63 women were aged from 15 to 75 years old ( $M = 29.83$ ,  $SE = 1.89$ ). These subjects were chosen according to musical training criteria: they had received no formal musical training, had never played an instrument and had never had extra-curricular school lessons in music. For this study, only the performances sang with the correct melodic contour and the correct number of notes (i.e. 21 notes) were selected.

The experienced singers were recruited from the Royal Northern College of Music in Manchester. These 14 women were aged from 19 to 54 years old ( $M = 24.29$ ,  $SE = 2.35$ ). They started training in singing from the age of 8 to 17 years old ( $M = 13$ ,  $SE = 0.85$ ). The number of years of practice is situated between 5 and 20 years ( $M = 9.57$ ,  $SE = 1.01$ ).

## ***Material***

The OS spontaneously sang the French version of the popular tune “Happy Birthday”, which has the same melody as the English version. No starting tone was given and participants could choose a comfortable tonality. They were asked to produce two glissandi (a continuous glide from a low note to a high note and vice versa). The aim of these glissandi was multiple: to warm up the vocal organs (Amir et al., 2005), to help the non-singer to sing the tune in their voice range, to verify the vocal capacity of the subjects which could explain missing notes in the auditory signal and to encourage a lack of inhibition in front of the experimenter and the recording equipment. The sound recordings were made using a head-worn microphone (Sennheiser HS2) positioned at a constant distance of 2 cm from the right corner of the mouth and a digital recorder (Marantz Professional Solid State Recorder PMD67).

The experienced singers sang “Happy Birthday” after two glissandi with the instructions to sing “naturally, whilst imagining a festive and friendly context”. This form of singing is referred to as “Without a specific vocal Technique” (WT). Thereafter, they sang a phrase from a song in their repertoire in order to warm-up for an operatic performance. To finish, we asked them to sing “Happy Birthday” in an operatic way “whilst imagining the context of a scene from an opera”. This singing is referred to as “with Lyrical Technique” (LT). This data was collected using a digital recorder (Zoom nh4) held at a distance of one metre away from each participant. The unidirectional microphone was placed directly in front of the mouth of the participant.

Since our study focuses on the interval accuracy and the respect of the tonal centre, we ensured that the tune was the same for all the subjects, whether it concerned the number of notes or the respect of the melodic contour.

## ***Analytical procedure***

The equal temperament is a compromise tuning scheme which has been developed for keyboard instruments. The equal tempered system uses a constant frequency multiple between the notes of the chromatic scale. This system seems to not suit a capella choirs (Howard, 2007a,

2007b) but stays a reference in Western music. The acoustic analyses of individual performances of the popular song “Happy birthday” were based on this temperament.

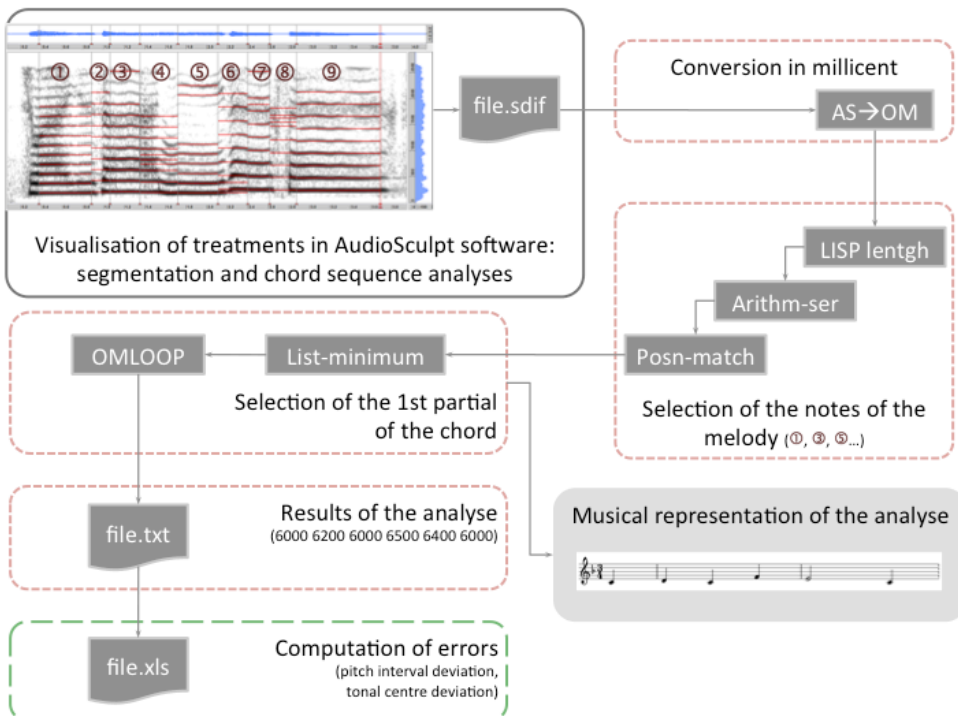
The popular song “Happy Birthday” is made up of four phrases, with each syllable corresponding to a note. It totals 25 notes corresponding to the 25 syllables of the song and 21 when the repeated notes are not taken into account. The vowels provide the maximum of voicing and stable pitch information (Murayama, Kashiwagi, Kashiwagi, & Mimura, 2004), mark the onsets of musical tones (Sundberg & Bauer-Huppmann, 2007) and constitute the best targets for analysis. Thanks to AudioSculpt and OpenMusic (Ircam, Paris, France), we were easily able to analyse the vowels sung of the 21 notes in two steps, using a MacBook Pro (Mac OS X, Version 10.6.5).

The first step involved the use of the AudioSculpt 2.9.4v3 software (designed and developed by the Ircam Analysis/Synthesis Team). This program enabled various processing and analysis of audio files (.wav). Note that analysis and treatments performed through this software use a Short Time Fourier Transform (STFT) analysis as an initial step. Therefore, the main parameters of the STFT analysis (window size, window type, FFT size and window step) can be modified. For an easy visualisation of the auditory signal a sonogram analysis was performed. The window size is the principal factor determining the analysis properties of spectral or partial analysis. It allows to precise the time and frequency resolution of the analysis that is performed. We have chosen a Window Size of 2048 samples (it represents 46.42 ms) and a Window Step of 8. This combination, recommended by experts in computer-assisted music and validated by specialists in music analysis, allows precise analysis of the singing voice.

The AudioSculpt software enables to not take into account the unharmonic frequencies (subharmonics and noises were selected and rejected) and therefore automatic and correct recognition of the partials by the software. Following this step, markers were manually placed on the spectrogram. Attacks were avoided and the stable parts of each vowel-group were isolated avoiding the slides toward the pitch at the beginning of a note. This step was possible thanks to the precise visualisation of the acoustic signal. To help the experimenter, a zoom

(focus on a part of the visual representation) or a time stretching (slowing the tempo without changing the spectral information of the sound) can be used. Finally, a Chord Sequence Analysis was performed with similar parameters and allowed to obtain a chord of all the partials analysed between each marker. The aim was to gain averages for each partial between two markers.

The second step was conducted using the Open Music 6.3 software (designed and developed by the Ircam Musical Representations Team). This programming tool allowed analysis and processing of musical structures. The program made it possible to skip the segments between two notes (i.e. to consider only the 21 notes of the melody), to select the lowest partial to be analysed and to transcribe into cents ( $1/2$  tone = 100 cents). This process allowed obtaining the averages of the  $F_0$  for the 21 notes of the tune.



**Figure 2.** Illustration of the procedure used for acoustic analyses of a sung performance.

As shown in Figure 2, different objects allowed the final analysis of the acoustic signal. The files (.sdif) of chord analysis were converted into

data (AS→OM). Then, only the 21 notes of the tune were conserved (LISP length, arthm-ser and Posn-match) and the lowest partial was selected (list-minimum and OMLOOP). The final analysis could be listened, visualised on a score or observed through the list of note's pitches (in cents). The file (.txt) was then exported to Excel in order to compute the pitch errors of the sung performance, regarding two different criteria: pitch interval deviation and tonal centre deviation.

## ***Calculations***

The average tempo of the singing was calculated on the basis of the performance's length. Two criteria were chosen to estimate the accuracy of the singing: the pitch interval deviation (for pitch interval errors) and the tonal centre deviation (for tonal centre error). In both cases, the criteria were estimated by measuring the difference between the size of intervals performed and the standard size based on the musical notation (see Figure 3). Small deviation reflects high accuracy in relative pitch. Note that the measurements were not computed in terms of semitones but on a continuous scale (in cents).

*Pitch interval deviation:* This measurement has often been adopted as a measure of singing proficiency (e.g. Berkowska & Dalla Bella, 2009; Dalla Bella & Berkowska, 2009; Dalla Bella, Giguère, & Peretz, 2007; Pfordresher, Brown, Meier, Belyk, & Liotti, 2010). Each interval was calculated by subtracting the pitch of the adjacent notes. Since the absence of contour errors was a selection criterion for our samples, the intervals were able to be calculated in absolute values. Each value was compared to a standard (a theoretical value estimated on the basis of the score) and represented in cents. For example, the interval between the 2<sup>nd</sup> and 3<sup>rd</sup> note of the tune should form a major second, which corresponds to a variance of 200 cents. If the subject made a quarter-tone mistake, i.e. if she sang an interval of 150 or 250 cents, we considered that the error was 50 cents in relation to the expected interval.



**Figure 3.** Score of the tune “Happy Birthday” with the number of notes used for calculating accuracy.

*Tonal centre deviation:* In terms of “tonal centre” criterion, Dalla Bella et al. (2007) observed the “pitch stability” by computing the deviation of two reproductions of a single phrase in a melody. This measurement was not applicable to the song “Happy Birthday” because of its structure (no repetition in the melody). To evaluate this criterion, we proposed here to adapt the “precision” criterion used in Pfordresher et al. (2010). They observed the consistency in repeating targets through imitation tasks of unfamiliar pitch sequences. A popular song is characterized by a strong tonal centre and the presence of important notes of the harmony given by the scale of the tune. In this study we observed the consistency of important notes of the harmony as a measure of the tonal centre deviation. Four notes (the first notes of the three first sentences and the last note, i.e. notes 1, 6, 11 and 21 on Figure 3) were selected. According to harmonic rules of Western music, these four notes establish the tonality centre of the tune and are considered as important and useful to determine the tonality of the tune. The last tone (note 21) refers undoubtedly to the first degree of the tonality of the song, namely the fundamental tone. The three first phrases begin with “Happy” and are sung at the same note (notes 1, 6, and 11). This tone can be analysed as the fifth degree of the scale in which the tune is sung, called dominant, or as the fundamental note (1<sup>st</sup> degree) of the dominant seventh chord of the principal tonality. In both cases, these three notes carry important information about the tonality of the melody. Intervals performed between the four notes selected were compared to the standard value, with the same procedure as for the criterion “pitch interval deviation”. A small deviation reflected a high respect of the tonal centre of the melody.

## Results

Means and standard error of tempo, pitch interval deviation and tonal centre deviation for OS, WT singers and LT singers are reported in Table 1.

A one-way ANOVA was applied to compare the three groups (OS, WT singers, LT singers) and to estimate the difference in averages for the measurements of tempo and accuracy (pitch interval deviation and tonal centre deviation).

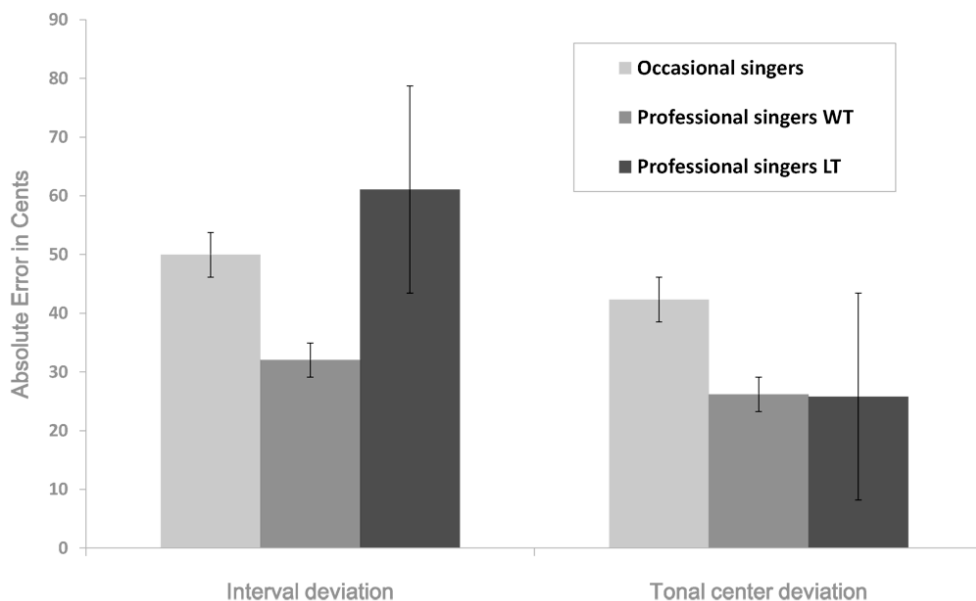
	Descriptive Statistics			One-way ANOVA
	Occasional singers <i>M (SE)</i>	Professional singers WT <i>M (SE)</i>	Professional singers LT <i>M (SE)</i>	
<b>Tempo</b> (beats per minute)	98.37 (1.79)	112.99 (4.59)	80.13 (4.53)	$F(2.88) = 16.61$ $p < .001$
<b>Interval deviation</b> (Absolute Error in Cents)	50.83 (3.29)	32.81 (2.25)	62.13 (7.66)	$F(2.88) = 5.07$ $p = .008$
<b>Tonal center deviation</b> (Absolute Error in Cents)	42.33 (4.14)	26.17 (3.32)	25.81 (6.17)	$F(2.88) = 3.00$ $p = .055$

**Table 1.** Measurements describing vocal performance (Tempo, Interval deviation criterion, Tonal centre deviation criterion) according to the group and comparison of averages between the sample groups.

As can be seen in Table 1, the means for each group were significantly different for the tempo [ $F(2.88) = 16.61, p < .001$ ], the pitch interval deviation [ $F(2.88) = 5.07, p = .008$ ] and tended to be different for the tonal centre deviation [ $F(2.88) = 3.00, p = .055$ ].

The performance tempo was quicker for WT singers ( $M = 112.99$  beats per minute,  $SE = 4.59$ ) than for OS ( $M = 98.37, SE = 1.79$ ). When the operatic vocal technique was applied, the tempo was slower ( $M = 80.13, SE = 4.53$ ) than for the two other groups.

As regards the measurements of accuracy, we conducted repeated measure T tests for the singing of WT and LT singers and independent measure T tests to compare the performances of OS with WT singers (see Figure 4).

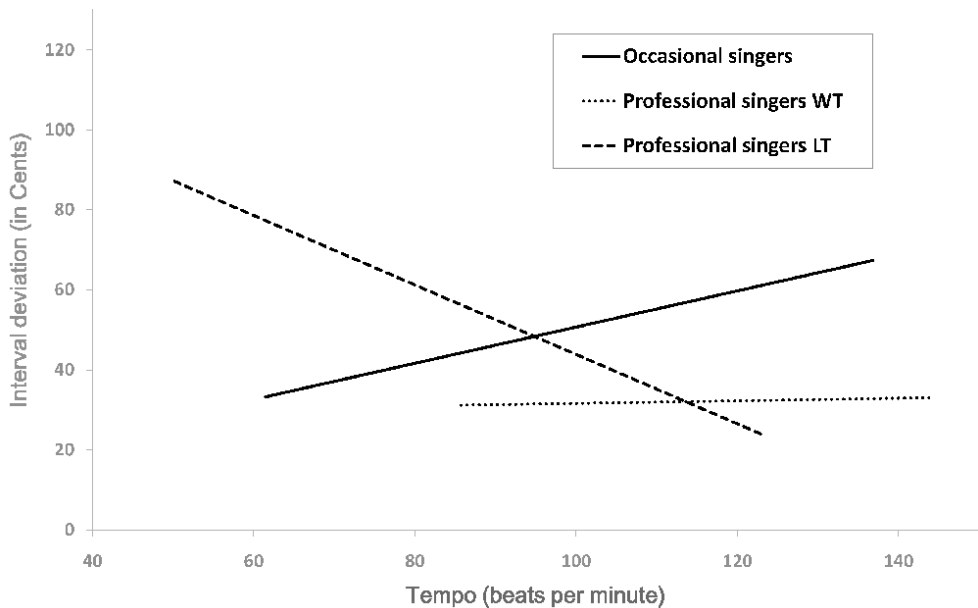


**Figure 4.** The average absolute value (in cents) of the “Interval deviation” and the “Tonal centre deviation” criteria for each group (OS, WT singers, LT singers), with standard error bars.

Regarding the tonal centre deviation, the OS performed worse ( $M = 42.33$ ,  $SE = 4.14$ ) than the WT singers ( $M = 26.17$ ,  $SE = 3.32$ ) [ $t = 3.05$ ;  $p = .004$ ] and LT singers ( $M = 25.81$ ,  $SE = 6.17$ ) [ $t = 2.22$ ;  $p = .035$ ]. There was no significant difference between the performance of the WT and LT singers.

Regarding the pitch interval deviation, the OS performed worse ( $M = 50.83$ ,  $SE = 3.29$ ) than the WT singers ( $M = 32.81$ ,  $SE = 2.25$ ) [ $t = 4.51$ ;  $p < .001$ ] but their performances were not significantly different from those of the LT singers ( $M = 62.13$ ,  $SE = 7.66$ ). When the operatic vocal technique was applied, more errors emerged than when the professional singers did not use this vocal technique [ $t = 4.27$ ,  $p = .001$ ].





**Figure 5.** Correlations between the Tempo and the “Interval deviations” for the three groups (OS, WT singers, LT singers).

As regards correlations between tempo and interval accuracy, we applied the Spearman’s rank correlation coefficient to each group of samples in order to observe the correlations between the tempo of the tune and the pitch interval deviation (see Figure 5). For the WT singers, no correlation was observed between these two measurements ( $r(14) = .13$ ;  $p = .659$ ) whilst for the OS, the accuracy increased when the singing tempo was slow ( $r(63) = .32$ ;  $p = .01$ ). The opposite was true for LT singers. Indeed, the singing was more accurate when the speed was quicker ( $r(14) = -.66$ ;  $p = .01$ ).

## Discussion

In order to validate our method for analysing the accuracy of a popular song, we analysed the performance of occasional and professional singers. For the occasional singers, we expected to confirm the observations of Dalla Bella et al. (2007) about the link between tempo and accuracy but not for professional singers. In this study, the WT singers sang quicker than the OS. Whereas we observed a correlation between the interval accuracy and the tempo for OS (Figure 5), this correlation did not appear for experienced singers. On average,

they had a quicker tempo than the OS but no link appeared between the speed of singing and the interval accuracy. The opposite is true for LT singers. On average, they sang more slowly than the OS but, in their case, the interval accuracy was better when the tempo was increased. We thus encountered in this instance the profile described in the literature (the slower it is, the more accurate it is) for the OS.

When comparing the accuracy of occasional singers who had not received any musical training with that of professionals well-versed in singing techniques, we expected to observe better accuracy amongst the latter and thus back up the relevance of the criteria for accuracy used and the quality of the analytical tools described in this study.

For analysis of accuracy, we chose to focus on the pitch interval deviation and the tonal centre deviation. By eliminating all singing in which contour errors appeared, we limited the cases of “poor singers”. Nevertheless, the results relating to the tonal centre deviation enabled us to confirm that the OS performed worse than the experienced singers (WT and LT). The effect of vocal expertise expected at this level was clearly visible. The criteria used in our method (first note of each phrase and final note) are therefore relevant and the type of analysis (lowest partial of the stable part of each vowel group) seems suitable for measurement of the “tonal centre” criterion.

For the interval accuracy, a difference appeared between the OS and WT singers. The results confirmed that vocal expertise leads to greater interval accuracy. The criterion (comparison of each interval produced with the target interval) as well as the type of analysis (lowest partial of the stable part of each vowel group) seem suitable for measurement of the “interval accuracy” criterion.

The possible artefacts reported during the mean  $F_0$  estimation (Dalla Bella et al., 2007; Pfordresher et al., 2010) were not found in our analytical procedure. The AudioSculpt software, used in our method, enabled the deletion of unharmonic frequencies and therefore automatic and correct recognition of the partials by the software. Selecting the lowest partial of segments, and the selected notes of the melody, the OpenMusic software reports automatically the mean  $F_0$  for the 21 notes, without any manual intervention. The verification was

possible thanks to the musical representation of the results of the analyses.

However, when “operatic” singing techniques are used, this difference no longer appears and the pitch interval accuracy is altered. Although we expected a good performance from LT singers, the results do not confirm this. This led us to discuss the criteria and analytical tools for judgement of the accuracy of the singing voice when the operatic vocal technique is applied.

We could not attribute these results to vocal expertise because the precision of intonation is better amongst musicians (Amir, Amir, & Kishon-Rabin, 2003; Dalla Bella, Giguère, & Peretz, 2007; Murry, 1990; Watts, Murphy, & Barnes-Burroughs, 2003) and also these same singers, without using a specific vocal technique, performed better than the OS.

Two interpretations can explain the results of the LT singers: the “pitch interval deviation” criterion may not be of paramount importance in judging these voices and/or our analytical tools may not be suitable for these voices.

Whilst the quality of singing is judged primarily on the basis of intonation (Watts et al., 2003), it seems that, in an operatic context, this criterion is no longer as evident and that accuracy within the meaning of “interval accuracy” may only be of secondary importance. Analysis of the accuracy was conducted on the stable segments of each vowel. We did our utmost to control the interpretation by not considering the liaisons between notes in our acoustic analysis. However, even if the composition is respected in classical music, the quality of the singing is also linked to the freedom of interpretation exercised by a singer. He or she intentionally modifies the size of the intervals (and the accuracy of the intervals is thus altered) without, however, modifying the musical phrase (the tonal centre is respected). This freedom of interpretation aimed at reinforcing the musicality may therefore be a criterion that would benefit from better control in future research aimed at objectively analysing the accuracy of singing.

In the case where the interval accuracy remains a criterion of paramount importance for estimating the quality of a vocal performance, the intention of LT singers would be to not alter this

precision. Our analytical tools provided averages for the fluctuations of the  $F_0$  and therefore calculated an average of vibrato for each note. Some studies highlighted that the pitch of the sound produced with vibrato is perceived as an average of the  $F_0$  and that the amplitude of the vibrato does not influence the perception of the sound's pitch (Shonle & Horan, 1980; Sundberg, 1978). On the contrary, this fluctuation in frequency is often considered as additional "colour", and represents a much sought after characteristic (LeBorgne, Lee, Stemple, & Bush, 2010). However, more recently, it has been showed that the perceived pitch with vibrato can be influenced by its phase and duration (van Besouw, Brereton, & Howard, 2008), and that range of acceptable tuning is accepted by the listener (van Besouw & Howard, 2008). In this study, the vibrato used by the LT singers was especially apparent. Here, calculating the average of the fluctuations may not be adequate and should be investigated in future researches. Whilst the vibrato contributes to the quality of the sound (Sundberg, 1994), our results indicate that it represents a bias during objective measurements of the accuracy of singing in a melodic context.

## Conclusion

In this study, two criteria (pitch interval deviation and tonal centre deviation) were observed by means of semi-automatic analytical tools. The performances of occasional and professional female singers enabled us to observe the expected effect of vocal expertise on accuracy but also to confirm the profile of occasional singers highlighted in the study by Dalla Bella et al. (2007). Whilst the application of an operatic singing technique renders the objective analysis of accuracy unreliable, the analytical tools as well as the criteria used in this study seem perfectly suited for an objective assessment of accuracy of a popular song in occasional and experienced singers.

## Acknowledgements

The authors thank the "Centre Henri Pousseur" of Liège and Guillaume Videlier for the technical support.

## ETUDE 2

# The evaluation of singing voice accuracy: A comparison between subjective and objective methods

---

Larrouy-Maestri P.<sup>1</sup>, Lévêque Y.<sup>2</sup>, Schön D.<sup>3</sup>, Giovanni A.<sup>2</sup>, & Morsomme D.<sup>1</sup>

<sup>1</sup> Logopédie de la Voix, Department of Psychology: cognition and behaviour, University of Liège, Belgium

<sup>2</sup> Laboratoire Parole et Langage, CNRS and Aix-Marseille University, France

<sup>3</sup> Institut de Neurosciences Cognitives de la Méditerranée, CNRS and Aix-Marseille University, France

In press in Journal of Voice

## Abstract

**Objective:** Vocal accuracy of a sung performance can be evaluated by two methods: acoustic analyses and subjective judgments. Acoustic analyses have been presented as a more reliable solution but both methods are still used for the evaluation of singing voice accuracy. This paper presents a first time direct comparison of these methods.

**Methods:** 166 untrained singers were asked to sing the popular song “Happy Birthday”. These recordings constituted the database analyzed. Acoustic analyses were performed to quantify the pitch interval deviation, the number of contour errors and the number of tonality modulations for each recording. Additionally, eighteen experts

in singing voice or music rated the global pitch accuracy of these performances.

**Results:** A high correlation occurred between acoustic measurements and subjective rating. The total model of acoustic analyses explained 81% of the variance of the judges' scores. Their rating was influenced by both tonality modulation and interval deviation.

**Conclusions:** This study highlights the congruence between objective and subjective measurements of vocal accuracy within this first time comparison. Our results confirm the relevance of the pitch interval deviation criterion in vocal accuracy assessment. Furthermore, the number of tonality modulations is also a salient criterion in perceptive rating and should be taken into account in studies using acoustic analyses.

## **Introduction**

The intonation of a sung melody, in terms of pitch accuracy, is an important factor to determine singing talent (Watts, Barnes-Burroughs, Andrianopoulos, & Carr, 2003). In experimental psychology, the accuracy of vocal performances has often been evaluated by music experts (Hébert, Racette, Gagnon, & Peretz, 2003; Racette, Bard, & Peretz, 2006; Schön, Lorber, Spacal, & Semenza, 2004; Wise & Sloboda, 2008). Vocal accuracy can also be objectively quantified by measuring the fundamental frequency ( $F_0$ ) variations along the performance. Since the SINGAD (SINGing Assessment and Development) system of Howard and Welch (1989), these acoustic methods have been developed (Elmer & Elmer, 2000) and presented as a more reliable solution to evaluate vocal accuracy, as they avoid the natural limits of a subjective judgment such as the imprecision of the exact deviation from the model pitch or the categorization of the pitch information with respect to the closest musical value (Dalla Bella, Berkowska, & Sowinski, 2011). Acoustic analysis consists in segmenting the auditory signal and extracting the  $F_0$  of sung vowels. Indeed vowels carry the maximum of voicing and stable pitch information (Murayama, Kashiwagi, Kashiwagi, & Mimura, 2004) and mark the onsets of musical tones (Sundberg & Bauer-Huppmann, 2007). In pitch-matching tasks, vocal accuracy is directly represented by

the difference between the produced pitch and the model (Wise & Sloboda, 2008; Alcock, Passingham, Watkins, & Vargha-Khadem, 2000; Alcock, Wade, Anslow, & Passingham, 2000; Amir, Amir, & Kishon-Rabin, 2003; Pfordresher & Brown, 2007; Watts, Moore, & McCaghren, 2005). In melodic contexts (Berkowska & Dalla Bella, 2009; Dalla Bella & Berkowska, 2009; Dalla Bella, Giguère, & Peretz, 2007; Pfordresher, Brown, Meier, Belyk, & Liotti, 2010), measures are rather based on relative pitch differences, linked to the succession of intervals, and so avoid the effect of a change in key at the beginning of a tune, which would lead to errors in the rest of the melody (Schön, Lorber, Spacal, & Semenza, 2004).

Though acoustic analyses have the advantage to provide an objective and reproducible measurement of vocal accuracy (Dalla Bella & Berkowska, 2009; Dalla Bella et al., 2007; Pfordresher et al., 2010; Terao, Mizuno, Shindoh, & al., 2006), they also have some limits. First, objective measurements require choosing a set of features or dependent variables that possibly describe vocal accuracy and/or what should be considered as an “error” (Dalla Bella et al., 2011). Second, computer-assisted methods seem limited when it comes to vocal accuracy assessment of a full song (Wise & Sloboda, 2008). While the interpretation is part of the quality of a musical performance, Wise and Sloboda (2008) believe that a rating scale suits the evaluation of a melody better than signal processing methods. Third, objective tools for acoustic analysis may not be adapted to all vocal data types. For example, objective measurements adapted for a song performed by untrained singers do not suit when an operatic technique is employed (Larrouy-Maestri & Morsomme, in press). Finally, despite the increase of automated tools, the acoustic analysis methods are still time-consuming and not easily implementable in contexts such as musical or clinical evaluation.

Although technical advances reversed the trend from perceptive methods to acoustic methods in research on singing (Dalla Bella et al., 2011), the judges’ assessment and the objective measurements have never been clearly compared. Otherwise the comparison between the self-evaluation and the objective measurements is well documented. Studies about tone deafness and poor-pitch singers showed the difficulty for participants to evaluate accurately their musical abilities

(Cuddy, Balkwill, Peretz, & Holden, 2005; Pfordresher & Brown, 2007; Wise & Sloboda, 2008).

In the present study, we compared objective and subjective methods assessing the vocal accuracy of a melody sung by 166 participants. Our aim was to (a) estimate the correlation level between the two methods and (b) find out which acoustic measurements can predict the judges' scores. For (a), we used the following variables: a global rating by experts versus a quantitative measurement of pitch interval deviation, the most common acoustic criterion to assess vocal accuracy (Berkowska & Dalla Bella, 2009; Dalla Bella & Berkowska, 2009; Dalla Bella et al., 2007; Larrouy-Maestri & Morsomme, in press; Pfordresher et al., 2010). For (b), we examined the predictive power of two additional variables: the number of contour errors and the number of tonality modulations. The importance of the melodic contour in music processing has been continually demonstrated employing a wide array of perceptual encoding, similarity, and memory paradigms (Schmuckler, 2009). The second variable has never been directly measured using acoustic methods although some previous studies showed its perceptive relevance for the vocal accuracy evaluation of children or adult performances (Flowers & Dunne-Sousa, 1990; Price, 2000).

## Methods

### *Participants*

One hundred sixty-six untrained singers (57 men, 109 women) were recruited among the Belgian population. Their age ranged from 14 to 76 ( $M = 29.89$  years,  $SD = 14.47$ ). The majority (64%) reported listening to music less than one hour a day, only 6% declared going to concerts more than once a month. None of them were professional musicians, 3% had studied a musical instrument in a conservatory of music and 24% reported a basic musical education in a local music school.

The jury was composed of eighteen subjects with expertise in singing voice and/or music. 4 music students ( $M = 20$  years) with an average of 12.5 years of musical training, were recruited in



conservatories and were all following piano classes. The professional musicians were 5 instrumentalists ( $M = 39.8$  years) with an average of 28.2 years of musical training and 5 singers ( $M = 39.4$  years) with an average of 16.6 years of vocal training. They all followed a classical music education in high institutions and were still performing in public when the study took place. The singing voice experts were 4 speech therapists specialized in singing voice treatment ( $M = 31$  years).

## ***Procedure***

*Song recording:* Participants were asked to produce two vocal glissandi (sliding up to the highest pitch and down to the lowest pitch in a comfortable range). The experimenter illustrated the vocal exercise by singing and imitating manually the movement of the voice. The aim of these glissandi was to warm up the vocal organs (Amir, Amir, & Michaeli, 2005), to verify the vocal capacity of the subjects and to encourage a lack of inhibition in front of the experimenter and the recording equipment. Then, they performed individually the popular French song “Happy Birthday” *a cappella*. Participants were instructed to sing “naturally, whilst imagining a festive and friendly context”. No particular starting note was given in order to let the participant choose his/her comfortable range. The sound recordings were made using a head-worn microphone (Sennheiser HS2) positioned at a constant distance of 2 cm from the right corner of the mouth and a Marantz Professional Solid State Recorder (PMD67). All these sung performances are listed in a database which can be viewed following the link: <http://sldr.org/sldr000774/en>.

*Self-evaluation:* After the sung performance, participants were instructed to rate their proficiency to sing in tune on a 9-point scale with 1 indicating “very inaccurate” and 9 “very accurate”.

*Judge’s rating:* Each jury member listened to the 166 recorded songs through headphones and rated them on a 9-point scale with 1 indicating “very inaccurate” and 9 “very accurate”. The instruction was to take into account the overall vocal pitch accuracy to make their judgments, and no other criteria than pitch. Four practice trials selected from the database were presented in order to verify the understanding

of the instruction and to allow the judges to adjust their assessment. They performed the evaluation individually.

## ***Description of the objective method***

*Acoustic analysis:* The popular song "Happy Birthday" is composed of four phrases. It has 25 notes (each note corresponding to a syllable), 21 when one ignores the rhythmically concurrent repeated notes (Figure 1). Data processing was semi automatically done in two stages on a MacBook Pro (Mac OS X, Version 10.6.5). Analyses were performed through AudioSculpt 2.9.4v3 and OpenMusic 6.3 software (Ircam, Paris, France) using a Short Time Fourier Transform (STFT) analysis. Markers were manually placed on the spectrogram, where the fundamental frequency and the 10 first partials were clearly visible, in order to avoid the attacks and the glides between notes. Then, the mean  $F_0$  estimation in the segments was automatically calculated and converted into cents (1/2 tone = 100 cents) for the 21 notes of the tune.



**Figure 1.** Score of the tune "Happy Birthday" with the number of notes used for calculating accuracy.

*Criteria observed:* The measurements are based on melodic intervals. The calculations are grounded on the equal temperament (i.e. constant frequency multiple between the notes of the chromatic scale) which is a compromise tuning scheme used in Western music. As Figure 2 illustrates, for each production, three criteria of vocal accuracy were quantified: pitch interval deviation, number of contour errors and tonality modulations.

**Pitch interval deviation:** We calculated the difference in cents between each performed interval and the theoretical one. We considered the absolute value of the differences (to avoid sharp and flat errors cancelling each other) and computed the average score across the entire melody. A small deviation reflects a high precision of intervals.

**Contour errors:** We counted each time that the produced interval direction deviated from the direction of the musical score.

Tonality modulation: We computed the number of modulations, defined as an interval error larger than 100 cents not followed by a corrective interval of at least 100 cents in the reverse direction. Thus, modulations indicated that tonality changed during at least three notes (2 intervals).

Figure 2 consists of four musical staves, each showing a sequence of five notes (numbered 1 to 5) under a single slur. The notes are: 1 (quarter), 2 (quarter), 3 (quarter), 4 (quarter), and 5 (half). The key signature has one flat (B-flat).

- Accurate version:** The notes are G4, A4, Bb4, C5, and Bb4.
- Contour error at the note 2:** The notes are G4, A4, Bb4, C5, and Bb4. Note 2 (A4) is slightly higher than in the accurate version.
- Interval errors of 200 cents between notes 1-2 and 2-3:** The notes are G4, A4, Bb4, C5, and Bb4. The interval between note 1 and 2 is 200 cents, and the interval between note 2 and 3 is 200 cents.
- Tonality modulation at the note 2:** The notes are G4, A4, Bb4, C5, and Bb4. Note 2 (A4) is slightly higher than in the accurate version, and note 5 (Bb4) is lower than in the accurate version.

**Figure 2.** Illustration of the errors observed through the objective method.

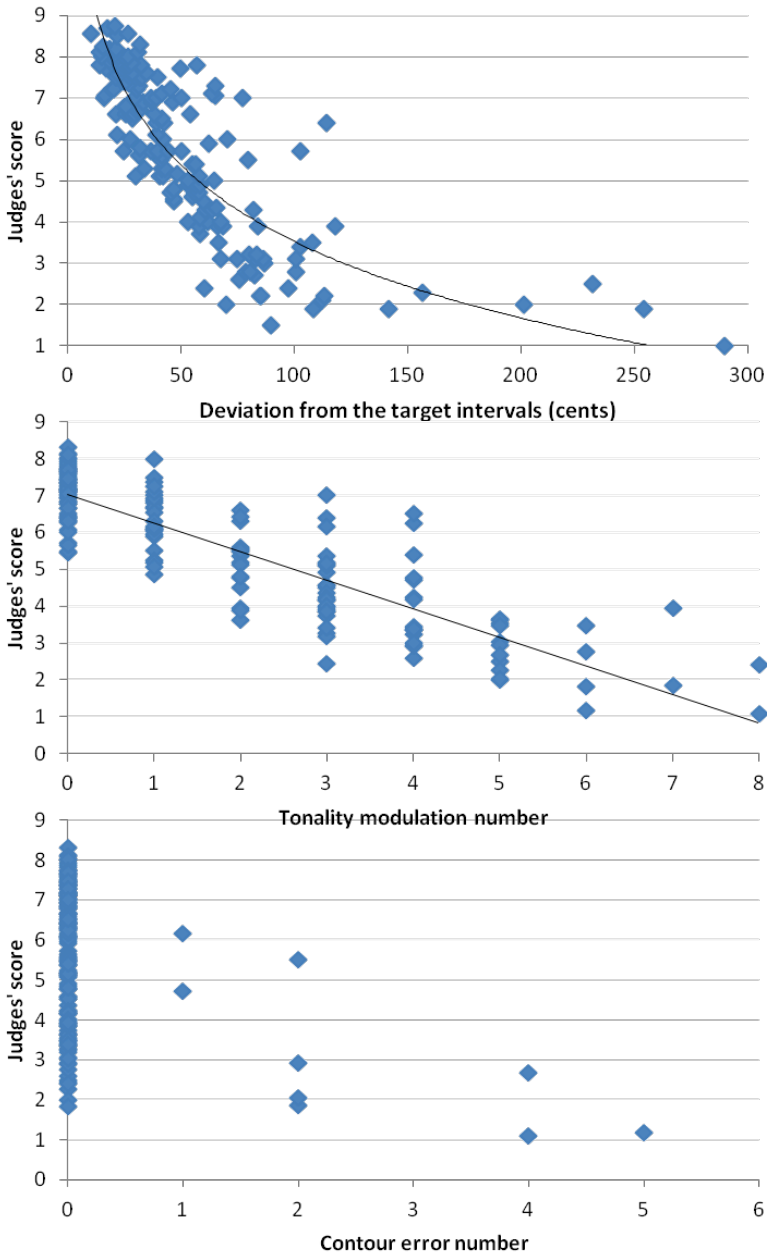
## Results

We computed a correlation matrix using Spearman coefficient to estimate the pairwise correlations between the 18 judges. We found a median correlation coefficient of  $r = 0.77$  ( $SD = 0.08$ ,  $p < .01$ ). As the 18 judges provided strongly and significantly correlated ratings, the mean rating from the whole group has been used in the following analyses.

### ***Interval deviation and subjective judgments of vocal accuracy***

Deviation from target intervals is the most common criterion used in acoustic analyses to estimate the vocal accuracy. In order to know to what extent this criterion is correlated with subjective judgments, we calculated correlations between interval deviation, scores of vocal accuracy given by the judges, and self-evaluation of vocal accuracy using the Spearman coefficient, for the 166 subjects.

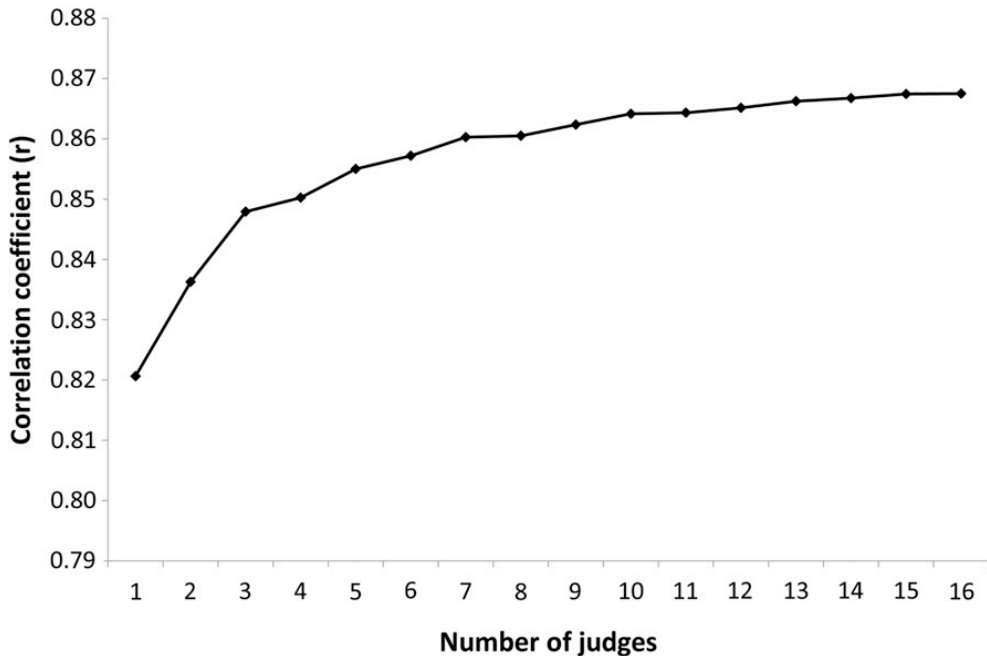
We found a high and significant correlation between the interval deviation criterion and the average score given by the judges ( $r = -0.87$ ;  $p < .01$ ), as illustrated in Figure 3 (top).



**Figure 3.** Relationship between Judges' scores and Interval deviation (top), Judges' scores and Tonality modulation number (middle), Judges' scores and Contour error number (bottom).

We then computed this correlation coefficient with sub-groups of judges of different size ( $n = 1$  to 16). For this purpose, we took random

sub-groups among the original 18 judges (100 random sub-groups of each size, from 1 to 16 judges) and computed the correlation between the mean rating of each subgroup and the semi-automatic interval deviation measure. We found a slight decrease of  $r$  for smaller groups of judges. Figure 4 shows the relationship between the correlation coefficient  $r$  and the number of judges. Data suggest that 3 judges are sufficient to get an average correlation coefficient superior to 0.85.



**Figure 4.** Evolution of the correlation Interval deviation/Average Judge score as a function of the Number of judges ( $n$ ) in the panel. We randomly selected 100 sub-groups of  $n$  judges among our 18 experts and iterated a Spearman correlation test between each sub-group of judges' mean rating and the interval deviation of singers.  $N$  varies between 16 and 1.

Self-evaluation by the subjects was moderately but significantly correlated to the interval deviation criterion,  $r = -0.35$ ;  $p < .01$ , and in the same extent to the judges' scores,  $r = -0.35$ ;  $p < .01$ . Self-evaluation appeared affected by some overrating. Notably, 41% of the inaccurate singers (judges' score  $< 4$ ,  $N = 39$ ) overestimated their vocal skill (self-evaluation  $> 4$ ). Conversely, a larger number of participants undervalued their vocal skill: 63% of the very accurate singers (judges' score  $> 6$ ,  $N = 84$ ) undervalued themselves (self-evaluation  $< 6$ ).

## *Acoustic parameters predicting the judge rating*

To determine whether acoustic variables other than the interval deviation criterion might contribute to predict the score of vocal accuracy given by the judges, we performed a multiple linear regression analysis entering interval deviation, number of tonality modulations, and number of melodic contour errors. As we observed a logarithmic relationship between judges' scores and interval deviation criterion (Figure 3, top), the logarithm of the interval deviation was used in the multiple regression.

The results of the analysis are shown in the Table, including beta weights and significance tests for each variable. In addition, the Table displays the results of the significance test for the multiple  $R^2$  for the set of 3 predictor variables.

$R^2$	0,81	
Adjusted $R^2$	0,81	p
F	232,17	2,00E-58

	Beta	t value	Pr(> t )
(Intercept)		16,30	5,68E-36
Intervals	-0,51	-6,61	5,43E-10
Tonality	-0,45	-6,32	2,39E-09
Contours	0,08	1,89	0,06

**Table 1.** Summary of multiple regression analysis on judges' scores with the three acoustic variables used as predictors. Intervals = mean size of interval deviation; Tonality = number of tonality modulations; Contours = number of contour errors. For each variable, the beta weights and significance tests are represented.

As shown in Table 1, the result of the regression analysis indicated that two variables predicted the score of vocal accuracy given by the judges: the pitch interval deviation ( $\beta = 0.51$ ;  $t = -6.61$ ;  $p < .001$ ) and the number of tonality modulations ( $\beta = -0.45$ ;  $t = -6.33$ ;  $p < .001$ ). The number of contour errors did not contribute to explain the judges'

scores ( $\beta = 0.08$ ;  $t = 1.89$ ;  $p = .06$ ), probably because this type of error was very rare. The correlation between these two former criteria and the judges' score is illustrated in Figure 3 (middle and bottom).

The total model explained 81% of the variance of judges' scores ( $F = 232.17$ ;  $p < .001$ ).

## Discussion

For the first time, this study directly compared two different methods to assess the vocal accuracy over the same material. We observed that the 18 judges provided significantly correlated ratings in the context of a popular song assessment. This result shows that the voice and/or music experts used similar subjective criteria despite their different backgrounds.

For the 166 sung performances, we found a high and significant correlation between the interval deviation and the scores given by the judges ( $r = 0.87$ ,  $p < .01$ ). Previous researches used objective or subjective methods, without comparing them. In this study, both objective and subjective approaches were used and compared, showing similar results in the assessment of vocal accuracy in a melodic context. Interestingly this correlation between objective and subjective methods was robust enough to survive downsampling to a number of 3 raters from the judge group. We observed a logarithmic relationship between the pitch interval deviation criterion and the subjective measurement, which means that the perceptual rating was more discriminatory among the accurate singers than for the singers with a low score according to the pitch interval deviation criterion. Judges seemed more critical when an accurate singer made a false note but did not make a difference between slightly inaccurate to very inaccurate sung performances.

When the perceptive assessment was done by the participants themselves, the correlation between interval deviation and subjective measurement was moderate ( $r = -0.35$ ;  $p < .01$ ). The present study confirms the difficulties in self-evaluation pointed out in previous studies (Cuddy et al., 2005; Pfordresher & Brown, 2007). These results could be attributed to a lack of musical expertise of the participants (e.g. a wrong representation of the target melody) or to an intrinsic difficulty



of the self-evaluation procedure (e.g. the social nature of self-assessments in general or specific auditory feedback integration). Comparing the subjective assessment of the present study with the rating of naive judges, without any music or vocal expertise, would allow a better understanding of the causes of such difficulties. Indeed, a similar pattern of results would indicate that difficulties in self-evaluation would not be due to a lack of musical expertise.

This study shows the correlation between objective and subjective measurements but also allows us to examine the predicting weight of the three criteria used in objective measurements. The regression analysis indicated that the pitch interval deviation criterion was related to the score of vocal accuracy given by the judges ( $\beta = 0.51$ ;  $t = -6.61$ ;  $p < .001$ ). This study thus supports the use of this criterion for the assessment of vocal accuracy in a melodic context (Berkowska & Dalla Bella, 2009; Dalla Bella & Berkowska, 2009; Dalla Bella et al., 2007; Larrouy-Maestri & Morsomme, in press; Pfordresher et al., 2010). In our study, two other criteria were observed: the number of contour errors and the number of tonality modulations. The regression analysis indicated that the number of contour errors did not explain the judges' scores ( $\beta = 0.08$ ;  $t = 1.89$ ;  $p = .06$ ). This pattern can be discussed regarding the choice of the melody, which leads to few contour errors. This tune has been chosen for its strong tonal center and the diversity of intervals. Also, the French version of the popular song "happy birthday" is very common and learned early in the childhood. Years of familiarity (ear and vocal) and the simplicity of the material could explain that untrained singers sang the good contour of the melody. It would be interesting to propose a more complex tune, which would induce more contour errors to be analyzed. Finally, this study highlights that the judges' score is partly explained by the number of tonality modulations ( $\beta = -0.45$ ;  $t = -6.33$ ;  $p < .001$ ). This variable has never been included in an acoustic analysis so far but this result supports the importance of the tonality component in vocal accuracy assessment (Flowers & Dunne-Sousa, 1990; Price, 2000).

Finally, even if the judges were asked to focus on the pitch of the melody, the unexplained 19% of variance in the regression model could be linked to other criteria such as rhythm accuracy or vocal quality, criteria that could unintentionally intervene in the accuracy judgment.

Future research would need to investigate these criteria to better understand the perceptual judgment process.

## **Conclusion**

In this study, we compared an objective and a subjective method to assess the vocal accuracy of a popular song performed by 166 untrained singers. Our results highlight the congruence between objective and subjective measures, whilst the subjective ratings are performed by expert judges. In analytical computer-assisted methods, the measured variables can be controlled but their use must be preceded by theoretical choices. Our results clearly confirm the weight of the pitch interval deviation criterion in the vocal accuracy assessment. Furthermore, this study also underlines that the number of tonality modulations is a relevant criterion in perceptive rating and should be taken into account for the objective vocal accuracy assessment.

## **Acknowledgements**

The authors thank the “Centre Henri Pousseur” of Liège and Guillaume Videlier for the technical support.

# ETUDE 3

## Singing ability is rooted in vocal-motor control of pitch

---

**Sean Hutchins<sup>1</sup>, Pauline Larrouy-Maestri<sup>2</sup>, and Isabelle Peretz<sup>1</sup>**

1. International Laboratory for Brain, Music and Sound Research (BRAMS), Université de Montréal, Canada
2. Department of Psychology: Cognition and Behavior, University of Liège, Belgium

Under review in Attention, Perception & Psychophysics

### **Abstract**

The inability to sing in tune can be caused by poor pitch perception or by poor vocal-motor control. Although previous studies have tried to examine the relationship between pitch perception and vocal production, they have failed to control for the timbre of the target to be matched. In the present study, we compare accuracy in instrumental and vocal pitch matching paradigms, designed such that each participant's own voice serves as both the target and the response. We also measured pitch accuracy in singing a familiar melody ("Happy Birthday") to assess the relationship between single pitch matching tasks and melodic singing. Our results showed that nonmusicians were significantly better at matching their own voices on the instrument than by singing, indicating that vocal-motor control is an important limiting factor on singing ability. We also found significant correlations between the ability to sing a melody in tune and vocal pitch matching, but not instrumental pitch matching. Better melodic singers also tended to have higher quality voices. These results provide important evidence about

the ultimate causes of poor singing ability, and demonstrate that single pitch matching tasks can be useful in measuring general singing abilities.

## Introduction

Music is an important and universal aspect of culture, and one of the most prevalent forms of musical activity is singing. Nevertheless, many people do not sing well, despite having no problems hearing or understanding music. Poor singing ability can have several manifestations, including problems with timing and with timbre. However, the most common manifestation of poor singing ability is poor pitch control (e.g. Dalla Bella, Giguère, & Peretz, 2007). In addition, music educators rank pitch intonation as the single most important factor in determining someone's singing talent (Watts, Barnes-Burroughs, Andrianopoulos, & Carr, 2003). Because of this, many studies of singing ability have focused on the ability to match one or more pitches (e.g. Estis, Coblenz, & Moore, 2009; Hutchins & Peretz, 2012; Pfordresher & Brown, 2007; Watts, Moore, and McCaghren, 2005). This is commonly measured acoustically, by taking the error of the sung note relative to the target note.

In order to vocally match a target note, a singer must perceive the note, must determine the configuration of their vocal apparatus that will create a note of the same pitch, and must enact that motor command. An error in any of these steps will lead to inaccurate vocal pitch matching (Berkowska & Dalla Bella, 2009; Pfordresher & Brown, 2007). Many studies searching for the cause of poor singing abilities have focused on perceptual abilities, generally by correlating vocal pitch matching accuracy with a measure of pitch perception. Although some such studies have found relationships between these two abilities (e.g. Estis et al., 2009; Estis, Dean-Claytor, Moore, & Rowell, 2011; Moore, Keaton, & Watts, 2007; Watts et al., 2005), several others have failed to do so (e.g. Bradshaw & McHenry, 2005; Dalla Bella et al., 2007; Moore, Estis, Gordon-Hickey, & Watts, 2008; Pfordresher & Brown, 2007). Ultimately, however, it is unlikely that problems with pitch perception can account for the majority of instances of poor pitch-matching abilities, given that errors in vocal pitch matching (when they occur) are often much larger than errors in measured pitch perception ability.

A recent study by Hutchins and Peretz (2012) used a novel method to investigate the relationship between the ability to perceive and vocally imitate pitches. This study used a new instrument called a slider as a non-vocal alternative to pitch matching. The slider is played by pressing on a horizontal touch sensitive strip, and creates a synthetic vocal tone based on the position at which it is pressed. Unlike a piano, this instrument is not divided into discrete steps, but can create any pitch within its range, just like the voice can. Hutchins and Peretz (2012) asked musicians and nonmusicians to match synthesized vocal tones on the slider and with their own voices. Participants were considerably better at matching pitches with the slider than with their voices, despite their unfamiliarity with the new instrument. As a control for timbre, participants were also asked to match examples of their own voice singing, with the target examples taken from prior recordings. Participants were more accurate at matching their own voices than synthesized tones, but were still less accurate than they were with the slider. Overall, the pattern of results indicated that poor vocal pitch matching ability was not generally caused by poor perception ability, as poor singers were generally able to match pitches accurately on the slider. Rather, of the 31 nonmusicians tested, 20% were impaired on both singing tasks, indicating a vocal-motor impairment, and 35% were impaired at matching synthesized vocal tones, but not at matching their own voice, indicating a sensorimotor problem involving translating between timbres.

An alternate explanation for the general superiority on the slider than on vocal self-matching is related to the timbres used in each task. Participants were able to make more accurate tuning judgments about the synthesized vocal timbre created by the slider than about natural vocal timbres (Hutchins & Peretz, 2012, Experiment 5). This general trend to be less discerning of tuning errors in the voice than in other instruments, termed the vocal generosity effect, was confirmed and extended in a later study (Hutchins, Roquet, & Peretz, 2012). Thus, participants' lower errors on the slider task than the self-matching task may only have been a function of a better ability to resolve tuning for the slider's timbre.

Another question that can be raised about this and many other studies of singing ability is the relationship between the ability to sing

single tones and whole melodies. Most studies of singing ability focus on pitch matching of single tones or small numbers of tones, with the assumption that this ability will scale up to the ability to accurately sing whole melodies (with a few exceptions, e.g. Dalla Bella et al., 2007). However, this assumption has not yet been validated experimentally. Pfordresher and Brown (2007) did show that error on single pitch matching is correlated with pitch errors in two- or four-tone contexts, but there are many reasons why single pitch matching might not predict melodic singing ability, such as the effect of establishing a tonal context, or the cumulative effect of corrective errors (a negative lag-1 correlation). Given that the ability to sing melodies accurately is generally of much greater importance than matching individual pitches when evaluating overall singing ability, it is important to quantify this relationship.

Finally, one other outstanding issue concerns the relationship between singing ability and other measures of vocal quality. Although the mean pitch error is the primary factor in most evaluations of singing ability (Watts et al., 2003), other factors can play a role in how we evaluate singing ability. Our intuition is that poor pitch singers would also tend to have lower quality vocal timbres, but it is possible that these abilities are unrelated, which would mean that there exists a category of singing problems which remains unevaluated by pitch error-based measurements. In addition, the vocal generosity effect (Hutchins, Roquet, & Peretz, 2012) also demonstrates that timbre can have a significant effect on our tuning judgments, making it important to evaluate these qualities separately.

To address these questions, we designed a new experiment to compare pitch matching ability on the slider and voice to each other and to melodic singing ability, while maintaining a consistent timbre. Nonmusicians (chosen here because they show a greater range of singing abilities and are more representative of the general population) first recorded single tones at different pitch heights to use as target stimuli. They imitated those recordings with their voice and on the slider. In contrast with the previous experiments, the slider sound was not synthesized, but rather was based off the recordings of each participant's own voice, and manipulated using a digital signal processor designed to tune vocal performances in real time. Thus, the vocal pitch

matching responses and the slider-based pitch matching responses had exactly the same timbre. Following the pitch matching tasks, participants were also asked to sing a complete version of a well-known song (“Happy Birthday”). All sung tones were analyzed acoustically for pitch, and vocal quality measurements (pitch stability, jitter, shimmer) were measured for the recorded target tones.

If timbral differences were responsible for participants’ better performance on the slider than singing tasks in Hutchins and Peretz (2012), we should see no difference in pitch error between singing and slider pitch matching tasks. This would indicate that pitch perception impairment could still be a cause of poor singing ability. However, we hypothesize that we will continue to observe better performance in the slider condition than the singing condition, even when the timbres are the same in both tasks, which would provide confirmatory evidence that singing problems are caused by sensorimotor and vocal-motor impairments, rather than perceptual impairments. We also hypothesize that vocal pitch matching ability, but not slider pitch matching ability should correlate with melodic singing ability, which would confirm the greater importance of vocal motor abilities than perception abilities in determining overall singing ability. Finally, we expect to see that participants with better measurements of vocal quality will be better at vocal pitch matching, as more practice singing would tend to benefit both vocal pitch and quality.

## Methods

### *Participants*

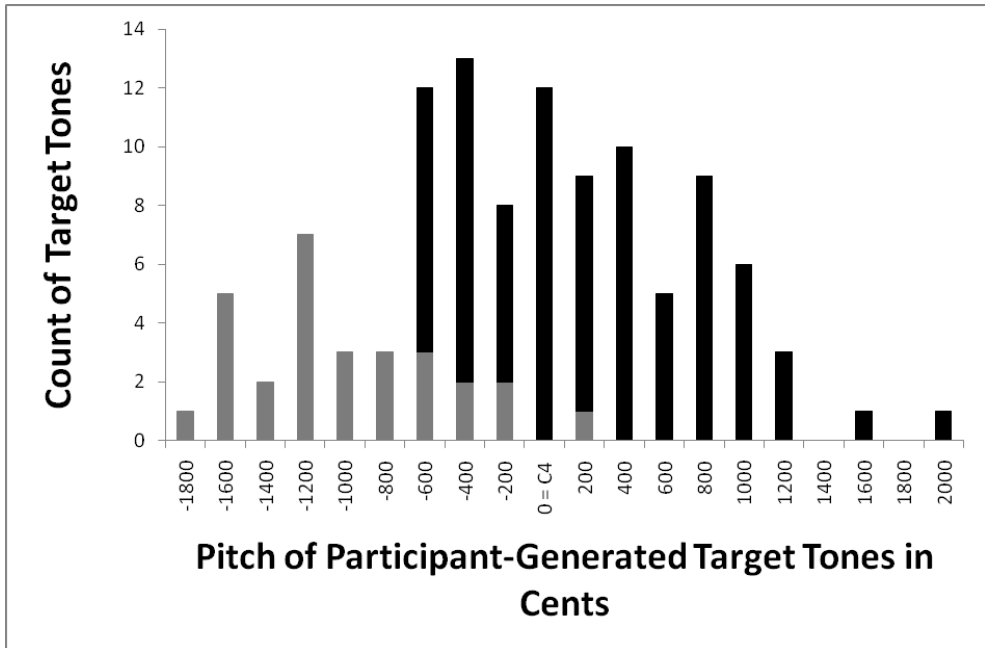
The participants were 22 nonmusicians (16 female), recruited from the Montréal population. All participants had 1 year or less of formal training ( $M = .2$  years), and ranged in age from 18 to 30 ( $M = 23$  years). They reported a mean of .32 years of informal musical experience, a mean of .32 years of group singing experience, and no formal singing training. No subjects reported any diagnosed hearing deficits or neurological disorders.

## ***Stimuli and procedure***

The session was divided into four sections: the recording of target tones from the participants' own voice, an instrument-based pitch-matching task, a vocal pitch-matching task, and a singing by memory task. All vocal stimuli and responses were recorded with a Neumann TLM 103 microphone (Georg Neumann GmbH, Berlin, Germany). Video examples of target recording and the two pitch matching tasks can be viewed at <http://www.brams.umontreal.ca/slidervoice>.

**Recording targets.** In the first section of the experiment, participants recorded their own voice at five different pitch levels, all on the syllable /ba/, sustained for 2-3 seconds. They sang a low tone, a medium-low tone, a medium tone, a medium-high tone, and a high tone. Tones were self-selected, to ensure that each was within the participant's comfortable range. Three different versions of each pitch level were recorded through Max/MSP (Cycling '74, San Francisco, California, USA). After all tones had been recorded, the experimenter chose the best example from each of the five categories, using the criteria of pitch stability, voice quality, and differentiability from other pitches. These tones were then normalized for amplitude and trimmed to remove silence from the beginning and end. The five targets were used as the five target tones for the pitch-matching tasks. Figure 1 shows the range of targets produced across all participants.





**Figure 1.** Histogram of the target tones chosen, sorted by frequency, in bins of 200 cents. Middle C (C4, 261 Hz) is represented as 0. The target tones of males are shown in grey, those of females in black.

**Slider pitch matching.** In the slider task, participants were presented with one of their previously-recorded sung tones as a target, and their task was to use the slider to find the same pitch as the target tone. The slider, designed by Hutchins and Peretz (2012), provides a non-vocal pitch-matching measurement that can be compared to vocal pitch-matching. The slider produces a sound with a pitch based on the position of a finger press, and is designed to be easily used by nonmusicians. The slider is made from a 50 cm position sensor overlaid on a pressure sensor (Infusion Systems, Montreal, Canada). The slider can register 1024 unique positions, making each position less than half a millimeter apart.

In each trial, one of these positions was randomly chosen as the target position (not including the top or bottom 1/6<sup>th</sup> of the slider- 170 positions on each side- to ensure adequate space for adjustment). Any press on the slider triggered the playback of the target tone. However, the output of the slider-triggered playback was routed through a VoicePro digital signal processor (TC-Helicon, Victoria, BC, Canada), programmed to apply a pitch shift to the target tone. The applied pitch

shift could range anywhere from plus 10 semitones to minus 10 semitones, and the specific shift was determined by the distance between the target position and the position of the finger press<sup>2</sup>. Finger presses on the slider to the right of the target position triggered a positive (sharp) shift, those to the left triggered a negative (flat) shift, giving the slider the same pitch orientation as a piano. Each discrete position on the slider represented a pitch shift of 1.17 cents, chosen so that the pitch distance between the two extremes of the slider was equivalent to one octave (1200 cents = 12 semitones = 1 octave). For example, pressing on the slider 4.17 cm to the right of the target position would yield a difference of +85 steps from the target, and would trigger the output of a tone shifted upwards by 1 semitone. The shifted output is identical to the original in duration, amplitude, and internal pitch change (e.g. instability, vibrato), but is shifted globally in pitch. Pressing precisely on the target position plays the original target tone, unshifted.

Each trial was initiated by the participant pressing the space bar, and began with an automatic presentation of the target tone, without any shift applied to it. Following this initial presentation, the participant pressed on the slider and heard the resulting shifted version of target. In order to minimize the role of pitch memory, the original target was played following every version triggered by a slider press, giving participants the ability to immediately compare the pitch of subsequent versions. Participants could also press the Enter key at any time to re-listen to the original target (although they rarely chose to). The target

---

<sup>2</sup> It is a possibility that the large pitch shifts at the extremes of the slider may have introduced distortions that could have provided a timbral cue to target finding. This is a problem for any method of shifting a non-synthesized pitch. However, we believe that this would have minimal effect on the results for three reasons. First, the digital signal processor (VoicePro) used to create the pitch-shifted versions of the target is designed for the voice, and other testing in our lab has shown very low levels of audible distortion for pitch shifts up to 2 semitones. Second, even unshifted versions of the tone are routed through the signal processor, with an applied shift of zero cents. Finally, the effects of the signal distortion would decrease as the participant moved closer to the target position on the slider, making timbral cues less useful for fine adjustments. Because participants were in general very accurate on the slider (13 cents error on average, with 98% of responses within 50 cents), it seems unlikely that timbral cues would be the primary source of feedback for determining the final response. Despite this, though, we cannot categorically rule it out as an explanation (though we should note that a vocal pitch match will also not have precisely the same timbre as the target either).

playback was halted whenever the slider was pressed, in order to avoid the superposition of the target sound and the slider-generated sound.

Participants were unaware of the randomly-chosen target position in each trial, and were instructed to find the position on the slider that generated a sound as close as possible to the target tone. They were told that they could respond as many times as they wished until they found the best pitch match for the target, and that their accuracy would be judged only by their final response. Participants indicated that they had finished matching the target by pressing on the spacebar, to end the trial.

Participants were familiarized with the slider before the task, and were allowed to use it freely to become acquainted with it. After this familiarization, they were presented with five practice trials. The main experiment consisted of 75 trials, with 15 instances of each of the five target tones, presented in pseudo-random order. Due to time limitations, not all participants could complete all the trials but each participant completed a minimum of 25 trials ( $M = 59.14$  trials). All data from each trial, including the position of the target, the time and position of each press on the slider, and the shift applied to each output, were recorded and saved as .txt files through Max/MSP.

**Vocal pitch matching.** The vocal pitch-matching task used the same trial design as the slider pitch-matching task but differed only in the medium of the response. The order of these two tasks was counterbalanced across participants. The vocal pitch matching section was preceded by five practice trials. This section consisted of 100 trials, with 20 instances of each of the five target tones (this is due to the greater variability in vocal pitch matching and the shorter average time to complete each trial in this condition; see Hutchins & Peretz, 2012), presented in pseudo-random order. Each participant completed a minimum of 77 trials ( $M = 97.23$  trials).

As in the slider pitch-matching task, each trial was initiated by the participant pressing the space bar, and began with a presentation of the target tone. Participants were instructed to match the target tone as closely as possible using their voice and told that they could make as many attempts as they liked to match the target. They were asked to wait until the self-recorded target tones had finished playing before

singing and could press on the enter key to re-listen to the target. Participants were instructed that their accuracy would be judged only by their final response. Participants indicated that they had finished matching the target by pressing on the spacebar, to end the trial. The participant's voices were recorded as .aif files.

**Melodic Singing.** In the final section of this experiment, participants were asked to perform the song "Happy Birthday", and were not given a particular starting note. The instruction was to sing "naturally, whilst imagining a festive and friendly context". Participants were informed about the aim of this task, i.e. the observation of the pitch accuracy in a melodic context. The participant's sung performances were saved as .wav files.

## ***Analyses***

**Vocal Quality Analysis.** Acoustic measurements of voice quality were made for each target tone performed by participants in the first section of the experiment using Praat (Boersma & Weenink, 2012). We measured the standard deviation (SD) of the fundamental frequency ( $F_0SD$ ) across the duration of the note, as an indication of its pitch stability, the jitter (a measurement of the period perturbation, in percentage) and the shimmer (a measurement of the amplitude perturbation, in percentage). These latter two are commonly used for the evaluation of voice disorders (Baken, 1987; Sataloff, 2005; Titze, 2000). Note that for those three measurements, a high score shows a high perturbation of the auditory signal, and is associated with a lower voice quality.

**Pitch Matching Analysis.** The vocal recordings (both targets and matching responses) were analyzed using a Matlab (The MathWorks Inc., Natick, Massachusetts, USA) implementation of YIN (de Cheveigné & Kawahara, 2002). These analyses provided information about frequency, amplitude, and aperiodicity at a rate of 1378 Hz, and the pitch information was converted to cents relative to the target pitch. As per the instructions to the participant, only the final response in each trial was evaluated for pitch, but we did compile the number of responses the participant chose to make in each trial.

Pitch matching accuracy was measured in three ways. First, we measured the signed pitch error, which measures how sharp or flat the response was. Second, we took the absolute value of the pitch error, to avoid sharp and flat errors from cancelling each other out in averaging (but see Pfordresher & Mantell, 2009, for a useful alternative to this). Finally, we measured accuracy by taking the percentage of final responses within 50 cents of the target. This criterion was chosen because it is the point at which non-musicians tend to notice tuning inaccuracies when comparing two vocal tones (Hutchins & Peretz, 2012; Hutchins, Roquet, & Peretz, 2012).

**Melodic singing analysis.** The song "Happy Birthday" is composed of 25 notes, with one syllable per note. Data processing was done in two stages, using AudioSculpt and OpenMusic (Ircam, Paris, France). The acoustical analyses were based on the extraction of the  $F_0$  from the stable middle part of each note (as determined by the experimenter); the onsets and offsets of each note were omitted. Immediately repeated pitches (four in total, each "-py" of "Happy" in this song) were not considered in this analysis, as they were typically very short in duration.

We measured pitch accuracy in "Happy Birthday" in two ways, based off of the methods presented in Larrouy-Maestri & Morsomme (in press). First, we calculated the mean interval error as the mean difference between the target interval and the produced interval (unsigned). For example, the first interval, between "Happy" and "Birth", should be two semitones, 200 cents. If the participant actually sang an interval 250 cents, that would result in an interval error of 50 cents. Second, we calculated the tonal drift in the same way, but ignored all but the most tonally stable notes. Here, only the first note of the first three phrases (i.e. the first syllable of "happy"), and the final note of the song were analyzed (the three target intervals were adjusted accordingly, yielding 0, 0, and +500 cents. This provides a measure of how accurately the singers returned to the tonally stable notes, with higher scores indicating that the singer had a greater tendency to allow these to drift over the course of the melody.

## Results

### *Slider-voice comparisons*

Participants were considered to be poor at a pitch matching task if they had a mean absolute pitch error of greater than 50 cents within one response modality (see Hutchins & Peretz, 2012, Hutchins, Roquet & Peretz, 2012 for more on the use of this criterion). Five participants (23%) reached this threshold in the vocal pitch-matching task, but only one (5%) did so in the slider pitch-matching task (who was not one of the five who failed the vocal task). Comparisons were carried out between results from slider and voice pitch-matching tasks for each of the three pitch measurements using three separate paired  $t$  tests. We also tested the correlations of each measurement between slider and voice. On average, participants had lower absolute pitch errors and higher percentage of accurate answers in the slider condition (Mean error = 13 cents,  $SE = 0.29$  cents, 98% accuracy) than in the vocal condition (Mean error = 38 cents,  $SE = 1.03$  cents, 86% accuracy),  $t(21) = 2.54, p = .02, d = 0.54$  (error measurement),  $t(21) = 2.86, p = .009, d = 0.62$  (percent accuracy measurement). Participants were able to match a pitch more accurately on the slider than with the voice. There was no significant correlation between slider and voice pitch matching abilities, with either measurement.

The signed pitch error, however, showed the opposite pattern of results. Here, there was no difference in the tendency to produce flat or sharp errors between slider ( $M = 2.01$  cents,  $SE = 0.24$ ) and voice ( $M = 0.41$  cents,  $SE = 0.92$ ) conditions,  $t(21) = .19, n.s.$  Neither modality had a significant tendency to elicit sharp or flat errors,  $t(21) = 1.10, n.s.$  (slider),  $t(21) = 0.05, p = n.s.$  (voice). However, with the signed errors, there was a moderate but significant correlation between the slider and the voice modalities with the signed pitch error measurement,  $r(20) = .45, p = .04$ , indicating that participants who showed the tendency to be sharp on the slider also tended to be sharp in their sung responses.

We also compared the average number of responses that participants chose to make using the slider and their voice during each trial. As in Hutchins and Peretz (2012), participants made more

responses in the slider condition ( $M = 10.73$ ,  $SE = 0.15$ ) than in the singing condition ( $M = 1.05$ ,  $SE = 0.01$ ),  $t(21) = 8.20$ ,  $p = .002$ ,  $d = 1.74$ . Despite being allowed to correct their responses freely, participants generally chose to respond only once in the singing condition, even though they were less accurate in this condition.

## Vocal quality

To assess whether the quality of the sung tone varied systematically across targets, we conducted three separate one-way repeated measures ANOVAs using the factor of Target Height (five levels), with the measurements of jitter, shimmer, and pitch stability as dependent variables (shown in **Table 1**, degrees of freedom are reported using Greenhouse-Geisser corrections). We found main effects of jitter,  $F(2.67, 55.93) = 18.58$ ,  $p < .001$ ,  $\eta_p^2 = .47$  and shimmer,  $F(2.44, 48.95) = 10.61$ ,  $p < .001$ ,  $\eta_p^2 = .37$  such that both tended to decrease (indicating better voice quality) when the pitch increased. There was no main effect of Target Height of the tone on pitch stability. Post hoc comparisons using the Tukey HSD test indicated significant differences in jitter between targets 1 (the lowest) and 5 (the highest), 1 and 4, 1 and 3, 2 and 5, and 2 and 4 ( $HSD > 0.18$ ,  $p < .05$ ), and differences in shimmer between targets 1 and 4, 1 and 5, and 2 and 5 ( $HSD > 1.49$ ,  $p < .05$ ). There was also a significant correlation between jitter and shimmer across the five target tones,  $r(20) = .49$ ,  $p = .02$ , but no correlations between pitch stability and either of the other two measurements.

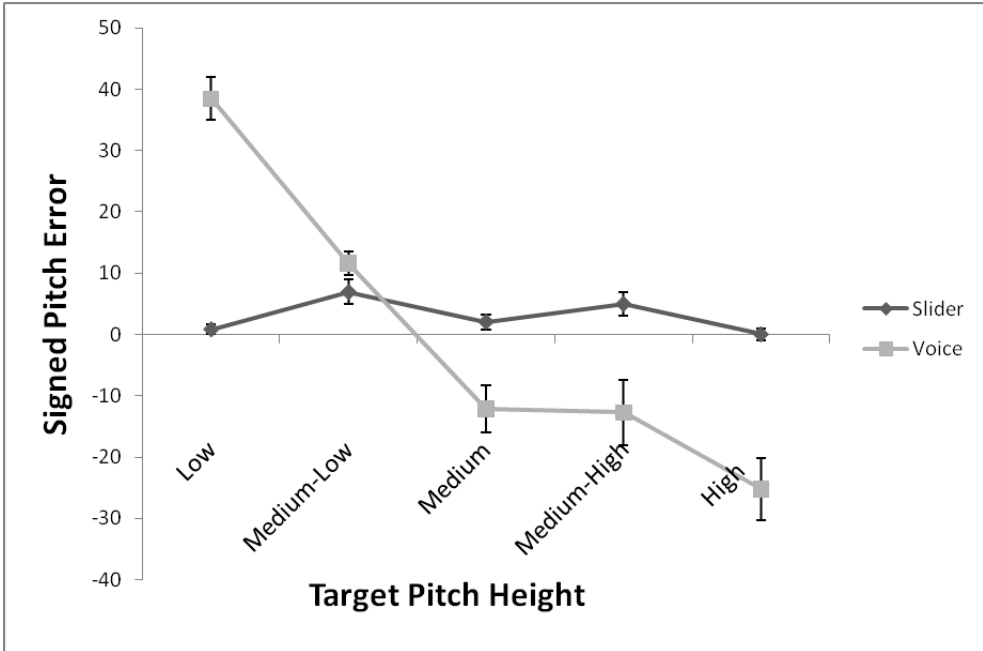
	Target Pitch Height				
	Low	Medium-Low	Medium	Medium-High	High
$F_0SD$	9.01	10.13	12.40	8.29	11.16
Jitter	0.72	0.56	0.46	0.37	0.33
Shimmer	4.44	3.93	3.16	2.59	2.24

**Table 1.** The mean of each vocal quality measurement (shimmer, jitter, and pitch stability [ $F_0SD$ ]), across the five target pitch heights.

## Target pitch height

To measure the effect of the height of the target tone on the ability to match that tone in each pitch-matching task, we conducted two one-

way repeated measures ANOVAs separately for slider and voice response modalities, across the five levels of Target Height. Target Height category (low, mid-low, medium, mid-high, and high) was used as the independent variable in this analysis, to allow comparison across participants with different vocal ranges (otherwise, this analysis would be dominated by the difference between males and females; see Figure 1 for the range of actual targets produced). Signed pitch error was used as the dependent variable, as flat or sharp errors were relevant across varying target pitches (see Figure 2). This measurement was different among targets in the voice condition,  $F(1.50, 31.45) = 4.60, p = .03, \eta_p^2 = .18$ . Participants were more likely to sing flat when matching high target notes, and more likely to sing sharp when matching low target notes, tending to err towards the middle of their vocal range. A Tukey post hoc test showed no significantly different pairs ( $HSD > 79.25, p < .05$ ), due mainly to the large sphericity correction, which increased the minimum significant difference by 69%. However, there was no difference in the ability to match the different target tones in the slider condition,  $F(2.46, 51.69) = .86, n.s.$ , which indicates that the vocal errors were due to vocal constraints, rather than perceptual constraints.

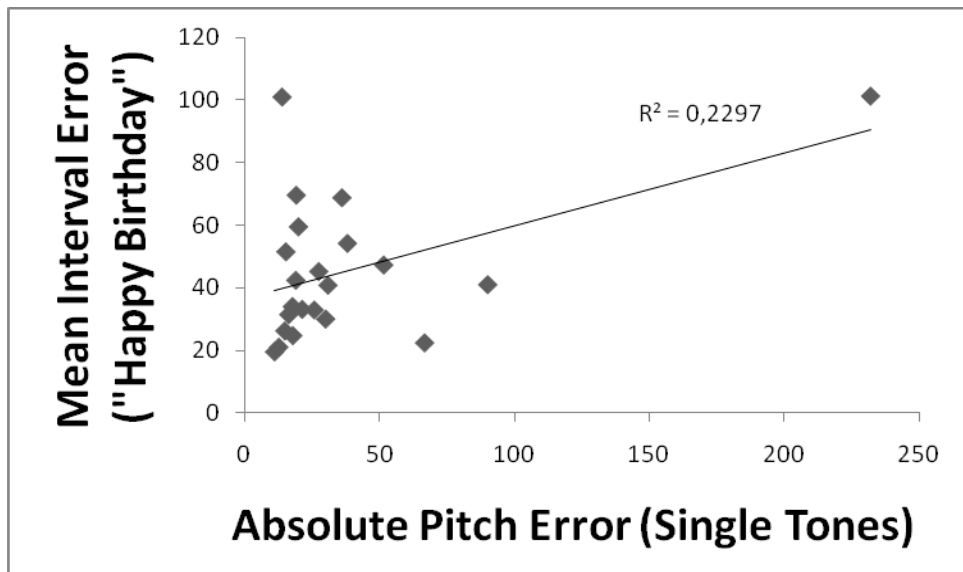


**Figure 2.** Signed errors for voice and slider conditions across the five different target heights, with standard error bars.



## ***Melodic singing ability and its relationship to other tasks***

To compare error between the melodic and vocal pitch matching tasks, we used the mean interval error and the absolute pitch error, respectively, as they are the most comparable measurements (both have unsigned cents as a unit). Mean error was not significantly different between the melodic ( $M = 45$  cents,  $SE = 4.92$ ), and vocal pitch matching conditions ( $M = 38$  cents,  $SE = 1.03$ ). These two measurements were correlated,  $r(20) = .48$ ,  $p = .02$  (see Figure 3). This correlation dropped slightly below the significance threshold when outliers were removed,  $r(16) = .44$ ,  $p = .07$ . There was no significant correlation between melodic singing ability and any measurement of slider pitch matching ability, nor was there a correlation between any measurements of single pitch matching ability and tonal drift in melodic singing.



**Figure 3.** A scatterplot of mean absolute error in the vocal pitch matching condition and mean interval error in the melodic singing condition. Note that the correlation drops to  $r = .44$ ,  $p = .07$  when outliers are removed

To see if there is a relationship between singing ability and vocal quality, we correlated our measurements of pitch matching ability and melodic singing ability with the vocal quality measurements (jitter, shimmer, and pitch stability). We found a significant correlation between the mean interval error during “Happy Birthday” and the pitch

stability ( $F_0SD$ ) of the sung target tones,  $r(20) = .43, p = .05$ . Participants with more stable individual tones also produced more accurate melodic intervals, on average. There was also a significant correlation between the tonal drift (measuring the long-term stability of melodic singing) and the target tone jitter,  $r(20) = .49, p = .02$ , and a non-significant trend towards a correlation of tonal drift and target tone shimmer,  $r(20) = .39, p = .07$ . In general, participants with a higher voice quality tended to sing Happy Birthday more accurately. There were no significant correlations between any vocal quality measurements and slider or vocal pitch matching abilities.

As a follow-up to this, we also computed the jitter and shimmer for singers during their productions of Happy Birthday (averaged across all tones). We found that singers' average jitter during Happy Birthday correlated with their melodic interval accuracy,  $r(20) = .46, p = .03$ , and with their tonal drift,  $r(20) = .50, p = .02$ . Singers' average shimmer during Happy Birthday correlated with the absolute error during the vocal pitch matching task,  $r(20) = .47, p = .03$ ; however, this correlation seems to have been driven entirely by outliers, and disappeared when they were removed,  $r(16) = .26, n.s.$  Interestingly, although the correlation between jitter during the Happy Birthday and in the single tones approached significance,  $r(20) = .40, p = .06$ , there was no correlation whatsoever between shimmer in these two conditions,  $r(20) = -.002, n.s.$

## Discussion

Despite the equivalent timbres in the slider and vocal pitch matching conditions, nonmusicians were still significantly better at matching pitches on the slider than with their voice. Participants were much more likely to make errors with their voice than on the slider, and there was no correlation between participants' average error in each modality. If perceptual problems were a common cause of poor singing, we would expect to see pitch errors on the slider that were similar in size to those in the voice, and correlations between these two measurements. On the contrary, our results confirm that poor singers are able to match the same pitches when using a different motor effector (the finger, in this case). Because there is only one timbre throughout

the experiment, for targets and responses, the proximal cause of poor singing here is likely to be poor vocal-motor control (see Hutchins & Peretz, 2012). In addition, the rate of poor singers here (23%) was similar to that found to have a vocal-motor control problem in Hutchins and Peretz (2012) (20%).

Addition evidence for the role of vocal-motor control can also be found in the effects of target pitch height. Although they were asked to produce target tones in a comfortable range, participants nevertheless showed a significant tendency to sing flat when imitating higher-pitched targets, and sharp when imitating lower-pitched targets. There was no similar tendency in responses on the slider, ruling out a perceptually-based explanation for this finding. This pattern (which was not present in previous self-matching paradigms, Hutchins & Peretz, 2012) indicates a clear role of vocal-motor control in singing errors- participants had trouble consistently reaching tones which were toward the extremities of their range. The analysis of vocal quality of the self-produced target tones (shown in Table 1) also shows an effect of vocal-motor control on target quality; lower-pitch tones were consistently of lower vocal quality. However, this lower quality did not seem to affect the ability to accurately perceive the pitch of these tones, as evidenced by the consistent, accurate pitch matching ability across all targets in the slider.

### ***Scaling up to melodies***

Another key aim of this study was to find evidence for the utility of pitch matching tasks in evaluating general singing ability. Although it has been assumed that the ability to match individual pitches is a foundational skill in singing, it had not yet been verified that this specific ability is related to the ability to sing whole melodies in tune. Our results showed a significant correlation between the average amount of error in singing melodies and in vocally matching single tones. Nonmusicians who were better at vocal pitch matching were also better at singing melodies in tune. Furthermore, there was no correlation between pitch matching ability on the slider and melodic singing ability, indicating that vocal single pitch matching is measuring a singing-specific ability, rather than only a general musical ability. This provides experimental evidence that measuring vocal pitch matching ability is a useful tool to gauge

overall singing ability. However, it should also be noted that this correlation was not one to one; there is still a good deal of variance in melodic singing ability unaccounted for in single pitch matching, some of which may concern tonality. Larrouy-Maestri et al. (2013) showed that, together, intervallic error and tonal drift could account for 81% of the variance in experts' judgments of singing ability; our results showing that single pitch matching error can predict intervallic error but not tonal drift lend quantitative confirmation to these expert judgments.

Our results also showed a significant correlation between melodic singing ability and vocal quality. People who were better at singing melodies in tune tended to have better vocal quality and sing individual tones more stably. Interestingly, no measurements of vocal quality had any significant correlations with vocal pitch matching ability. Although this is only correlative evidence, this seems to indicate that the relationship between vocal quality and in-tune melodic singing is not mediated by vocal pitch matching ability; there is an aspect of melodic singing which is influenced by vocal timbre separately from single pitch matching. It is also important to note that, because the experimenter selected the tones to use as targets partially on the basis of vocal quality, these represent the upper ranges of quality for each singer.

There are several explanations for the relationship between vocal quality and the ability to sing in tune. First, it seems somewhat likely that those with better vocal quality are also more likely to sing more often; the extra practice may lead to better overall singing ability and better vocal quality. Those who practice singing melodies more often may not practice single pitch imitation to the same extent, which could be a reason for the lack of correlation between vocal quality and single pitch matching. Another explanation is that a poor singing quality could lead people to misinterpret the actual pitch they are producing within a melodic framework. Differences in timbre (in this case between their actual production and an ideal melodic template) can cause more difficulty in pitch discrimination (Krumhansl & Iverson, 1992; Melara & Marks, 1990a, 1990b, 1990c). A poor vocal timbre might cause singers to be less sure if they were singing a correct pitch within a melody, and the resulting adjustments could cause successive melodic intervals to be out of tune. This would also explain why the effects were limited to interval errors, rather than tonality errors, which would be less

susceptible to this type of misperception, given the greater strength of the mental representation of the tonic tone (Krumhansl, 1990).

It should also be noted that the aspect of vocal quality measured here (pitch stability, jitter, and shimmer) represent only some of the possible measurements of vocal quality and timbre. Of these, both pitch stability and jitter, which are different measurements of short-term fluctuation in pitch, are strongly related to the mean pitch. Although these do not directly affect mean pitch in single tone or melodic contexts, it is less surprising that there would be relationships between these variables (the same does not hold for shimmer, on the other hand, which is a measurement of short-term fluctuation in amplitude). These measurements were chosen because they represent a few standard variables used in assessing vocal quality, but other measurements (e.g. breathiness, spectral centroid, spectral envelope, etc.) may reveal a different pattern of effects. In addition, because we measured the vocal quality of only one vowel (/a/), there is the possibility that vocal quality of any given subject may be markedly different in different vowels. Nevertheless, exploring the relationship between other aspects of timbre and singing ability should only serve to strengthen the relationship between these two abilities shown here- this would seem to be a fruitful avenue for future exploration.

## Conclusion

By using a touch-based measurement of peoples' sensitivity to pitch variations of their own voice, we were able to provide clear evidence that timbral factors are not responsible for the difference between pitch perception and production abilities. Rather, it is increasingly clearer that vocal-motor problems are a primary cause for many types of singing problems. In addition, we have shown that the types of single-tone pitch matching tasks used here and in many previous studies are a good proxy for melodic singing ability, and shown that differences in timbre and vocal quality can also be associated with the ability to sing melodies in tune.

## **Acknowledgments**

This work was supported by grants from the Canadian Institutes of Health Research, the Natural Sciences and Engineering Research Council of Canada and a Canada Research Chair in neurocognition of music to I. P., and a travel grant from the French Community of Belgium to P.L.-M. We would also like to thank Dominique Morsomme, who kindly supported the work of P.L.-M. on this paper.

# ETUDE 4

## The effects of stress on singing voice accuracy

---

**Larrouy-Maestri, P. & Morsomme, D.**

Logopédie de la Voix, Department of psychology: Cognition and Behavior, University of Liège, Belgium

In press in the Journal of Voice

### **Abstract**

**Objective:** The quality of a music performance can be lessened or enhanced if the performer experiences stressful conditions. In addition, the quality of a sung performance requires control of the fundamental frequency of the voice, which is particularly sensitive to stress. The present study aimed to clarify the effects of stress on singing voice accuracy.

**Methods:** Thirty-one music students were recorded in a stressful condition (i.e., a music examination) and a non-stressful condition. Two groups were defined according to the challenge level of the music examination (first and second music levels). Measurements were made by self-reported state anxiety (CSAI-2R questionnaire) and by observing heart rate activity (electrocardiogram) during each performance. In addition, the vocal accuracy of the sung performances was objectively analyzed.

**Results:** As expected, state anxiety and heart rate were significantly higher on the day of the music examination than in the non-stressful condition for all the music students. However, the effect of

stress was positive for the first-year students but negative for the second-year students, for whom the music examination was particularly challenging. In addition, highly significant correlations were found between the intensity of cognitive symptoms and the vocal accuracy criteria.

Discussion: This study highlights the contrasting effects of stress on singing voice accuracy but also the need to consider the challenge level and perception of the symptoms in experimental and pedagogical settings.

## Introduction

Singing voice accuracy depends on variations in fundamental frequency ( $F_0$ ) during a sung performance (Berkowska & Dalla Bella, 2009; Dalla Bella & Berkowska, 2009; Dalla Bella, Giguère, & Peretz, 2007; Larrouy-Maestri & Morsomme, in press; Pfordresher & Brown, 2009; Pfordresher, Brown, Meier, Belyk, & Liotti, 2010). In a melodic context, the relation between the tones (i.e., musical intervals) and the tonal center of the melody must be respected if the performer is to be considered as singing in tune. Indeed, two kinds of errors are considered in the judgment of vocal accuracy: the precision of the musical intervals and the respect of the tonality of the tune (Larrouy-Maestri, Lévêque, Schön, Giovanni, & Morsomme, 2013). The former is assessed by measuring the difference between each interval produced and the theoretical interval given by the musical score (Berkowska & Dalla Bella, 2009; Dalla Bella & Berkowska, 2009; Dalla Bella et al., 2007; Larrouy-Maestri et al., 2013; Larrouy-Maestri & Morsomme, in press; Pfordresher & Brown, 2009; Pfordresher et al., 2010). The latter is based on the consistency of the harmonically important notes performed in the course of the tune (Larrouy-Maestri & Morsomme, in press). In order to obtain low scores for these criteria (i.e., a high accuracy level), multiple components are required (Dalla Bella, Berkowska, & Sowinski, 2011; Hutchins & Peretz, 2012; Pfordresher & Brown, 2007). Indeed, the perceptual, motor, sensorimotor and memory components must all function properly to control the  $F_0$  variations in the sung performance and thus to sing in tune. Note also that there is a relationship between singing voice accuracy and the tempo of the



performance for both untrained singers (the slower, the more accurate) (Dalla Bella & Berkowska, 2009; Dalla Bella et al., 2007) and professional singers (the faster, the more accurate) (Larrouy-Maestri & Morsomme, in press).

Most studies of singing voice accuracy have observed untrained singers in an experimental context but this condition does not reflect stage performance, which can be stressful. For instance, performing in public or under pressure (e.g., in an examination) is reported to be stressful for musicians (Kenny, 2011), particularly in a Western classical context (Papageorgi, Creech, & Welch, 2013; Iusca & Dafinoiu, 2012). Some authors use the names “stage fright” and “music performance anxiety” interchangeably to describe the consequences of stress (Brotsky, 1996; Papageorgi, Hallam, & Welch, 2007; Salmon, 1990), whereas others distinguish these terms depending on the severity of the stress level (Fehm & Schmidt, 2006). The somatic and cognitive symptoms induced by a stressful situation and their interpretation by the performer can be observed by using questionnaires. For example, the performer may report perceived physiological arousal or negative thoughts and interpret them as facilitative or debilitating. Objective measurements of physiological manifestations of stress can also be observed (Abel & Larkin, 1990; Brotons, 1994; Craske & Craig, 1984; Fredrikson & Gunnarsson, 1992; Hamann & Sobaje, 1983; Kenny, 2011; LeBlanc, Chang Jin, Obert, & Siivola, 1997; Yoshie, Kudo, & Ohtsuki, 2008; Yoshie, Shigemasu, Kudo, & Ohtsuki, 2008). Indeed, the heart rate is dramatically higher during a stressful condition such as a competition than during rehearsals (Abel & Larkin, 1990; Brotons, 1994; Craske & Craig, 1984; Fredrikson & Gunnarsson, 1992; Hamann & Sobaje, 1983; LeBlanc, Chang Jin, Obert, & Siivola, 1997; Yoshie, Kudo, & Ohtsuki, 2008; Yoshie, Shigemasu, Kudo, & Ohtsuki, 2008).

As heart rate variations have a significant influence on the  $F_0$  of the speaking voice and of sustained voice productions (Bermudez de Alvear, Baron-Lopez, Alguacil, & Dawid-Milner, in press; Giddens, Barron, Byrd-Craven, Clark, & Winter, 2013), we hypothesize that a stressful condition should have an impact on vocal accuracy. However, this impact could be positive or negative. Studies of the speaking voice have shown that  $F_0$  increases with stress (Giddens et al., 2013; Hagenaaers & van Minnen, 2005; Ruiz, Absil, Harmegnies, Legros, & Poch, 1996; Scherer, 1977) but

vocal accuracy in a melodic context depends on the relation between the notes and cannot be limited to the increase in  $F_0$ . An objective analysis of the vocal accuracy criteria would be necessary to observe the positive or negative effect of stress on the accuracy of a vocal performance. More generally, the relationship between physiological arousal and performance has been investigated since the inverted-U hypothesis of Yerkes and Dodson (1908). According to Yerkes and Dodson's law (1908), performance increases with physiological or mental arousal and decreases when the level of arousal becomes too high. More recently, Yoshie, Kudo, Murakoshi, and Ohtsuki (2009) reported a negative change in performance quality among pianists when they performed in a stressful condition compared to without pressure. Their results did not confirm the findings of previous studies that reported no significant differences in performance quality between evaluative and non-evaluative conditions (Craske & Craig, 1984; Yoshie, Kudo, & Ohtsuki, 2008), or even an improvement under jury conditions (Hamann & Sobaje, 1983). These contrasting findings may be explained by a lack of control over stress levels, in terms of level or type. A combination of physiological and psychological measurements is necessary to evaluate the stress level manifestations. In addition, the difference in results could depend on the challenge level of the stressful condition for the participants. In past studies, college-level music students (Hamann & Sobaje, 1983), competent (Craske & Craig, 1984), skilled amateurs (Yoshie, Kudo, & Ohtsuki, 2008), and highly trained pianists (Yoshie et al., 2009) were observed in non-stressful and stressful conditions but the implications of the context of the musical performance (i.e., crucial examination or just part of the musical training) were not examined (Craske & Craig, 1984; Hamann & Sobaje, 1983; Yoshie et al., 2008, 2009).

In order to clarify the effects of stress on sung performances, the present study observed the physiological and psychological manifestations induced by a stressful situation and their effects on singing voice accuracy in music students at two levels (i.e., low and high challenge levels). For this purpose, we analyzed the sung performances of first- and second-year students in a stressful (music examination) and a non-stressful condition. Our aim was (a) to confirm that a music examination represents a stressful condition through the observation of

heart rate activity and self-reported state anxiety, (b) to observe the changes in the vocal performances by objectively analyzing the vocal accuracy of each recording, and (c) to compare two music levels (low versus high challenge levels) in terms of the changes in stress and vocal accuracy.

## Methods

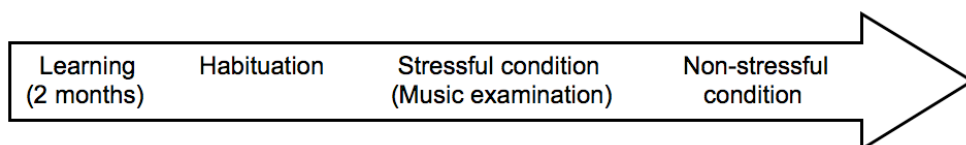
### *Participants*

Thirty-one music students (19 men, 12 women) from the Royal Conservatories of Belgium made up the sample. Their age ranged from 14 to 24 ( $M = 19.29$ ). In order to be accepted into these institutions, music students have a singing audition, in addition to their instrumental audition. Their ability to read and sing musical scores was thus assured. The solfeggio classes are held during the first two years of the music program (i.e., first and second levels). Note that the two solfeggio examinations are similar, with the same jury, but that the second-level examination is more important as it represents the end of the solfeggio classes, which is a condition for continuing in the music program. Eighteen participants were in their first year (aged from 14 to 24 years old,  $M = 19.22$ ) and 13 were in their second year (aged from 16 to 23 years old,  $M = 19.38$ ). Participants were instructed about the goal of the study, received an information form, signed the consent form and knew that they could stop their participation at any time.

### *Procedure*

The collaboration with the Royal Conservatories and the support of the teaching staff meant we could have students learn a melody during the solfeggio classes, once a week for two months (Learning phase in Figure 1) before the first recording. The performance was then recorded several times, with three weeks between sessions (see Figure 1). “Habituation” corresponds to a condition which allowed music students to get acquainted with the protocol, the experimenter and the equipment. Participants were instructed to sing the melody they had learned (Figure 2) a cappella, after hearing the first note played on a

piano by the experimenter. As indicated on the score, they had to perform with repeats, at a tempo of 80 beats per minute. The melody was sung with the syllable /no/ in order to avoid hesitations due to the spelling of the words and to provide a clear auditory signal to segment and analyze. In the stressful condition, the participants had to sing the melody in front of a jury composed of four music experts: two solfeggio teachers, one musician with experience in solfeggio evaluation and the pianist who accompanied the singing examination. The melody was performed at the beginning of the solfeggio examination. In the non-stressful condition, participants sang the melody in front of the experimenter in a quiet room at the Conservatories. This condition was similar to the “habituation” condition.



**Figure 1.** Illustration of the experimental procedure.

## ***Material***

### **Audio data**

The melody the students learned (Figure 2) was inspired by the popular song “Happy Birthday” and created in order to use the criteria and analysis tools validated by Larrouy-Maestri and Morsomme (in press). Audio was picked up by a Sennheiser HS2 head-worn microphone (Wedemark, Germany) positioned at a constant distance of 2 cm from the right corner of the mouth and recorded on a Marantz PMD67 recorder (Kanagawa, Japan). The sung performances were digitalized with a sampling frequency of 44.1 kHz and a 16-bit resolution.



**Figure 2.** Musical score learned, recorded and analyzed. The melodic line is composed of different intervals from the second minor to the octave, with a tonal center particularly marked in F Major. The rhythm is composed of quarter notes and half notes (crotchets are considered as ornaments). Two indications are visible: the repeat and the tempo (80 beats per minute).

### Stress level evaluation

*Objective measurements: electrocardiogram.* Participants wore a finger sensor, linked to a DATEX OHMEDA S/5 monitor (Madison, USA) connected to a ThinkPark IBM laptop. The heart rate (beats per minute) was recorded in the habituation, stressful and non-stressful conditions (Figure 1). In order to observe the changes in heart rate between the two conditions, we selected the non-stressful condition as a baseline and observed the difference between the stressful condition and the baseline for each participant.

*Subjective measurements: questionnaire.* Cognitive and somatic symptoms were assessed just before the examination (i.e., the stressful condition) and in the non-stressful condition (Figure 1), with the Competitive State Anxiety Inventory – 2 Revised (CSAI-2R) questionnaire (Cox, Martens, & Russell, 2003), translated and validated in French (Martinent, Ferrand, Guillet, & Gauthier, 2010). Although this material was designed to assess state anxiety in sports competitions, it has also been used for music performance (Iusca & Dafinoiu, 2012; Yoshie, Shigemasu, Kudo, & Ohtsuki, 2008). Because of the limited time available for students to fill in the questionnaire, we selected 11 items, which correspond to the two subscales designed to investigate somatic and cognitive symptoms. Participants had to rate the intensity of each symptom on a scale of 1 (not at all) to 4 (very much) at the time of the performance. Thus, the questionnaire was administered immediately

before each sung performance. For the direction scale, the participants rated on a scale ranging from -3 to +3 the degree to which the experienced intensity of each symptom was either debilitating or facilitative, with 0 indicating that it was considered to be unimportant. A positive score represents a state of facilitation and a negative score a state of debilitation. In order to observe the changes in anxiety symptoms between the different conditions, we selected the non-stressful condition as the baseline and observed the difference between the stressful condition and the baseline for each participant.

## ***Evaluation of sung performances***

### **Acoustical analyses**

The data were processed semi-automatically in two stages on a MacBook Pro (Mac OS X, Version 10.6.5). Analyses were done with AudioSculpt 2.9.4v3 and OpenMusic 6.3 software (Ircam, Paris, France) using a Short-Time Fourier Transform (STFT) analysis. Following the analytical procedure of Larrouy-Maestri and Morsomme (in press), markers were manually placed on the spectrogram in order to avoid the attacks and the glides between notes and minimize the influence of musical interpretation of the melody. Then, the average  $F_0$  was automatically calculated and converted into cents (1/2 tone = 100 cents) for the 18 notes selected from the melody (**Figure 2**).

### **Vocal accuracy criteria**

Concerning the intonation of the melody, the measurements were based on  $F_0$  differences between the tones in an equal temperament (i.e., equality of the semitones). For each sung performance, the interval deviation and tonal center deviation were examined according to the procedure described by Larrouy-Maestri and Morsomme (in press). For the interval deviation criterion, each interval was computed by subtracting the  $F_0$  of the adjacent notes and comparing the values to the theoretical ones (i.e., on the basis of the musical notation). The final score corresponds to the mean of the deviations between the performed intervals and the theoretical ones. To determine the tonal center deviation criterion, the intervals performed between eight notes, which convey important information about the tonality of the melody (i.e.,

notes 1, 4, 7, 14, 15, 16, 17, 18 in Figure 2), were compared to the theoretical value given by the musical notation. As for the interval deviation, the final score is measured in cents. Note that a small deviation reflected good interval precision and good respect for the tonal center of the melody. In order to observe the differences between vocal accuracy in the different conditions, we selected the non-stressful condition as the baseline and observed the difference between the stressful condition and the baseline for each participant.

### **Tempo**

The mean tempo was computed on the basis of the length of each performance and the number of beats of the melody (beats per minute).

## ***Statistical analyses***

Given the small number of participants at each music level, non-parametric tests were chosen to compare the two conditions (stressful and non-stressful). A Wilcoxon signed-rank test was used to assess differences in the variables under study (physiological, psychological and vocal accuracy) between the two conditions. This non-parametric test was run for each music level separately, as well as for the two groups combined in a single sample. Note that three participants performed the melody with contour errors (the direction of the intervals was not correct). As we focused on singing voice accuracy, these three participants were removed from the statistical analyses. Therefore, we had 16 participants at the first level and 12 at the second level.

In order to observe the changes in the variables (physiological, psychological and vocal accuracy) between the non-stressful and stressful conditions, we selected the non-stressful condition as the baseline and observed the differences between the stressful condition and the baseline for each participant. A Mann-Whitney U test was run to look for a significant difference in the evolution of these variables between the two groups of participants.

Finally, Spearman correlation matrices were computed for each music level, in order to evaluate the relationships between the changes in stress manifestations (heart rate, intensity and direction of cognitive

and somatic symptoms) and those in vocal parameters (tempo, interval deviation and tonal center deviation).

## Results

### *Evaluation of stress level*

Table 1 shows the descriptive statistics of the stress measurements (heart rate, CSAI-2R questionnaire) separately for each music level.

--

**Table 1.** Mean and standard error (in parentheses) of the stress measurements (heart rate, CSAI-2R questionnaire) for each music level (first and second), in the stressful and non-stressful conditions.

As can be seen in Table 1, the objective and subjective measurements of stress were higher on the day of the examination (i.e., stressful condition) than in the non-stressful condition.

--

**Table 2.** Results of the Wilcoxon signed-rank test for the comparison between the stressful and non-stressful conditions, for each music level (first and second).

For each music level, the Wilcoxon signed-rank test showed statistically significant differences in the heart rate measurement, as well as on the questionnaire (intensity of the somatic symptoms, intensity and direction of the cognitive symptoms). As Table 2 shows,



the direction of somatic symptoms was not statistically different between the stressful and the non-stressful conditions for participants at the first music level.

			Descriptive Statistics		Comparison	
			First level	Second level	U	p-value
<b>Heart rate (bpm)</b>			17.02 (6.41)	32.01 (6.74)	62.00	.114
<b>CSAI-2R</b>	<b>Somatic symptoms</b>	Intensity	10.00 (1.38)	11.11 (1.91)	87.50	.687
		Direction	-4.06 (3.67)	-14.86 (3.29)	40.50	.010
	<b>Cognitive symptoms</b>	Intensity	14.37 (1.80)	10.33 (2.10)	63.50	.129
		Direction	-10.50 (3.73)	-15.83 (4.83)	78.50	.416

**Table 3.** Mean and standard error (in parentheses) for the changes in the stress measurements (heart rate, CSAI-2R questionnaire) between the non-stressful and the stressful conditions, for each music level (first and second).

Differences between the two music levels were computed with the Mann-Whitney U test.

The changes between the non-stressful and stressful conditions for each music level are summarized in Table 3, together with the results of the Mann-Whitney U test and the p-values. No statistically significant difference occurred between the first and second music levels, except for the direction of the somatic symptoms. For the second-year students, the change in the perception of somatic symptoms between the non-stressful and stressful conditions was greater and more negative than for the first-year students.

### ***Evaluation of sung performances***

*Interval deviation criterion.* As can be seen in Figure 3, the first-year students performed better in the stressful condition ( $M = 17.05$ ,  $SE = 1.45$ ) than in the non-stressful condition ( $M = 28.63$ ,  $SE = 5.87$ ). The Wilcoxon signed-rank test confirmed this difference ( $z = -2.36$ ,  $p = .019$ ). For the second-year students, no significant difference appeared between the stressful ( $M = 39.35$ ,  $SE = 8.54$ ) and non-stressful ( $M = 27.53$ ,  $SE = 8.14$ ) conditions for this criterion ( $z = -1.18$ ,  $p = .24$ ).



**Figure 3.** Illustration of the mean scores for the interval deviation criterion (in cents). The scores for the two music levels (first and second) are represented for the two conditions (stressful and non-stressful). Note that a low score represents a low deviation and thus a better accuracy.

*Tonal center criterion.* As can be seen in Figure 4, the second-year students performed better in the non-stressful condition ( $M = 19.67$ ,  $SE = 3.17$ ) than in the stressful condition ( $M = 40.61$ ,  $SE = 11.39$ ). The Wilcoxon test confirmed this difference ( $z = -2.59$ ,  $p = .01$ ). In the case of the first-year students, no significant difference appeared between the stressful ( $M = 20.06$ ,  $SE = 2.45$ ) and non-stressful ( $M = 38.65$ ,  $SE = 11.33$ ) conditions for this criterion ( $z = -0.74$ ,  $p = .46$ ).



**Figure 4.** Illustration of the mean scores for the tonal center deviation criterion (in cents). The scores for the two music levels (first and second) are represented for the two conditions (stressful and non-stressful). Note that a low score represents a low deviation and thus better accuracy.

The different profiles of the music students, according to their level, were confirmed with the Mann-Whitney U test. Indeed, the changes between the non-stressful and stressful conditions differed according to the music level for the interval deviation criterion ( $U = 50.50$ ,  $p = .035$ ) and the tonal center criterion ( $U = 48.00$ ,  $p = .026$ ). Whereas the stressful condition was positive for the first-year students in terms of the interval deviation criterion, this condition was negative for the second-year students' performance, as measured by the tonal center deviation criterion.

### ***Relationships between stress level and vocal accuracy***

For each variable under study, we computed the change between the stressful condition and the baseline (i.e., non-stressful condition) for each participant and observed the relationships between the variables. For each music level, no correlation was found between the objective measurement of stress (i.e., heart rate) and singing voice accuracy (interval deviation and tonal center deviation). For students at the first

music level, a significant correlation was found between the interval deviation criterion and the intensity of cognitive symptoms ( $r(16) = .520, p = .039$ ), and also the direction of these symptoms ( $r(16) = .608, p = .013$ ). In other words, the more intense the cognitive symptoms were, and the more negatively they were interpreted, the worse the precision of the intervals was along the melody. Note that no significant correlation was found between the interval deviation criterion and the intensity of somatic symptoms ( $r(16) = .297, p = .264$ ) or the direction of these symptoms ( $r(16) = .438, p = .089$ ). For students at the second music level, a significant correlation was found between the tonal center deviation criterion and the intensity of cognitive symptoms ( $r(12) = .773, p = .003$ ). For these students, the more intense the cognitive symptoms were, the worse their respect of the tonal center became. However, there was no significant relationship between the interpretation of cognitive symptoms (facilitative or debilitating) and vocal accuracy in this group of music students ( $r(12) = .452, p = .140$ ). Note that there was no significant correlation between the tonal center deviation and the intensity of somatic symptoms ( $r(12) = .340, p = .279$ ) or the direction of these symptoms ( $r(12) = .218, p = .495$ ). Concerning tempo, there was no significant correlation with stress measurements or vocal accuracy, at either music level.

## Discussion

In order to observe the effects, positive or negative, of stress on singing voice accuracy, we analyzed the sung performances of students at two music levels in a stressful and a non-stressful condition. These results have both pedagogical and experimental implications.

### *Changes in stress level*

Our results confirmed objectively that musicians are more likely to feel anxious in conditions in which they are being evaluated (Kenny, 2011). As expected, the heart rate and the intensity of cognitive and somatic symptoms were significantly higher in the stressful condition than in the non-stressful situation, for students at both music levels. In addition, the cognitive and somatic symptoms were experienced as more negative in the stressful condition for both groups. However,

members of the two groups differ significantly in the manner in which they interpret their somatic symptoms. In fact, this change was significantly higher for the second-year students. Note that the solfeggio classes are given during the first two years of the music program at the Royal Conservatories of Belgium. The experimental conditions were similar for students at both music levels (i.e., same musical score, same habituation, and same jury) and the main difference concerned the level of the challenge. Indeed, whereas second-year students must pass this examination in order to continue their musical education, the first-year students have the possibility of continuing their education regardless of their results on the first examination. This study revealed that the heart rate of first- and second-year students was faster during the music examination than during the non-stressful condition, with a similar evolution between the two groups. However, the difference observed in the interpretation of somatic symptoms (larger negative evolution for the second-year students) could be attributed to the challenge level of the situation. In addition to this explanation, one can hypothesize that the kind of stress perceived by the first- and second-year students differs. It would therefore be interesting to compare different kinds of stressful situations (e.g., concerts, competitions). Also, we focused on the physiological and psychological manifestations induced by a music examination, but these observations would have to be completed in future researches. For example, the self-confidence subscale of the CSAI-2R or trait anxiety questionnaires (Elwood, Wolitzky-Taylor, & Olatunji, 2012; Rossi & Pourtois, 2012) could be proposed to develop the findings of the present study.

However, the present experimental design provides a pertinent and ecological context for examining the effect of stress on the quality of a music performance. In addition, the results regarding the interpretation of somatic symptoms confirmed the relevance of taking challenge level into account in the observation of differences in singing voice accuracy.

### ***Changes in singing voice accuracy***

Two important vocal accuracy criteria were observed for the participants whose performance respected the melodic contour ( $n =$

28). Previous studies highlighted the impact of stress on  $F_0$  in the speaking voice (Giddens et al., 2013; Hagedaars & van Minnen, 2005; Ruiz et al., 1996; Scherer, 1977) and on the quality of music performance (Craske & Craig, 1984; Hamann & Sobaje, 1983; Yoshie et al., 2009; Yoshie et al., 2008; Kokotsaki & Davidson, 2003). Because vocal accuracy is based on  $F_0$  variations and must be respected in a musical performance, it is not surprising that stress has an impact on vocal accuracy. However, the effect was different depending on the music level of the participants: positive for the first-year students (i.e., greater interval precision under stress) and negative for the second-year students (i.e., tonal center less respected under stress). The results also pointed out the relationship between vocal accuracy and the psychological manifestations of stress. Interestingly, no correlation was found between the changes in heart rate or somatic symptoms and the quality of the sung performances. However, the precision of the intervals was linked with the intensity and direction of the cognitive symptoms of the first-year students whereas the intensity of these symptoms was linked to the tonal center criterion for the second-year students. For students at each music level, high-intensity cognitive symptoms had a negative impact on their vocal accuracy. These findings confirm that the relationship hypothesized by Yerkes and Dodson (1908) cannot be limited to the observation of physiological manifestations. Note that these manifestations might not be strong enough to have consequences for the vocal quality. We therefore encourage future researchers to induce more stress, for example, with the presence of a larger jury or audience and/or more difficult musical material in order to confirm that physiological manifestations of stress are not linked to vocal accuracy. In addition to the psychological manifestations (i.e., intensity of the cognitive symptoms), the challenge level must be considered. Indeed, two profiles appeared regarding the participants' musical level, with stress having a positive effect on the first-year students and a negative effect on the second-year students. Although the results have to be interpreted with caution because of the limited number of participants in each group (16 first-year students and 12 second-year students), they show that a high challenge level leads to a greater negative evolution of the perception of somatic symptoms between the two conditions. This could explain the differences between

the two groups in this study and the contrasted findings of previous studies of the impact of stress on the quality of a music performance (Craske & Craig, 1984; Hagenars & van Minnen, 2005; Hamann & Sobaje, 1983; Kokotsaki & Davidson, 2003; Ruiz et al., 1996; Scherer, 1977; Yoshie et al., 2009; Yoshie et al., 2008).

## ***Implications***

### **Educational applications**

The ecological but controlled conditions of the present study have direct implications for music education. As interval precision and respect of the tonal center of a melody are linked to the intensity of cognitive symptoms, it would be interesting to focus on the prevention of these symptoms in solfeggio classes, for instance, by providing information about coping strategies in conservatories' education programs. More generally, a more distanced approach to the challenge would help to generate a positive effect of stress on music performance (i.e., as observed for the first-year students). The second-year students interpreted their somatic symptoms much more negatively than the first-year students and the results showed that stress had a negative effect on them. In addition to information on coping strategies, this study suggests that music institutions should organize examinations on a more regular basis in order to reduce the challenge level (and its negative effect) of the last year of the music program.

### **Implications for singing research**

The present findings highlight the effect of stress – which can be positive or negative – on singing voice accuracy and thus the relevance of including emotional state among the multiple components required to sing in tune (Dalla Bella et al., 2011; Hutchins & Peretz, 2012; Pfordresher & Brown, 2007). In addition, we observed a dissociation between two criteria that are relevant in the evaluation of singing voice accuracy (i.e., interval deviation and tonal center deviation) (Larrouy-Maestri et al., 2013). In other words, this study indicates that emotional state and the dissociation between the vocal accuracy criteria should be taken into account in theoretical models explaining singing voice accuracy.

Finally, the relationship between tempo and vocal accuracy, observed with both occasional singers (Dalla Bella & Berkowska, 2009; Dalla Bella et al., 2007) and trained singers (Larrouy-Maestri & Morsomme, in press), is not visible when participants are instrumentalists (i.e., people who have a high level of musical knowledge but not much training in vocal production). Consequently, participants' music level should be taken into account in future investigations of singing voice accuracy.

## **Conclusions**

By observing physiological, psychological and acoustical parameters in music students in stressful and non-stressful conditions, this study shows the effects of stress on singing voice accuracy. These effects can be either positive or negative, depending on the challenge level of the music examination. In addition to the pedagogical implications of these findings, this study provides relevant information about the singing voice accuracy process and the effects of stress on a sung performance.

## **Acknowledgments**

The authors thank the Centre Henri Pousseur in Liège, Guillaume Videlier, and David Magis for technical support. We also thank Céline Clijsters for her help with the data collection and the Royal Conservatories of Belgium for their collaboration.



# ETUDE 5

## The effect of melody and technique on the singing voice accuracy of trained singers

---

**Pauline Larrouy-Maestri<sup>1</sup>, David Magis<sup>2</sup>, & Dominique Morsomme<sup>1</sup>**

1. Logopédie de la Voix, Department of psychology: Cognition and Behaviour, University of Liège, Belgium

2. Department of Education, University of Liège, Belgium

In press in Logopedics Phoniatrics Vocology

### **Abstract**

A previous study highlighted the effect of vocal technique on the singing voice accuracy of trained singers (Larrouy-Maestri & Morsomme, in press). The intervals' precision between the notes of the tune was altered when the singers used Western operatic singing technique. In order to better understand these results, we have recorded two different melodies sung with two different vocal techniques. A large panel of trained singers (N = 50) participated in the study. The analytical method described in the reference paper has been applied. The results confirm the effect of vocal technique on the vocal accuracy of trained singers. In addition, these results provide an answer about the melodic effect and guide future work on the perception process of operatic voices.

### **Introduction**

In a melodic context, the vocal accuracy is observed through the fundamental frequency variations in the tune and has been shown as an

important parameter to assess the quality of a sung performance (Watts, Barnes-Burroughs, Andrianopoulos, & Carr, 2003). The most common criterion used in objective assessment of a vocal performance is the pitch interval deviation, estimated by measuring the difference between the size of intervals performed and the standard size based on the musical notation (Berkowska & Dalla Bella, 2009; Dalla Bella & Berkowska, 2009; Dalla Bella, Giguère, & Peretz, 2007; Pfordresher & Brown, 2009; Pfordresher, Brown, Meier, Belyk, & Liotti, 2010).

In a recent study (Larrouy-Maestri & Morsomme, *in press*), the authors have compared the sung performances of 63 untrained singers and 14 trained singers of the popular song “Happy Birthday”. As expected, the professional singers were more accurate than the untrained ones. However, when a Western operatic singing technique was used by the trained singers, the size of the intervals was particularly deviated from the musical notation. The effect of the technique observed in this study could be discussed regarding the small number of participants, but also regarding the interpretation effect of the trained singers. Indeed, the simple and familiar melody “Happy Birthday” is not part of the classical vocal repertory. An over-interpretation may be the cause of the inaccuracy of the trained singers who would exaggerate the melodic contour of the song and add ornaments. From roughly the first decade of the 19th century to the middle 20th century, the romantic era music was characterized by increased attention to the expressivity of the melodic line (e.g. Samson, 2001). Within the stylistic characteristics of this musical period, one can note complex modulations, tonal ambiguity with harmonic tensions and surprising cadences, warm and expressive melodies as well as ornaments with *appoggiatura*, trill, turn, mordent or *arpeggio* (e.g. Hanning, 2011). The transfer of these stylistic characteristics on a simple popular song would consequently alter the intervals’ precision and thus this vocal accuracy criterion. Indeed, the singers would be less strict as far as the precision of intervals is concerned in order to make it sound like a Western romantic song.

The present study aimed (a) to confirm the effect of singing technique observed in the previous study and (b) to observe if the choice of the melody has an effect on the vocal accuracy. For (a), a large group of trained singers was instructed to perform the Happy Birthday song with two different vocal techniques (“naturally” and with a

Western operatic singing technique). For (b), the same participants were asked to select and to sing a romantic melody of their musical repertory, with two contrasting singing techniques.

## **Methods**

### ***Participants***

A total of 50 professional or semi-professional singers (38 women, 12 men) from 19 to 66 years old ( $M = 36.94$  years) participated in the study. They all had an education in classical music and regularly engaged in classical solo vocal performances. They began singing lessons between 6 and 49 years of age ( $M = 20.18$  years) and they had between 5 and 51 years of singing experience ( $M = 19.86$  years). They reported to practice their vocal instrument 13.68 hours/week on average.

### ***Material and procedure***

The sound recordings were made using a head-worn microphone Sennheiser HS2 (Wedemark, Germany) positioned at a constant distance of 2 cm from the right corner of the mouth as well as a Marantz Professional Solid State Recorder PMD67 (Kanagawa, Japan). All these sung performances have been listed in a database, which can be viewed following the link: <http://sldr.org/sldr000792/en>.

As in the previous study, participants were first asked to produce two vocal glissandi to warm up the vocal organs (Amir, Amir, & Michaeli, 2005) and to encourage a lack of inhibition in front of the experimenter and the recording equipment. Then, they individually performed, a cappella, two different melodies (the popular song "Happy Birthday" and a romantic melody of their choice) with 2 different techniques (singing without any particular technique and with a Western operatic singing technique). Concerning the chosen melody, the participants had to choose a romantic melody of their musical repertory (i.e. Lieder of Schubert, Schumann). Except for the chosen melody, no particular starting note was given in order to let the participant perform in his/her comfortable range.

The four vocal performances were recorded following a precise order: the French version of the popular birthday melody “naturally” (i.e. whilst imagining a festive and friendly context) (*birthday*), the chosen melody “naturally” (*romantic*), the chosen melody with a Western operatic singing technique “like on stage” (*ROMANTIC*) and finally the birthday song with a Western operatic singing technique (*BIRTHDAY*).

### ***Acoustic analysis***

For the four performances of each participant (*birthday*, *romantic*, *ROMANTIC*, *BIRTHDAY*), the pitch interval deviation was computed following the analytical procedure of Larrouy-Maestri and Morsomme (in press). In order to observe the respect of the tonality of the tune, we have chosen to compute the number of tonality modulations. Indeed, the tonal centre was not always clear because of the length of the musical excerpts. Also, the number of tonality modulations appeared as a better criterion to observe the respect of the tonality of the tune than the tonal centre deviation used in the reference paper (Larrouy-Maestri & Morsomme, in press). Note that, in order to compare the chosen melodies and the popular song (containing 21 notes analysed), the excerpts selected contained from 14 to 28 notes ( $M = 20.76$ ,  $ES = 3.61$ ). One tonality modulation corresponds to an interval error of more than 100 cents (one semitone), which is not compensated by an error in the other direction (for details, see Larrouy-Maestri, Lévêque, Schön, Giovanni, & Morsomme, 2013). All the acoustic analyses were run on a Macintosh (Mac OS X, Version 10.6.8).

### ***Statistical analysis***

The pitch interval deviation and the number of tonality modulations were modelled separately by means of a multiple regression model, with the type of melody and the technique as covariates. An interaction effect between the melody and the technique was also included in the initial model. The model was then simplified by removing all not statistically significant tests, using the likelihood-ratio test for nested models (Agresti, 1996). The significance level was fixed

to 5%. The simplest model was retained and discussed for further investigation of statistically significant trends.

## Results

For each of the four conditions (2 melodies X 2 techniques), the pitch interval deviation and the number of tonality modulations criteria for the 50 singers are represented in Table 1.

Var.	Stat.	Popular song		Chosen melody	
		birthday	BIRTHDAY	romantic	ROMANTIC
Intervals	Mean	28.788	37.606	34.307	42.568
	SE	3.364	4.136	3.093	2.769
Modulations	Mean	0.240	0.360	0.500	0.540
	SE	0.089	0.103	0.109	0.126

**Table 1.** Descriptive analyses of the pitch interval deviation (in cents) and the number of tonality modulations for the four sung performances.

Table 2 displays a summary of the p-values for both main effects (melody and technique) to each response variable modelling. Note that the interaction between melody and technique variables was never statistically significant and is therefore skipped from this table.

Variables	Melody	Technique
Intervals	0.118	0.012
Modulations	0.040	0.453

**Table 2.** P-values of the effects of the melody and the technique on the variables.

The effect of the technique is significant for the pitch interval deviation criterion: the Western operatic singing technique yields an increase of the pitch interval deviation ( $p = 0.012$ ). However, this vocal technique does not disturb the tonal centre of the tune ( $p = 0.453$ ). The effect of the melody is statistically significant only for the number of tonality modulations ( $p = 0.040$ ). It can be observed from the final model that singing the chosen melody leads to more tonality modulations than in the popular song. Moreover, as can be seen in Table

2, no effect of melody occurred for the pitch interval deviation criterion ( $p = 0.118$ ).

## Discussion and conclusion

This study confirms the findings of Larrouy-Maestri and Morsomme (in press) concerning the effect of technique on the singing voice accuracy (see Table 2). Indeed, among the 50 trained singers, the use of Western operatic singing technique leads to a high pitch interval deviation in the sung performance. We also found that the respect of the tonal centre (number of tonality modulations) did not vary according to the singing technique. However, this criterion was sensible to the melodic effect. The number of tonality modulations was significantly more important for the chosen melody than for the popular song. This result is not surprising regarding the size of the excerpts selected. For experimental reasons, we wished to observe two different melodies containing an equivalent number of notes: one popular song (21 notes) and one from their musical repertory. The participants were thus asked to choose excerpts in accordance with the musical notation (respect of phrases and articulation) but also with an equivalent length as the popular song. This selection allowed comparing the two melodies but the harmonic component of the romantic melody was less clear as it was not performed in its complete version. The trained singers showed more difficulties to respect the tonal centre in the excerpt from the romantic melody than in the full popular song.

The fact that the singers were particularly out of tune (high pitch interval deviation) when they sang with a Western operatic singing technique was already surprising in the previous study. We pointed out that an over-interpretation could explain the inaccuracy of the trained singers by the exaggeration of the melodic contour of the song and the addition of ornaments. Thanks to the analysis of two different melodies (popular song and romantic melody chosen by each participant), we observed no effect of the melody on the singing voice accuracy (see Table 2). Indeed, for both melodies, the singers were more out of tune when they used the Western operatic singing technique than when singing "naturally". The hypothesis of an over-interpretation of the popular song, which is not adapted to the daily practice of the trained

singers, has not been confirmed. For future researches on trained singers, the choice of the melody would not have consequences on the vocal accuracy.

Finally, regarding the profile of the trained singers (low interval precision when the Western operatic singing technique is employed), the accuracy of the pitch intervals is perhaps less important when this vocal technique is employed than for untrained singers or trained singers singing “naturally”. This criterion is commonly used in experimental researches (Berkowska & Dalla Bella, 2009; Dalla Bella & Berkowska, 2009; Dalla Bella et al., 2007; Pfordresher & Brown, 2009; Pfordresher et al., 2010) and explains a large part of music experts’ evaluation in a context of untrained singers (Larrouy-Maestri et al., 2013). However, listeners could be willing to accept pitch alterations if it adds or is completed by other components. This question about pitch importance could be a relevant subject for further investigation, for example by the confrontation of the objective evaluation of the sung performances of this study with music experts’ rating.

## **Acknowledgements**

The authors thank the “Centre Henri Pousseur” of Liège and Guillaume Videlier for the technical support of acoustic analyzes. We also thank the singers who took part in this experiment and Marion Nowak for her help with the data collection.





# ETUDE 6

## Effects of melody and technique on acoustical and musical features of Western operatic singing voices

---

**Pauline Larrouy-Maestri\*, David Magis<sup>x</sup>, & Dominique Morsomme\***

\* Department of psychology: Cognition and Behavior, University of  
Liège, Belgium

<sup>x</sup> Department of Education, University of Liège, Belgium

Submitted in the Journal of Voice

### **Abstract**

The operatic singing technique is frequently employed in classical music. Several acoustical parameters of this specific technique have been studied but how these parameters combine remains unclear. This study aims to further characterize the Western operatic singing technique by observing the effects of melody and technique on acoustical and musical parameters of the singing voice. Fifty professional singers performed two contrasting melodies (popular song and romantic melody) with two vocal techniques (with and without operatic singing technique). The common quality parameters (energy distribution, vibrato rate and extent), perturbation parameters (standard deviation of the fundamental frequency, signal-to-noise ratio, jitter and shimmer) and musical features (fundamental frequency of the starting note, average tempo, and sound pressure level) of the 200 sung performances were analyzed. The results regarding the effect of melody and technique on the acoustical and musical parameters show that the

choice of melody had a limited impact on the parameters observed, whereas a particular vocal profile appeared depending on the vocal technique employed. This study confirms that vocal technique affects the majority of the parameters examined. In addition, the observation of quality, perturbation and musical parameters contributes to a better understanding of the operatic singing technique.

## Introduction

The vocal technique used in singing differs depending on musical style. The Western operatic singing technique is employed by classically trained singers. Such singing requires a highly refined use of physiological components such as the laryngeal, respiratory and articulatory muscles (Collyer, Kenny, & Archer, 2011; Lam Tang, Boliek, & Rieger, 2008; Sataloff, 2005; Sundberg, 2009; Thalén & Sundberg, 2001; Thorpe, Cala, Chapman, & Davis, 2001). These components lead to a complex combination of acoustical parameters (see Sundberg, 2013, for a review), which differs from natural or untrained singing (Brown, Rothman, & Sapienza, 2000; Hunter, Svec, & Titze, 2006; Mendes, Rothman, Sapienza, & Brown, 2003).

The concept of good vocal quality is naturally very difficult to define as it consists of various components. Among the acoustical parameters observed in the operatic singing technique, particular attention has been paid to vibrato and energy distribution.

Vocal vibrato has been the subject of several acoustical and perceptual studies since the beginning of the twentieth century, and is highly correlated with overall vocal beauty (Ekholm, Papagiannis, & Chagnon, 1998; Rothman, Rullman, & Arroyo, 1990). It can be defined as a slightly tremulous effect imparted to vocal tone and corresponds to a quasi-periodic modulation of the frequency of a tone. Vibrato is mainly characterized by two parameters: rate and extent. Vibrato rate specifies the number of frequency fluctuations per second and ranges between 5 and 7 Hz depending on the singer (Prame, 1994). Observing the variability in vibrato rates across 10 prominent artists singing in the Western classical music tradition, Prame (1994) found an average rate of 6 Hz, with a mean variation from the singer's average of 8%. More recently, the values were defined between 4.55 and 6.25 Hz for classical

voices (de Almeida Bezerra, Cukier-Blaj, Duprat, Camargo, & Granato, 2009) and between 6.28 and 7.14 Hz for operatic singing voices (Howes, Callaghan, Davis, Kenny, & Thorpe, 2004). However, Stone, Cleveland, Sundberg, and Prokop (2002) found a mean vibrato rate of 5.7 Hz with this particular technique, slower than for the Broadway style of singing (6.1 Hz). Note that mean vibrato rate is considered to be a constant for a singer (Shipp, Leanderson, & Sundberg, 1980). It does not depend on the context of solo or choir performance (Reid, Davis, Oates, Cabrera, Ternström, Black, & Chapman, 2007) but can be slightly modified deliberately (King & Horii, 1993) or by musical training (Mürbe, Zahnert, Kuhlisch, & Sundberg, 2007) or warming up (Moorcroft & Kenny, 2012). Vibrato extent describes how far the frequency fluctuates around the pitch of the tone. Unlike vibrato rate, vibrato extent varies considerably for individual singers (Prame, 1997) and between singers (Bretos & Sundberg, 2003). The variability pointed out by Seashore (1938) has been confirmed since then, with, for example, a mean extent of 71 cents according to Prame (1997), and a variation of between 0.38 and 3.26 semitones (Hirano, Hibi, & Hagino, 1995) or between 0.54 and 1.66 semitones (de Almeida Bezerra et al., 2009). Mean vibrato extent has been found to be greater for the operatic singing style ( $\pm 98$  cents) than the Broadway style ( $\pm 78$  cents) (Stone, Cleveland, Sundberg, & Prokop, 2002).

The long-term average spectrum (LTAS) provides information on the spectral distribution of the sound. Operatic singing is associated with an increase of energy of between 2 and 4 kHz (Barnes, Davis, Oates, & Chapman, 2004; Omori, Kacker, Carroll, Riley, & Blaugrund, 1996; Thorpe et al., 2001). This frequency range has been identified as being of prime importance for opera soloists' vocal projection (Bloothoof & Plomp, 1988; Sundberg, 2001) and is often associated with good quality, as it allows the audibility of the voice to be optimized against an orchestra. Note that energy distribution can be developed with training (Brown et al., 2000; Lundy, Roy, Casiano, Xue, & Evans, 2000; Omori et al., 1996) and depends on the sound level of the performance (Ferguson, Kenny, & Cabrera, 2010; Sundberg & Nordenberg, 2006) as well as on the vocal technique employed (Stone, Cleveland, Sundberg, & Prokop, 2002; Sundberg, Gramming, & Lovetri, 1993). Several parameters, such as the singing power ratio (Omori et al., 1996), the energy ratio (Thorpe

et al., 2001), the  $\alpha$  coefficient (Sundberg & Nordenberg, 2006) and the difference in energy between different frequency bands, have been proposed to quantify the spectral balance from an LTAS. While these measurements are not intended to provide information about the singer's formant, they illustrate differences in energy distribution.

In addition to these acoustical parameters describing the operatic singing technique (i.e., vibrato rate, extent and spectral distribution), Butte et al. (2009) pointed out the relevance of perturbation analysis for operatic voices. By comparing 26 recordings of different singers and singing styles, they confirmed that singing in the operatic style exhibits more "chaos" than other styles of singing. The high perturbation values could be due to vibrato, which leads to considerable pitch variability. They also highlighted the fact that the operatic singing style is less periodic than other styles of singing. Perturbation analysis (i.e., parameters such as jitter, shimmer, and signal-to-noise ratio), which is commonly used for the evaluation of voice disorders (Sataloff, 2005), is thus relevant for non-periodic voices (Lieberman, 1961, 1963; Zhang, Jiang, & Wallace, 2005) such as operatic voices.

As well as investigating the vocal and quality parameters of the operatic singing technique (i.e., vibrato rate and extent, spectral distribution and vocal perturbation), this study examined some musical characteristics such as the loudness of the voice, fundamental frequency of the starting note and tempo. In their comparison of the performance of a popular song with and without an operatic singing technique, Larrouy-Maestri and Morsomme (in press) observed a particularly slow tempo when the song was performed with an operatic singing technique. In order to observe the operatic singing technique, the present study investigated the effect of technique on the commonly examined acoustical parameters but also on vocal perturbation and the musical parameters mentioned above, as well as the relationship between the various parameters.

Previous studies had compared the physiological and acoustical parameters of different vocal techniques. They applied three kinds of methods: the comparison of a single professional singer in different musical styles (Borch & Sundberg, 2011; Evans & Howard, 1993; Stone et al., 2002; Sundberg, Gramming, & Lovetri, 1993; Sundberg & Thalén,

2010; Sundberg, Gu, Huang, & Huang, 2012; Thalén & Sundberg, 2001); the comparison of different groups specializing in one particular musical style (Björkner, 2008; de Almeida Bezerra et al., 2009; Estill, 1988; Schutte & Miller, 1993; Sundberg et al., 2012); or the examination of one melody performed by one group of participants with two different vocal techniques (Barlow & Lovetri, 2010; Barlow, Lovetri, & Howard, 2007; Lebowitz & Baken, 2011). However, all of these experimental designs have their limitations. First, an investigation that focuses on a single subject makes it difficult to generalize the effect of vocal technique and calls into question the general validity of the findings. Second, the use of different musical material, representing different musical styles, does not allow one to control for the effect of melody on the acoustical parameters of the vocal performances. Third, when the same musical material is performed by a group of participants, with different styles, only one melody is considered and the replicability of the findings for other melodies cannot be examined. To sum up, previous studies provided useful information on the Western operatic singing technique but further investigations, with more participants, the observation of several parameters, and the control of melodic effects, would allow for a better understanding of this particular singing technique.

The present study examined the effects of melody and technique on quality, perturbation and musical parameters among a large panel of Western professional singers. For this purpose, we analyzed the sung performances of singers performing two different melodies, with and without an operatic singing technique. This experimental design allows one to observe what is attributable to the musical material used versus to the technique employed and thus to develop a theoretical model clarifying this particular singing technique.

## **Methods**

### ***Participants***

Fifty singers (38 women, 12 men) between 19 and 66 years old ( $M = 36.94$  years) participated in the study. They had all received a classical music education in higher institutions such as music conservatories and regularly engaged in classical solo vocal performances. They began their

singing lessons between 6 and 49 years of age ( $M = 20.18$  years) and they had between 5 and 51 years of singing experience ( $M = 19.86$  years). They reported practicing with their vocal instrument for 13.68 hours/week on average.

## ***Material and procedure***

The sound recordings were made using a head-worn microphone (Sennheiser HS2, Wedemark, Germany) positioned at a constant distance of 2 cm from the right corner of the subjects' lips and a Marantz Professional Solid State PMD67 Recorder (Kanagawa, Japan). Note that the recording equipment was never removed during each participant's recording session in order to keep a constant distance between the mouth and the microphone.

Prior to taping, participants were asked to produce two glissandi to warm up their vocal organs and to encourage a lack of inhibition in front of the experimenter and the recording equipment. Then, each one produced four vocal performances: the popular song "Happy Birthday" without any particular technique, for example while imagining a festive and friendly context (birthday); a romantic melody of their choice without any particular technique (romantic); the same melody with a Western operatic singing technique (ROMANTIC); and finally, the birthday song with a Western operatic singing technique (BIRTHDAY). Concerning the romantic melody, the participants had to choose a melody in their musical repertory (e.g., Lieder of Schubert or Schumann). Except for the chosen melody, no particular starting note was given so each participant could perform in his/her comfortable range. The performances were sung a cappella to avoid altering the auditory signal with musical instrumentation and recorded with a sampling frequency of 44.1 kHz and a 16-bit resolution (i.e., sufficient to extract normal vocal perturbation measurements).

## ***Acoustical analysis***

For the 200 performances (birthday, romantic, ROMANTIC, BIRTHDAY from the 50 participants), ten variables were observed,

grouped into quality, perturbation and musical parameters. All the acoustical analyses were run on a Macintosh (Mac OS X, Version 10.6.8).

### **Quality parameters**

For the acoustic analyses, we extracted the longest note of each melody (the last note of “Happy Birthday” and the longest note of the romantic melody). The quality parameters (energy distribution, vibrato rate and extent) were analyzed with AudioSculpt and OpenMusic software (Ircam, Paris, France). The variable known as *Energy distribution* allows one to observe the spectral balance and thus the energy in the band containing the singer’s formant. This measure was computed in two steps. First, we separated the total frequency range into two bands following the procedure used in several studies of trained singers (e.g., Collyer et al., 2007, 2009; Fergusson et al., 2010; Thorpe et al., 2001). However, we adapted the frequency range according to the participants’ gender, considering that the female formant pattern is scaled upwards about 20% compared to the average male formant pattern. The two bands were 0–2 kHz and 2–4.5 kHz for the male participants and 0–2.4 kHz and 2.4–5.4 kHz for the female ones. The energy of the second band was then divided by the energy of the first band. Note that a high score for the Energy distribution variable shows a strong reinforcement of the band containing the singer’s formant.

The *Vibrato Rate (VR)* and *Vibrato Extent (VE)* were based on the fact that vibrato is essentially a periodic modulation of fundamental frequency, which is generally found to be closely sinusoidal (e.g., Prame, 1994, 1997). VR specifies the number of undulations per second and VE corresponds to the difference between the minimum and the maximum of the undulations (in cents).

### **Perturbation parameters**

Perturbation analysis was performed on the same selected note using Praat software (version 5.1.44, Boersma, 2001). We measured the *Standard Deviation of the fundamental Frequency (SDF<sub>0</sub>)* across the duration of the note, as an indication of the F<sub>0</sub> variations over a short period (in Hz). With the *Signal-to-Noise Ratio (SNR)*, the degree of acoustical periodicity in the vocal waveform (in dB) can be assessed.

The *Jitter* (a measurement of the period perturbation) and *Shimmer* (a measurement of the amplitude perturbation) were reported as percentages.

Note that for  $SDF_0$ , Jitter% and Shimmer%, a high score shows a high perturbation of the auditory signal and is associated with a low vocal quality.

### **Musical parameters**

The *fundamental Frequency* ( $F_0$ ) of the starting note was extracted using Praat software and was reported in Hertz (Hz). The average *Tempo* of the performance was calculated on the basis of its length, by dividing the number of pulsations by the duration of the performance (beats per minute). The *Sound level* (in dB) was also observed for the longest note of each melody using Praat software (version 5.1.44, Boersma, 2001).

### ***Statistical analysis***

Each variable was modeled separately by means of multiple regression models, with the type of melody and the technique as covariates. An interaction effect between the melody and the technique was also included in the initial model. Then, the model was simplified by removing all statistically non-significant tests, using the likelihood-ratio test for nested models (Agresti, 1996). The significance level was set at 5%. The final model was retained and discussed for further investigation of statistically significant trends.

In a complementary analysis, we attempted to observe the profile of the operatic singing technique by modeling it as the dependent variable, and by including all the acoustical and musical parameters as model covariates. Logistic regression modeling of the probability of making use of the Western operatic singing technique was done. The model included all acoustical and musical parameters as main effects only. The “strength” of the relationships between the parameters and the type of singing technique was summarized by the p-values derived from the logistic regression model. Note that the variables related to the participants’ age and musical experience were also included in the study. However, none of them exhibited statistically significant effects.



Therefore, they were not further considered in the analysis of the results.

## Results

Table 1 displays the means and standard deviations of the ten parameters analyzed for each of the four vocal performances (birthday, romantic, ROMANTIC, BIRTHDAY).

Parameters		Stat.	Sung performance			
			birthday	BIRTHDAY	romantic	ROMANTIC
Singing	F <sub>0</sub> (in Hz)	Mean	303	339	318	317
		SD	87.50	98.50	90.22	89.13
	Tempo (in beats per min)	Mean	93	76	80	73
		SD	20.39	16.31	33.50	31.28
	Sound level (in dB)	Mean	68.41	71.07	71.24	73.45
		SD	10.43	8.62	9.49	7.38
Quality	Energy distribution (ratio)	Mean	1.88	1.60	2.20	1.95
		SD	0.37	0.35	0.65	0.60
	Vibrato rate (in Hz)	Mean	5.87	5.95	5.75	5.95
		SD	1.23	0.83	0.72	0.82
	Vibrato extent (in cents)	Mean	96.84	161.36	117.56	178.92
		SD	59.68	65.46	62.37	69.72
Perturbation	SDF <sub>0</sub> (in Hz)	Mean	5.22	9.37	7.30	10.39
		SD	3.70	4.74	6.94	5.17
	SNR (in dB)	Mean	24.76	21.35	27.29	21.82
		SD	3,93	5,59	5,55	5,29
	Jitter (in%)	Mean	0.31	0.38	0.27	0.43
		SD	0.19	0.19	0.16	0.24
Shimmer (in%)	Mean	1.97	2.83	2.01	2.78	
	SD	0.91	2.54	1.53	1.83	

**Table 1.** Means and standard deviations (SD) of the ten parameters analyzed for each of the four vocal performances (i.e., the birthday and romantic melodies performed without and with an operatic singing technique).

The correlation matrix of the ten parameters was computed to observe the relationships between the acoustical and musical features observed for the 200 sung performances. As Table 2 shows, in addition to the significant correlations between the perturbation parameters (SDF<sub>0</sub>, SNR, Jitter, Shimmer), several significant correlations were found between these parameters and the F<sub>0</sub>, Energy distribution, Vibrato Extent and Sound level parameters. Regarding the relationships between the musical and quality parameters, the Energy of the band containing the singer's formant increased with the Sound level and the F<sub>0</sub> of the starting note, whereas the Vibrato Rate increased with the Tempo.

	1	2	3	4	5	6	7	8	9	10
1 F <sub>0</sub>	--									
2 Tempo	-.06	--								
3 Sound level	.07	.09	--							
4 Energy	<b>.14*</b>	-.05	<b>.59**</b>	--						
5 Vibrato rate	.10	<b>.22**</b>	-.11	-.11	--					
6 Vibrato extent	.03	-.10	.08	<b>-.14*</b>	<b>-.16*</b>	--				
7 SDF <sub>0</sub>	<b>.47**</b>	-.12	.07	-.04	-.09	<b>.72**</b>	--			
8 SNR	<b>.35**</b>	-.02	<b>.23*</b>	<b>.54**</b>	-.05	<b>-.61**</b>	<b>-.26**</b>	--		
9 Jitter	<b>-.32**</b>	.01	-.02	<b>-.25**</b>	-.03	<b>.59**</b>	<b>.29**</b>	<b>-.75**</b>	--	
10 Shimmer	<b>-.28**</b>	-.11	<b>-.36**</b>	<b>-.35**</b>	-.07	<b>.40**</b>	<b>.18*</b>	<b>-.73**</b>	<b>.63**</b>	--

\*  $p < .05$

\*\*  $p < .01$

**Table 2.** Observed correlation matrix of musical and acoustical parameters for the 200 sung performances.

### ***Effect of melody and technique on the singing parameters***

Table 3 displays a summary of the p-values for both main effects (melody and technique) in each response variable model. Note that the interaction between melody and technique never reached statistical significance and is therefore omitted from this table.

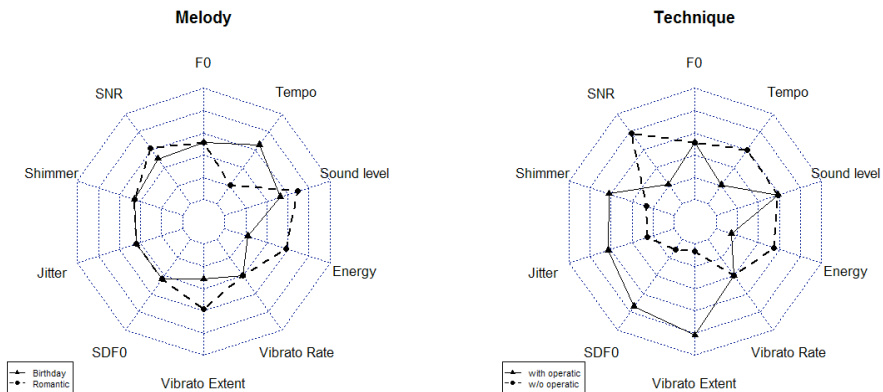
Parameters	Melody	Technique
$F_0$	0.807	0.175
Tempo	0.033	0.001
Sound level	0.043	0.058
Energy distribution	< 0.001	< 0.001
Vibrato rate	0.675	0.291
Vibrato extent	0.037	< 0.001
$SDF_0$	0.12	0.012
SNR	0.041	< 0.001
Jitter%	0.971	< 0.001
Shimmer%	0.991	0.001

**Table 3.** P-values of the melody and technique effects for each parameter.

The effect of melody is statistically significant only for SNR, Tempo, Sound level, Energy distribution and VE. The final models reveal that singing the popular song with or without an operatic singing technique led to a smaller SNR ( $p = 0.041$ ), a slower Tempo ( $p = 0.033$ ), a lower Sound level ( $p = 0.043$ ), a smaller Energy distribution ( $p < 0.001$ ) and a smaller VE ( $p = 0.037$ ).

The effect of technique, however, is significant for all variables but  $F_0$ , Sound level and VR. The response variables can be combined in two groups, as follows: the operatic singing technique yields an increase in  $SDF_0$  ( $p = 0.012$ ), Jitter% ( $p < 0.001$ ), Shimmer% ( $p = 0.001$ ) and VE ( $p < 0.001$ ) compared to the performance without any particular singing technique. Conversely, the operatic singing technique results in a decrease in SNR ( $p < 0.001$ ), Tempo ( $p = 0.001$ ) and Energy distribution ( $p < 0.001$ ) compared to the performance without any particular technique.

A complete visual representation is given in Figure 1. The ten parameters are displayed as extremes of the spider plots. The two melodies (left panel) and the two techniques (right panel) are represented by different points and trait types. Points that overlap refer to parameters without any statistically significant effect on the melody/technique.



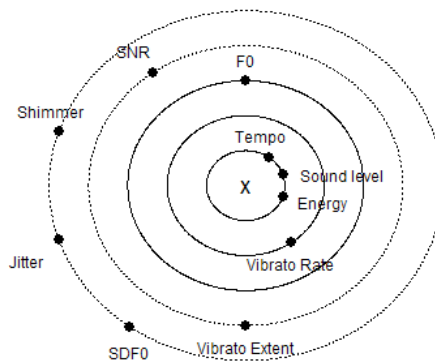
**Figure 1.** Illustration of significant and non-significant effects of melody (left) and technique (right) on the ten parameters observed. Each axis represents a parameter, with the minimum value at the center of the spider plot. Significant differences in melody (left) and technique (right) are displayed by point differences in the respective axes. Superimposed points refer to non-significant differences.

As illustrated in Figure 1, the choice of melody (“Happy Birthday” versus romantic melody) had fewer consequences for the parameters observed than the choice of technique (with versus without an operatic singing technique). Regarding the quality parameters usually observed (Energy distribution, VR, and VE), only vibrato rate was not affected by the effects of melody and technique. Vibrato extent was larger with the romantic melody and with an operatic singing technique, whereas energy distribution was higher with a romantic melody (i.e., strong reinforcement of the band containing the singer’s formant) but lower with an operatic singing technique. The greatest differences were found with the perturbation parameters. The technique effect is clearly visible for all the parameters ( $SDF_0$  and SNR, Shimmer, and Jitter), which showed a high perturbation of the signal when the singers were asked to sing with an operatic singing technique. Otherwise, the perturbation parameters did not differ greatly depending on the melody, and only the signal-to-noise ratio was significantly better when performing a romantic melody. Concerning the musical parameters, the  $F_0$  of the starting note did not differ significantly as a function of the melody performed or the technique employed. However, the tempo was slower with the romantic melody and with the operatic singing technique. The

sound level did not differ significantly depending on the technique but was higher with the romantic melody than with the birthday song.

### ***Profile of the operatic singing technique***

The results of the complementary analyses are displayed in Figure 2 as follows. P-values of the F-test of statistical significance of each parameter were merged into five categories, referring to the usual ranges of highly significant ( $]0, 0.001[$ ), very significant ( $[0.001, 0.01[$ ), significant ( $[0.01, 0.05[$ ), borderline ( $[0.05, 0.1[$ ) and non-significant ( $[0.1, 1[$ ) p-values. These five categories are displayed in a target plot, with the center of the target referring to a p-value of zero and each circle accounting for one category (circles further away from the center correspond to larger p-values). The ten parameters are then displayed on their respective circles (i.e., their respective categories of p-values) similarly to the spider plots in Figure 1.



**Figure 2.** Illustration of the degree of significance of the ten parameters observed to clarify the operatic technique. Each circle represents a range of p-values. Full circles refer to highly significant  $]0, 0.001[$ , very significant  $[0.001, 0.01[$ , and significant p-values  $[0.01, 0.05[$  (from smaller to larger circles).

Dotted circles refer to non-significant p-values.

As Figure 2 shows, five of the ten parameters observed (Energy distribution, Sound level, Tempo, VR and  $F_0$ ) seemed particularly important in the operatic singing technique profile. Note that p-values of the SNR and the VE parameters were not significant but borderline, which can be interpreted as indicating that they are moderately important in the operatic singing profile. Conversely, the other parameters (SDF $_0$ , Shimmer, and Jitter) do not appear to be significant predictors of this particular vocal technique.

## Discussion

The analysis of 50 singers performing two different melodies, with and without an operatic singing technique, provides a better understanding of this singing technique. Indeed, the experimental design of the present study allows us to observe the effects of melody and technique on acoustical and musical parameters. The correlation matrix confirms the relationships between the perturbation parameters and the link between musical and acoustical parameters. However, a closer examination of these parameters in a logistic regression model reveals the relevance of several specific parameters in explaining the Western operatic singing technique.

### *Effects of melody and technique*

#### **Quality parameters**

The observation of the vibrato rate in the four kinds of performances (two melodies with two vocal techniques) confirms its consistency for a given singer (Shipp et al., 1980). With a mean of 5.88 undulations per second, the vibrato rate observed here is similar to that in several previous studies (de Almeida Bezerra et al., 2009; Prame, 1994; Stone et al., 2002) but lower than what Howes et al., (2004) found. The results show that the melody or technique employed by the singer did not affect the vibrato rate, which confirms the participants' training level (Mürbe et al., 2007; Prame, 1994) and the consistency of this parameter regardless of the melody performed or the singing technique used by the singer. The values for vibrato extent, on the other hand, depended on the melody (higher in the romantic melody) and the

technique (higher with the operatic singing technique) used by the singer. Ranging between 14 and 370 cents ( $M = 139$  cents), these values confirm the considerable variability of vibrato extent (de Almeida Bezerra et al., 2009; Hirano et al., 1995; Prame, 1997; Seashore, 1938).

The melodic effect highlights the relevance of selecting melodies that are specific to the singing technique investigated in order to observe acoustical parameters in future studies. The melodic effect observed for the Energy distribution variable also points to the necessity of focusing on appropriate melodic material for observing the spectral balance of trained singers. Indeed, our result shows that the band containing the singer's formant is strongly reinforced when he or she performs a romantic melody. However, despite the participants' training level, we observed considerable variation in spectral balance. Other measurements, such as the singing power ratio (Omori et al., 1996), energy ratio (Thorpe et al., 2001), and  $\alpha$  coefficient (Sundberg & Nordenberg, 2006) could provide more information about this variability, as could controlling the sound pressure level.

### **Perturbation parameters**

As expected, the jitter and shimmer measurements were particularly high with the operatic singing technique, which confirms the aperiodicity of this technique (Butte et al., 2009). The other two perturbation parameters examined (SNR and  $SDF_0$ ) generated similar results and support the hypothesis that the operatic singing technique leads to a high perturbation in intensity, amplitude and frequency. Note that there was no melodic effect and the perturbation parameters were only sensitive to the technique employed. However, these findings should be compared with perceptual judgments. In fact, parameters such as jitter, shimmer, and signal-to-noise ratio have been shown to be relevant for non-periodic voices (Lieberman, 1961, 1963; Zhang et al., 2005), but the analytical tools may be limited; for example, a strong perturbation may lead to a failed  $F_0$  extraction in perturbation analysis, and thus provide irrelevant information (Dejonckere, Giordano, Schoentgen, Fraj, Bocchi, & Manfredi, 2012). Regarding the significant correlations between vibrato extent and the perturbation parameters, our results suggest the need to control for vibrato extent in future research on the aperiodicity of operatic singing voices.

### **Musical parameters**

The  $F_0$  of the starting note of the performance was not affected by melody or technique. Note that the participants had the opportunity to perform the two different melodies in their comfortable range (no starting note was given for “Happy Birthday” and the romantic melody was chosen by the participant), which explains the absence of melody effect. For the popular song, the  $F_0$  seemed higher when the participants performed with an operatic singing technique (Table 1). However, the technique effect is not shown in the statistical analyses (Table 3), as they included the chosen melody, for which the starting note was controlled. Conversely, we observed a different tempo depending on the melody (slower for the romantic melody) and the technique employed (slower with an operatic singing technique). Slow melodies thus seem to be representative of the romantic music era according to the participants, who chose the songs. Note that songs were sung more slowly with an operatic singing technique despite the tempo indication in the musical score.

Because vibrato is relevant in the operatic singing technique (see Sundberg, 2013, for a review) and takes place on vowels, trained singers may extend the duration of the vowels. It is therefore not surprising that the duration of the performance was longer with an operatic singing technique and when performing the romantic melody (i.e., representative of the classical repertory). The technique effect observed confirms Larrouy-Maestri and Morsomme’s (in press) findings with a larger panel of professional singers. In their study, they also found a relationship between tempo and singing voice accuracy when professional singers performed with an operatic singing technique (the faster they sang, the more accurate they were). As the present study highlights the effect of melody on the tempo, it seems relevant to observe similar melodies (in a singing research or music evaluation context) to control for this musical parameter and its consequences for the quality of the sung performance.

As for the loudness of the voice, the results show an effect of melody (louder with the romantic melody) but no effect of technique. The experimental design (each singer performed both melodies with both singing techniques, without removing the recording equipment)



allows us to observe the correlation between this parameter and the energy of the band containing the singer's formant, as expected (Ferguson, Kenny, & Cabrera, 2010; Sundberg & Nordenberg, 2006). However, this variable should be controlled in order to accurately define the energy distribution of the operatic singing technique.

### ***Profile of the operatic singing technique***

The results clearly show the effect of technique on several parameters. However, the analysis of the ten parameters in the logistic model identifies five parameters that explain the operatic singing technique: Sound level, Tempo, Energy distribution, Vibrato Rate and  $F_0$  of the starting note. Note that two parameters were marginally significant (Vibrato Extent and Signal-to-Noise Ratio). This findings have pedagogical implications and provide some guidelines for singing teachers. The relationships between these acoustical parameters and physiological components have been examined (Collyer et al., 2011; Lam Tang et al., 2008; Sataloff, 2005; Sundberg, 2009; Thalén & Sundberg, 2001; Thorpe et al., 2001) and music teachers could focus on the components linked to the parameters highlighted by the logistic model in this study (Figure 2). They could, for instance, focus on subglottal pressure to improve vocal projection and increase the energy in the band containing the singer's formant as well as the sound level of the performance. Because vibrato rate can be somewhat modified deliberately (King & Horii, 1993) by musical training (Mürbe et al., 2007) or by warming up (Moorcroft & Kenny, 2012), this parameter could be to the subject of specific training so singers can learn to reach a frequency around 5.95 Hz. In addition, although the tempo and the  $F_0$  of the starting note are given by the musical score, prolonging the vowels (i.e., slowing down the tempo) and choosing melodies in the high part of the singer's range would enhance the operatic character of a performance.

Still, this model needs to be developed and expanded before it can provide a precise definition of the operatic singing technique. For example, a larger number of participants would refine the present results. In addition, this study focused on semiprofessional or professional singers and the number of years of musical training was

shown to have no significance in the model. However, this pattern could vary as a function of other characteristics such as singing level or vocal ability. It would therefore be interesting to compare different singing levels, defined on the basis of Bunch and Chapman's (2000) taxonomy of singers or the phonetogram for singers (Lamarche, Ternström, & Pabon, 2010). Furthermore, the parameters examined in the present study are not exhaustive. Additional parameters such as the variation in vibrato rate and extent along the note (Ferrante, 2011), articulation (Sundberg, 2009), text intelligibility (Sundberg & Romedahl, 2009), vocal accuracy (Larrouy-Maestri & Morsomme, in press), formant tuning (Sundberg, La, & Gill, 2013) or other measurements of spectral distribution (Ferguson et al., 2010) could complete this framework. Finally, the statistically significant parameters describing the operatic singing voice could be observed in order to assess this singing technique objectively. However, this model would have to be compared with the perceptual judgment of expert judges in order to confirm the relevance of these parameters in classical music.

## Conclusion

Several parameters (quality, perturbation and musical) were examined in two different melodies sung with and without an operatic singing technique. The results confirm the effect of technique on the vibrato extent, energy distribution, perturbation parameters and tempo of the performance. By examining these parameters in a theoretical model, this study highlights the relevance of vibrato rate, sound level, energy distribution, fundamental frequency of the starting note and tempo in describing the Western operatic singing technique. Although this model needs to be further developed in future research, it has already generated some implications for research and teaching.

## Acknowledgments

The authors thank the Centre Henri Pousseur in Liège for technical support. We also thank Marion Nowak for her help with the data collection and the singers who allowed us to run this experiment.

# ETUDE 7

## The evaluation of vocal accuracy: The case of operatic singing voices

---

**Pauline Larrouy-Maestri\*, David Magis<sup>x</sup>, & Dominique Morsomme\***

\* Department of psychology: Cognition and Behavior, University of Liège, Belgium

<sup>x</sup> Department of Education, University of Liège, Belgium

Under review in Music Perception

### **Abstract**

The objective analysis of Western operatic singing voices indicates that professional singers can be particularly “out of tune”. This study aims to better understand the evaluation of operatic voices, which have particularly complex acoustical signals.

Twenty-two music experts were asked to evaluate the vocal accuracy of 14 sung performances with a pairwise comparison paradigm in a test and a retest. In addition to the objective measurement of vocal accuracy, several performance parameters (average tempo, fundamental frequency of the starting note) and quality parameters (energy distribution, vibrato rate and extent) were observed and compared to the judges’ perceptual rating.

The results show high intra- and inter-judge reliability when rating the vocal accuracy of operatic singing voices. Surprisingly, all the parameters were significantly related to the ratings and explain 78.8% of the variability of the judges’ evaluation. These findings moderate the importance of vocal accuracy when an operatic singing technique is

employed and provide information about the perception of operatic voices.

## Introduction

In classical music, singing in tune is key, as it enables singers to create expressive melodic lines and harmonize with other singers and instrumentalists. In addition, precision of intonation has been shown to be an important factor associated with the perception of singing talent (Watts, Barnes-Burroughs, Andrianopoulos, & Carr, 2003). Vocal accuracy, in terms of the precision of musical intervals, can be objectively measured by observing fundamental frequency variations in the course of the tune (Berkowska & Dalla Bella, 2009; Dalla Bella & Berkowska, 2009; Dalla Bella, Giguère, & Peretz, 2007; Larrouy-Maestri & Morsomme, in press; Pfordresher & Brown, 2009; Pfordresher, Brown, Meier, Belyk, & Liotti, 2010). A previous study found that objective and subjective measurements of the vocal accuracy of untrained singers are highly correlated (Larrouy-Maestri, Lévêque, Schön, Giovanni, & Morsomme, 2013). That study also reported that the objective measurements of vocal accuracy explained 81% of the judges' rating.

Interestingly, renowned professional singers are not always found to sing in tune when their performances are objectively measured (Sundberg, Prame, & Iwarsson, 1996; Vurma & Ross, 2006). Recently, Larrouy-Maestri and Morsomme (in press) compared the sung performances of 63 untrained singers and 14 trained singers of the popular song "Happy Birthday" with a computer-assisted method. The trained singers had to perform the melody twice, with and without a Western operatic singing technique. As expected, the professional singers were more accurate than the untrained ones. However, the professional singers manifested an effect of vocal technique: musical intervals were less precise when a Western operatic singing technique was employed than when it was not. This pattern was confirmed with a larger group of professional singers and two different melodic materials (Larrouy-Maestri, Magis, & Morsomme, in press). Indeed, the pitch interval deviation was greater with the operatic singing technique than without it, no matter which melody was performed (a popular song or a

melody from the romantic period). Given that professional musicians' voices, when measured objectively for accuracy, show low pitch interval deviation, this study aims to examine the perceptual judgment of the overall vocal accuracy of operatic singing voices.

The operatic singing technique involves several acoustical components such as vibrato and reinforcement of the energy band containing the singer's formant (Sundberg, 2013). Vibrato has been extensively investigated since Seashore (1938) and has often been observed among classical singers (e.g., Bretos & Sundberg, 2003; de Almeida, Cukier-Blaj, Duprat, Camargo, & Granato, 2009; Ekholm, Papagiannis, & Chagnon, 1998; Hirano, Hibi, & Hagino, 1995; Howes, Callaghan, Davis, Kenny, & Thorpe, 2004; Prame, 1994, 1997; Stone, Cleveland, Sundberg, & Prokop, 2002; Sundberg, 1994). It corresponds to a quasi-periodic modulation of the frequency of a tone and is mainly characterized by its rate (number of fluctuations per second) and its extent (amplitude of the fluctuations around the pitch of the tone). Regarding the spectral distribution, the operatic singing technique is associated with an increase in energy between 2 and 4 kHz (Barnes, Davis, Oates, & Chapman, 2004; Omori, Kacker, Carroll, Riley, & Blaugrund, 1996; Sundberg, 1995, 2001, 2013; Thorpe, Cala, Chapman, & Davis, 2001), which allows the voice to be heard above the sound of the orchestra. These two components (i.e., vibrato and reinforcement of the energy band containing the singer's formant) seem to be much appreciated by listeners. For instance, Garnier, Henrich, Castellengo, Sotiropoulos, and Dubois (2007) observed the large vocabulary used to describe the operatic singing voice and pointed out that vibrato and timbre are the acoustical parameters used by judges (with a consensus) to define a "good" vocal performance in Western operatic singing.

However, the acoustical features developed in operatic singing techniques can alter the waveform of the tone or the spectrum of the acoustical signal and thus affect how the pitch height of the voice is perceived. Indeed, the pitch perception of tones differs depending on spectral composition (Hutchins, Roquet, & Peretz, 2012; Russo & Thompson, 2005; Vurma, Raju, & Kuuda, 2010; Warrier & Zatorre, 2002) and vibrato (van Besouw, Brereton, & Howard, 2008). As vocal accuracy is mainly based on variations in the fundamental frequency during the tune, the perception of which is altered by the acoustical

components involved (and appreciated) in operatic singing voices, it may be particularly complex to evaluate such voices. The consequence of this complexity has been observed in expert listeners' unexpected tolerance of both "in tune" and "out of tune" intervals (Vurma & Ross, 2006). In Vurma and Ross's study, 17 expert listeners were asked to rate intervals performed by trained singers. The authors remarked that the judges sometimes considered intervals as "in tune" even when they were off by 20 to 25 cents. In a melodic context, Sundberg et al. (1996) observed the judgment of vocal accuracy in 10 commercial recordings (Schubert, *Ave Maria*). The seven listeners had to circle each tone they perceived to be "out of tune." Analyzing the mean  $F_0$  of circled and non-circled tones, they observed a tolerance zone of about 10 cents. However this zone was wider for some tones, and there was considerable variability among the judges. Although these studies highlighted the tolerance of music experts regarding professional singers' vocal accuracy, the experimental designs did not allow them to observe the evaluation process used by the music experts. In addition, the cause of this tolerance could be observed by analyzing and comparing the acoustical parameters involved in the operatic singing technique.

The present study aims to (a) assess the ability of expert listeners to evaluate the vocal accuracy of operatic singing voices, and (b) find out which acoustical measurements can predict the evaluation of vocal accuracy in operatic singing voices. In pursuit of goal (a), we observed intra- and inter-judge reliability in assessing the overall vocal accuracy of operatic singing voices and hypothesized that high intra-judge reliability would support the ability of expert listeners to rate the overall vocal accuracy of complex voices. In addition, the observation of inter-judge reliability would allow us to investigate consistency between the judges. For goal (b), we objectively measured several acoustical features and incorporated them in a predictive model of vocal accuracy evaluation in order to describe the process of evaluating the vocal accuracy of operatic singing voices.

## Methods

### *Participants*

The judges were 22 professional musicians (8 women, 14 men) aged 26 to 73 years old ( $M = 45.68$  years,  $SD = 11.16$ ). They had between 15 and 55 years of musical experience ( $M = 35.77$  years,  $SD = 10.74$ ). They had all received a classical music education (all instrument families were represented) in higher institutions such as music conservatories. When the study took place, all participants were still performing in public and reported practicing their instrument(s) 18.68 hours/week on average.

### *Material*

All the sung performances used in the perceptual task were selected from the database: <http://sldr.org/sldr000792/en>. Professional or semi-professional singers were asked to produce two vocal glissandi. Then they individually performed, a cappella, two different melodies (the French version of “Happy Birthday” and a romantic melody of their choice) with two different techniques (singing without any particular technique and with a Western operatic singing technique), in a specific order. For the popular song, no particular starting note was given so the participants could perform in their comfortable range. In this study, we examined the last sung performance recorded, namely the popular song “Happy Birthday” performed with an operatic singing technique. We therefore presumed that the singers were warmed up and were comfortable enough to perform a characteristic sample of operatic singing technique.

The sound recordings were made using a head-worn microphone (Sennheiser HS2, Wedemark, Germany) positioned at a constant distance of 2 cm from the left corner of the mouth as well as a Marantz Professional Solid State Recorder PMD67 (Kanagawa, Japan). The performances were recorded with a sampling frequency of 44.1 kHz and a resolution of 16 bits.

In order to apply a pairwise comparison paradigm, we needed relatively short stimuli and thus extracted the last sentence of “Happy

“Birthday” (Figure 1). From the initial corpus, we selected female sung performances (from 245.42 Hz to 449.26 Hz,  $M = 352,55$  Hz,  $SD = 21.13$ ), where the last tone of the tune was long enough to examine its vocal quality (from 1.13 s to 1.98 s,  $M = 1.45$  s,  $SD = 0.09$ ). Based on these two conditions, we selected 14 excerpts sung by 14 different professional or semi-professional singers, aged 21 to 66 years old ( $M = 34.86$ ;  $SD = 12.72$ ). They all reported having received an education in classical music and giving regular solo vocal performances in a classical style. They had begun singing lessons between 10 and 49 years old ( $M = 20.79$ ;  $SD = 9.23$ ) and reported between 6 and 17 years of singing lessons ( $M = 10.57$ ;  $SD = 3.58$ ). They practiced their vocal instrument 14.07 hours/week on average.

## ***Acoustical analysis***

For the 14 sung performances, six variables were observed, grouped into performance parameters (vocal accuracy, average tempo and fundamental frequency of the starting note) and quality parameters (energy distribution, vibrato rate and vibrato extent). All the acoustical analyses were run on a Macintosh (Mac OS X, Version 10.6.8).

### **Performance parameters.**

The *vocal accuracy* measurement was based on the relative pitch differences ( $F_0$  variations) between the 5 tones analyzed in the musical excerpt (in cents). Using AudioSculpt 2.9.4v3 software (Ircam, Paris, France), a Short-Time Fourier Transform (STFT) analysis was performed and markers were manually placed on the spectrogram in order to select the stable part of each tone (i.e., avoiding the attacks and the glides between tones). The mean  $F_0$  of each tone was then extracted with OpenMusic 6.3 software (Ircam, Paris, France). Note that when a vibrato occurred (mainly for the last tone), we selected complete vibrato cycles for the mean  $F_0$  measurement. We then calculated the size in cents of each interval between tones 1 and 5 (Figure 1). Each value was compared to the standard given by the musical score. For instance, the last interval (between tones 4 and 5) should be two semitones, corresponding to 200 cents. If the singer produced an interval of 150 or 250 cents, that would mean a deviation of 50 cents. We then averaged each deviation to generate the vocal accuracy criterion. A small



deviation indicated that the musical intervals were very precise. Note that this measurement was computed on the basis of an equal temperament, which is a compromise tuning scheme developed for keyboard instruments. Although the equal tempered system seems not to be appropriate for a cappella choirs (Howard, 2007a, 2007b), it remains a reference in Western tonal music.



**Figure 1.** Musical score representing the last sentence of the tune “Happy Birthday” with the number of the tones used to calculate the vocal accuracy parameter.

As performance parameters, we also observed the *fundamental frequency* ( $F_0$ ) of the starting tone (in Hertz) and the average *tempo*, computed on the basis of the excerpt’s length, by dividing the number of pulses by the duration (in beats per minute).

### Quality parameters.

For these acoustical analyses, we extracted the last tone in each melody, which is also the longest tone of the sung excerpt (tone 5 in Figure 1), sung on the vowel / $\varepsilon$ /.

Energy distribution, vibrato rate and extent were measured with AudioSculpt and OpenMusic software (Ircam, Paris, France). *Energy distribution* allows one to observe the spectral balance and thus the energy in the band containing the singer’s formant. This measure was computed in two steps. First, we selected three bands from the total frequency range, 0–2.4 kHz, 2.4–5.4 kHz and 5.4–10 kHz. Then, the energy of the second band was divided by the energy across all three bands. Note that a high score for the energy distribution variable shows a strong reinforcement of the band containing the singer’s formant. The *vibrato rate* (VR) and *vibrato extent* (VE) were based on the fact that vibrato is essentially a periodic modulation of fundamental frequency, which is generally found to be closely sinusoidal (e.g., Prame, 1994, 1997). VR specifies the number of periodic modulations of the  $F_0$  per second (in Hertz) and VE corresponds to the difference between the minimum and the maximum of the  $F_0$  within the same tone (in cents).

## ***Perceptual task***

The 14 sung performances were presented with a pairwise comparison paradigm. Each one was compared to the other ones in a random order. Thus, the total number of pairs was  $N(N-1) / 2$ , or 91 pairs to compare in the present study.

The participants pushed buttons on a graphical user interface to listen to the two sung performances as often as they wished. For each pair, they were asked to indicate which excerpt was the most “in tune,” considering only vocal accuracy as a parameter (and not the quality of the voice). Participants also had the possibility of labeling both excerpts in a pair as “equal” regarding the vocal accuracy criterion. After each pair, they saved their choice and a new trial was presented, that is, another pair of sung performances to compare.

The perceptual task was scored in two steps. For each pair to be compared, the total score of the sung performance considered as the most “in tune” was increased by one point (no points were awarded to the “out of tune” excerpt in the pair). If both sung performances of the pair were judged to be “equal,” the total score of both sung performances was increased by 0.5 points. After the presentation of the 91 pairs, the total score of each musical excerpt was computed. The result was a ranking of the 14 singers, with higher scores for the sung performances considered to be the most “in tune.”

The auditory stimuli were presented through professional headphones (K271 MKII, AKG, Vienna, Austria) at a comfortable loudness level. The entire perception task was approximately 25 minutes long, and participants were given a break halfway through.

The same procedure was administered twice (test-retest) with 8 to 15 days in between, in order to observe the intra-judge reliability.

## ***Statistical analyses***

Intra-judge reliability was evaluated as follows. For each judge, the rankings of the singers were derived from that judge’s ratings, separately in the test and the retest. Differences in rankings (test minus retest) were obtained, and the sample variance of the 14 differences in

rankings was derived as a measure of intra-judge reliability. Small variances indicated minor changes in rank between the test and the retest, and hence strong intra-judge reliability, while larger variances reflected larger differences in rank and lower intra-judge reliability.

Inter-judge reliability was evaluated in the same way as intra-judge reliability. Each singer was assigned a set of rankings, with each one corresponding to her performance (compared to the other singers) for one judge. The variance of the rankings was then computed. Small variances indicated small differences in the ranking of singers' performances across judges (and therefore strong inter-judge reliability), while larger variances indicated lower agreement among the judges' rankings.

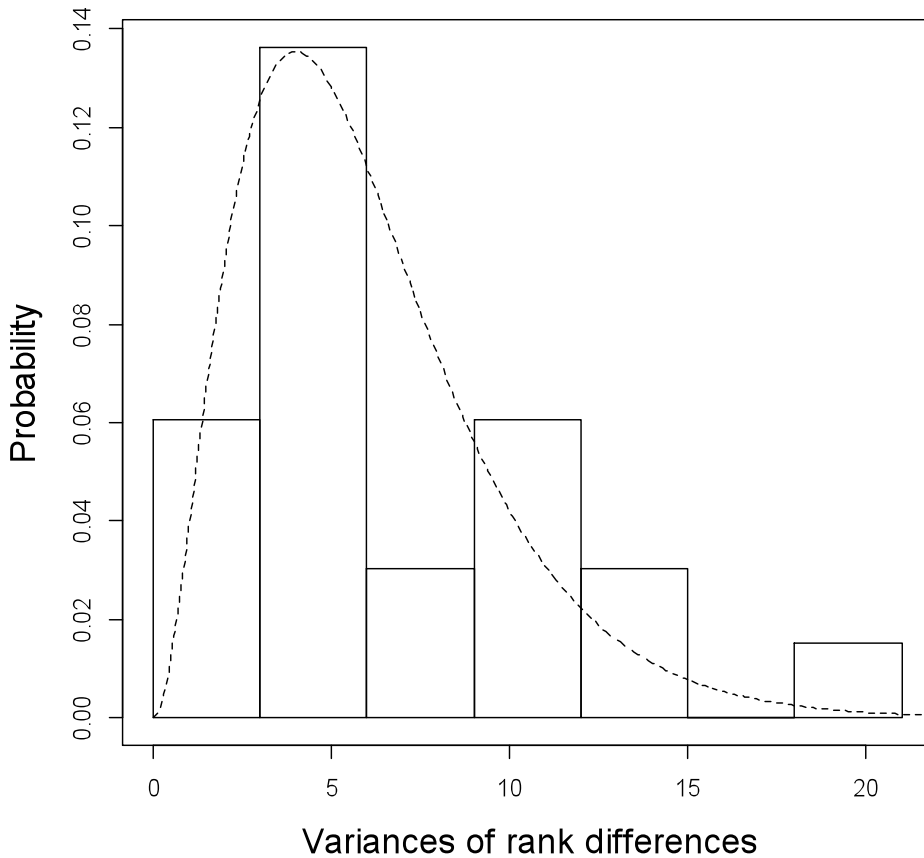
In practice, issues involving intra- and inter-judge reliability were detected by a similar process. First, the sample distributions of the respective variances were compared to a chi-squared distribution by selecting the degrees of freedom that maximized the p-value of a Kolmogorov-Smirnov test. Then, a threshold was set as the quantile of the chi-squared distribution with a lower tail probability of 0.95. Judges whose variance of differences in rankings (test minus retest) was greater than that threshold were considered to have been affected by an intra-judge reliability problem and were excluded from further analyses.

Finally, the relationships between the judges' ratings and the objective measurements (performance and quality parameters) were investigated through a multiple linear modeling approach. The variables (judge, vocal accuracy, tempo,  $F_0$ , energy, VR and VE) were included as covariates, in both main effects and pairwise interactions. The model was then simplified to retain the most statistically significant terms while retaining an acceptable  $R^2$  coefficient. The final model was further interpreted in terms of the main effects and interactions found in that model (through an analysis of variance), as well as the direction and magnitude of meaningful terms.

## Results

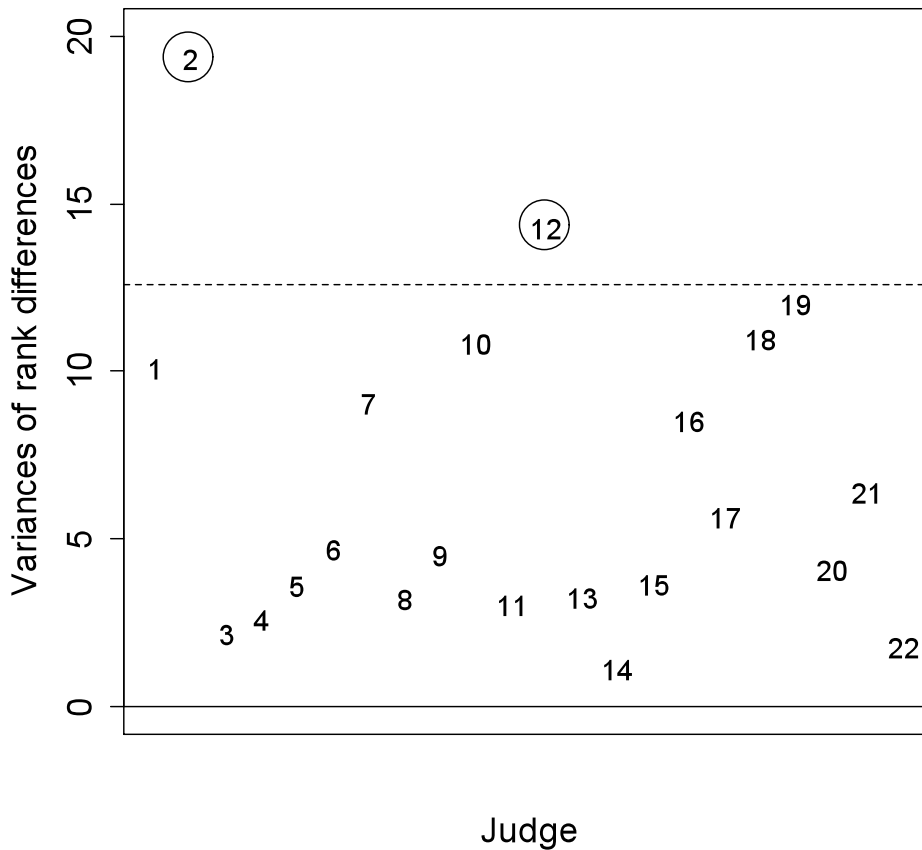
### *Intra-judge reliability*

The distribution of the 22 variances of rank differences is displayed in Figure 2 in a histogram. Because the nature of the variance is an index of intra-judge reliability, an adjustment of the sample distribution by a chi-squared distribution is expected. The best adjustment is obtained by considering six degrees of freedom, for which the Kolmogorov-Smirnov test returns a p-value of 0.535. The corresponding chi-squared distribution is shown in Figure 2 by a dashed line.



**Figure 2.** Sample distribution of the variances of rank differences. The dashed line represents the probability of the chi-squared distribution with six degrees of freedom.

The quantile of this chi-squared distribution with six degrees of freedom and lower tail probability of 0.95 equals 12.592 and serves as a detection threshold for intra-judge reliability issues. Figure 3 presents the 22 variances of rank differences for each judge, represented by their respective numbers. The horizontal dashed line represents the detection threshold.

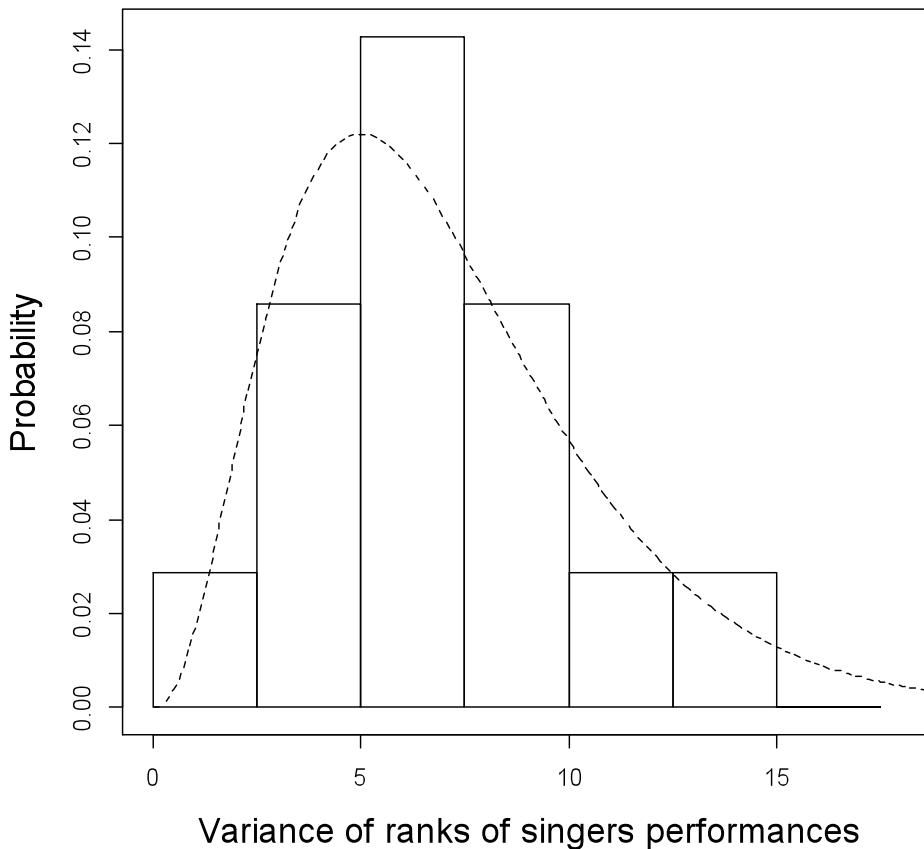


**Figure 3.** Variances of rank differences by judge, displayed by their number. The dashed line represents the cutoff value for identifying intra-judge reliability issues.

As Figure 3 shows, judges 2 and 12 exhibited a sample variance larger than the threshold, meaning that their ranking differences were too dispersed with respect to the overall set of variances. In other words, the differences in performance ranks (between the test and the retest) varied more for these two judges compared to the overall variability of these rank differences. Judges 2 and 12 were therefore discarded from further analyses.

## *Inter-judge reliability*

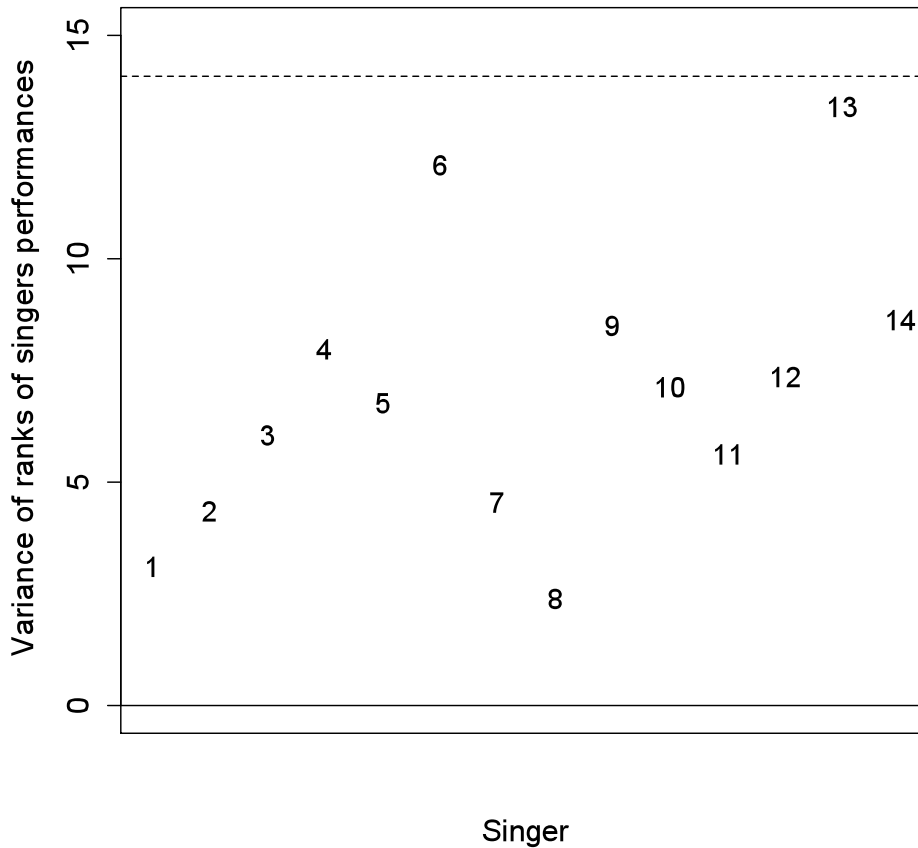
Figure 4 is a histogram showing the variances of ranks for each singer's performance. Recall that judges 2 and 12 were removed from this analysis because of their low intra-judge reliability level. The adjustment by a chi-squared distribution is best when seven degrees of freedom are chosen (Kolmogorov-Smirnov  $p = 0.924$ ).



**Figure 4.** Sample distribution of the variances of rank of the sung performances. The dashed line represents the probability of the chi-squared distribution with seven degrees of freedom.

The corresponding quantile with a lower tail probability of 0.95 equals 14.067. As Figure 5 shows, none of the variances of singers' ranks are greater than that threshold, which indicates quite strong inter-judge

reliability across all the singers' performances and thus the use of similar judgment process by all 20 judges.



**Figure 5.** Variances of the rankings of sung performances, displayed by their number. The dashed line represents the cutoff value for identifying inter-judge reliability issues at the singer level.

### ***Predictive model of vocal accuracy evaluation***

The multiple regression model including all the aforementioned covariates (means of the six objective parameters are presented in Table 1) and main effects has an R<sup>2</sup> coefficient of 82.7%. The final model retained has an R<sup>2</sup> coefficient of 78.8%. This model contains all main effects as well as several pairwise interaction terms. The output of the analysis of variance (ANOVA) is displayed in Table 2.



All covariates appear at least once in a pairwise interaction term. This means that the judges' rating depends on all the covariates (including the judge himself or herself) but always by means of an interacting effect with (at least) one other covariate. Note that some interaction terms were kept in the final model although the Table 2 lists them as non-significant interactions. The reason is that, although not statistically significant, their contribution to the overall R<sup>2</sup> coefficient cannot be neglected. This is probably a sign of collinearity between the covariates.

Singers	Performance parameters			Quality parameters		
	Vocal accuracy (in cents)	Tempo (bpm)	F <sub>0</sub> (in Hertz)	Energy distribution (ratio)	Vibrato rate (in Hertz)	Vibrato extent (in cents)
1	87.5	53.33	245.44	1.58	5.40	246
2	115.5	46.45	438.54	2.07	5.52	148
3	14.5	53.33	398.91	1.47	6.25	186
4	17	57.60	449.26	1.89	4.96	154
5	29	53.33	434.05	1.74	5.43	134
6	36.5	46.45	245.42	2.14	5.58	168
7	18	51.43	405.30	1.61	5.33	150
8	15.5	65.45	436.28	1.76	5.61	132
9	9.5	49.65	398.92	1.50	6.56	130
10	21	57.60	268.31	1.60	6.14	174
11	64	53.33	249.07	2.17	5.08	204
12	15	65.45	402.91	1.27	6.45	118
13	99.5	60.00	396.51	1.39	5.81	270
14	25.5	68.57	318.94	1.48	6.67	170

**Table 1.** Values of Each Singer's Objectively Measured Performance and Quality Parameters

Variable	Sum Squares	DF	Mean Square	F-value	p-value
Vocal accuracy	234.38	1	234.38	58.04	0.000
Judge	0.00	21	0.00	0.00	1.000
Tempo	20.32	1	20.32	5.03	0.030
F <sub>0</sub>	89.86	1	89.86	22.25	0.000
Energy	93.34	1	93.34	23.12	0.000
VR	16.13	1	16.13	3.99	0.050
VE	39.98	1	39.98	9.90	0.000
Vocal accuracy:F <sub>0</sub>	46.60	1	46.60	11.54	0.000
Vocal accuracy:					
Tempo	607.97	1	607.97	150.56	0.000
Judge: Tempo	141.08	21	6.72	1.66	0.040
Judge:F <sub>0</sub>	82.27	21	3.92	0.97	0.500
Judge: VR	120.98	21	5.76	1.43	0.110
Judge: VE	267.36	21	12.73	3.15	0.000
Tempo:F <sub>0</sub>	2.94	1	2.94	0.73	0.390
Tempo: Energy	33.11	1	33.11	8.20	0.000
Tempo: VR	175.86	1	175.86	43.55	0.000
Tempo: VE	666.12	1	666.12	164.97	0.000
Residuals	767.21	190	4.04	///	///

**Table 2.** Analysis of Variance of the Retained Multiple Linear Regression Model

Scores cannot therefore be explained by a single covariate, or by one covariate that does not interact with the others. For instance, the vocal accuracy parameter does not explain the judges' rating as a main effect only but in association with, on the one hand, F<sub>0</sub> of the starting tone, and on the other hand, tempo. More specifically, the fitted model coefficients indicate that, conditionally upon vocal accuracy, the judges' ratings increase with lower F<sub>0</sub>; moreover, the lower the vocal accuracy, the larger the increase with lower F<sub>0</sub>. Similarly, conditionally upon vocal accuracy, the judges' ratings increase with faster tempo; the greater the vocal accuracy, the larger the increase with faster tempo. As Table 2 reveals, tempo is involved in many significant pairwise interaction terms, which is a sign that it plays a central role in judges' ratings. Moreover, these interactions confirm that there are no direct relationships between the performance and quality parameters and the judges' ratings.

Concerning the effect of judge, it is noticeable that, although this factor interacts with several other covariates, the significant interaction terms relate to only a few judges (at most three of them). This means that for most judges there is absolutely no interaction effect; in other words, scoring does not vary across all judges. Rather, a few of them tend to exhibit significant interaction effects with one or more other covariates ( $F_0$ , tempo, VR and VE), meaning that, for these judges, some quality parameters are most significant for the scoring process. Note that there was no interaction between vocal accuracy and judge. Indeed, the overall effect of vocal accuracy on the judges' ratings did not vary across judges. This result is in line with the inter-judge reliability analysis in terms of vocal accuracy.

## Discussion

This study supports the hypothesis that music experts are able to rate the vocal accuracy of operatic voices. Indeed, by observing the intra-judge reliability, we found that only two judges were not consistent between the test and the retest. In other words, each music expert seems to use similar strategies to rate the overall vocal accuracy of the musical material at different times. Future research on operatic singing voices could therefore be limited to one task instead of two. Note that the two inconsistent judges (judges 2 and 12 in Figure 3) did not show any particular features concerning their biographical data and musical background or training compared to the other participants. Thus, we hypothesized that they had changed their rating strategy between the test and the retest and did not include these participants in the inter-judge analysis. The high inter-judge reliability does not confirm the variability pointed out by Sundberg et al. (1996). On the contrary, this result suggests that the definition of vocal accuracy was clear enough for all the music experts despite their different musical backgrounds. The high intra- and inter-judge reliability could have been facilitated by the use of a pairwise paradigm. Indeed, the presentation of each pair of sung performances to be compared allows an indirect rating, whose efficiency has already been tested in the evaluation of disordered voices (Kacha, Grenez, & Schoentgen, 2005). The use of a pairwise comparison paradigm therefore seems to be appropriate for

the perceptual judgment of operatic singing voices. Given that previous studies of such complex voices found it difficult to obtain a consensus among listeners (Ekholm et al., 1998; Garnier et al., 2007; Howes et al., 2004), this paradigm could be suggested as a means of achieving better reliability among the judges. However, note that this experimental method involves the use of a limited number of short musical excerpts to ensure that the perceptual task is a reasonable length.

These first analyses highlighted the music experts' ability to evaluate operatic voices and thus supported the possibility of computing a predictive model on the basis of the judges' ratings. Although it is visible in the model, the effect of the individual judge on the ratings is not great. Only a few judges exhibited a significant interaction with some of the acoustical parameters. This supports the hypothesis that some judges pay more attention or are more sensitive to certain acoustical parameters than others. This has an indirect effect on their ratings, but those effects are very limited according to our analysis.

The multiple regression modeling explains 78.8% of the variance in the judges' ratings. Compared to the study by Larrouy-Maestri et al. (2013) with untrained voices, the objective measurement of vocal accuracy is not sufficient to predict the judges' rating in the case of operatic voices. Indeed, all covariates appeared at least once in the final model but the judges' rating cannot be explained by a single covariate, or by one covariate that does not interact with the others. For example, the vocal accuracy parameter affects the judges' rating only when associated with the  $F_0$  of the starting tone or with the tempo. This result confirms the tolerance effect observed in intervallic and melodic contexts (Sundberg et al., 1996; Vurma & Ross, 2006) and moderates the importance of vocal accuracy in the perceptual judgment of operatic singing voices. Note that the musical material in the present study (i.e., last sentence of "Happy Birthday") is not representative of the repertory of Western classical music. However, given that vocal accuracy is not affected by the melodic effect (Larrouy-Maestri et al., in press), we can imagine that the performance of a melody from the Western classical repertory would lead to a similar pattern of results; it would be interesting to confirm this hypothesis.

Although a sung performance could be “out of tune” but not perceived as such when an operatic singing technique is employed, the present study shows that the combination of several acoustical parameters influences the perceptual judgment of the sung performance. Consequently, the objective measurement of vocal accuracy based on fundamental frequency variations during the tune cannot be applied without considering the performance and quality parameters of the operatic singing voice. These findings are directly applicable in pedagogical settings. For instance, when a singing student is considered to be “out of tune” in a melodic context, he or she could be advised to improve intonation but also to take vocal quality and interpretation into account in order to be perceived as “in tune.” The observation of a larger sample of singers and music experts would make it possible to improve the predictive modeling and thus the pedagogical implications of this study. Note that the unexplained 21.2% of variance may be linked to other criteria such as rhythm accuracy or vocal perturbation (Butte, Zhang, Song, & Jiang, 2009), which could unintentionally intervene in the judgment of vocal accuracy.

Concerning the perceptual process, several studies reported that spectral composition (Hutchins et al., 2012; Russo & Thompson, 2005; Vurma et al., 2010; Warrier & Zatorre, 2002) and vibrato (van Besouw et al., 2008) affect the perception of tones. The several interactions observed in our multiple regression modeling showed that the individual acoustical parameters cannot be observed separately. Thanks to the analysis of natural sung performances, this study observed music experts’ perceptual judgment of vocal accuracy in a melodic context and pointed out the relevance of performance and quality parameters in the evaluation of operatic voices. However, further studies using different combinations of the performance and quality parameters in synthesized material would allow one to manipulate these parameters and thus to clarify the perception of vocal accuracy in the case of operatic singing voices.

## **Conclusion**

This study focused on the evaluation of vocal accuracy in operatic singing voices. By observing the perceptual judgment of 14 sung

melodies performed by professional singers, this study highlights the high intra- and inter-judge reliability of music experts when evaluating the vocal accuracy of operatic voices. However, the perceptual rating was not directly linked with vocal accuracy, as objectively analyzed for each sung performance, which moderates the importance of vocal accuracy in the evaluation strategy of operatic singing voices. Interestingly, several interactions between performance and quality parameters contribute to explain the judges' ratings. Indeed, both vocal quality (energy, vibrato rate and extent) and interpretation of the tune (tempo,  $F_0$ ) contributed to the judgment of vocal accuracy. In addition to the pedagogical applications of these findings, this study opens up certain research perspectives concerning the perception of complex stimuli such as operatic singing voices.

## **Acknowledgments**

The authors thank the Centre Henri Pousseur in Liège, the Laboratories of Images, Signals and Telecommunication Devices (LIST) in Brussels, and Guillaume Videlier for technical support. We also thank the singers and judges who allowed us to run this experiment and Marion Nowak for her help with the data collection.







# **Discussion générale**



# DISCUSSION

---

L'objectif de ce travail était d'explorer l'évaluation de la justesse vocale en contexte mélodique. Si la définition musicale et perceptive de la justesse, exposée au chapitre I, peut paraître simple, l'évaluation de la justesse d'une performance musicale est complexe. Comme en témoignent les conclusions du chapitre II, les méthodes d'évaluation (subjective versus objective) utilisées actuellement montrent des limites. D'une part, la méthode "subjective" ne permet pas d'évaluer précisément la justesse d'une performance. D'autre part, la méthode "objective" manque de pertinence, étant donné que les critères musicaux à quantifier ne sont pour l'instant pas spécifiés. Afin d'élaborer une méthode objective, pertinente et cohérente avec la méthode subjective d'évaluation de la justesse vocale en contexte mélodique, nous avons tenté de répondre à trois questions :

- Comment évaluer objectivement la justesse de performances vocales en contexte mélodique ?
- Quelle est la pertinence d'une telle méthode ?
- Quelles clarifications cette méthode objective apporte-t-elle à la notion de justesse mélodique ?

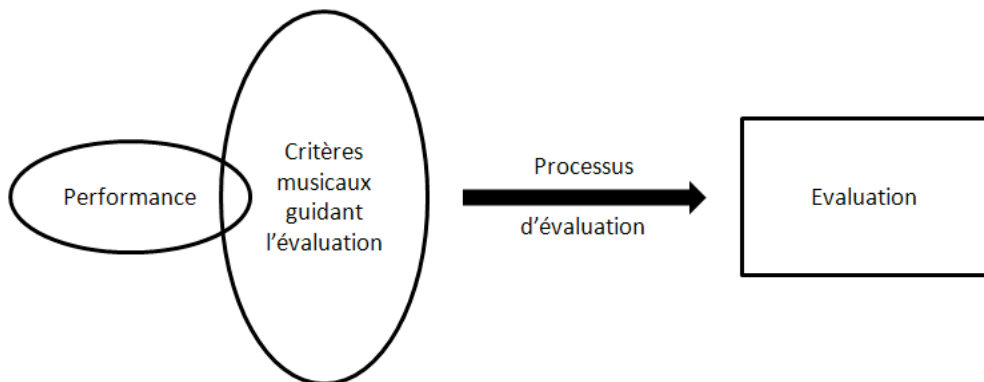
Le point de départ de nos recherches a été l'élaboration d'une méthode d'évaluation objective qui puisse s'appliquer aux performances de chanteurs occasionnels et qui tienne compte de critères pertinents, tant au niveau des règles musicales dictées par le système tonal occidental que de la perception mélodique qu'un auditeur en a (Etude 1). Pour affiner ces outils, nous avons souhaité examiner les critères à la base de l'évaluation effectuée par des juges (Etude 2). Cette méthode a ensuite été appliquée dans des contextes expérimentaux, afin d'en observer sa pertinence (Etude 3) mais également d'en explorer ses limites (Etude 4). Finalement, la voix est un instrument particulièrement accessible qui peut également être développé à l'aide d'un entraînement

spécifique, comme dans le cas des chanteurs lyriques. Nous avons donc souhaité examiner l'évaluation de ces voix particulières dont le signal acoustique et la dimension esthétique diffèrent des voix de chanteurs occasionnels (Etudes 5, 6 et 7).

Avant de discuter de l'évaluation objective de la justesse d'une performance vocale en contexte mélodique, de la pertinence de cette méthode et de la clarification de la notion de justesse, nous présentons une synthèse des principaux résultats de chaque étude.

## Synthèse des études

L'**étude 1** peut être illustrée par le schéma idéal d'évaluation d'une performance musicale de McPherson et Thompson (2004), c'est à dire par la présence de deux composantes (les critères et l'évaluation), auxquelles nous avons ajouté la variable "performance" (Figure 1).



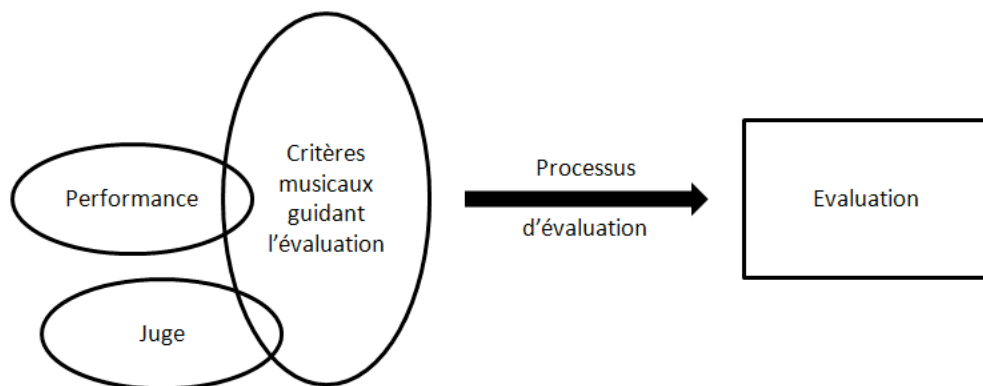
**Figure 1.** Schéma idéal d'évaluation d'une performance musicale de McPherson et Schubert (2004), avec ajout de la variable "performance" : évaluation objective de la justesse vocale.

Dans cette étude, nous avons analysé la justesse des performances vocales en considérant la variable "critères musicaux" comme fixe. En effet, ces critères ont été prédéfinis d'après les règles musicales régissant le système tonal occidental. Chez les chanteurs occasionnels, il a été montré à plusieurs reprises que chanter dans un tempo lent favorise une bonne justesse (Dalla Bella et al., 2007, 2009). Par ailleurs, un entraînement musical permet une expertise musicale générale et spécifique à l'instrument (Zatorre et al., 2007) et ainsi l'amélioration de la justesse (Amir et al., 2003 ; Dalla Bella et al., 2007 ; Murry, 1990 ;

Watts et al., 2003). Dans le but d'observer la pertinence globale de notre méthode objective d'évaluation de la justesse, nous avons souhaité vérifier le bénéfice d'un tempo lent et de l'expertise vocale sur la justesse de la mélodie "Joyeux Anniversaire" chantée par 63 chanteurs occasionnels et 14 chanteurs professionnels. Notons que les chanteurs professionnels avaient pour consigne de chanter à deux reprises, une fois de manière "naturelle" et une seconde fois en utilisant une technique lyrique.

Chez les chanteurs occasionnels, les résultats confirment le lien entre le tempo et la justesse. De plus, nous avons observé l'effet de l'expertise vocale sur les critères de justesse observés (précision des intervalles et respect du centre tonal). Par ailleurs, chez les chanteurs professionnels, l'utilisation d'une technique lyrique mène à une altération de la précision des intervalles par rapport aux performances chantées sans technique lyrique. Outre la pertinence globale des outils utilisés, cette étude met en évidence la particularité des voix lyriques qui a été approfondie dans les cinquième, sixième et septième études de cette thèse.

L'**étude 2** avait pour objectif de cibler les critères musicaux utilisés par des juges pour évaluer une performance vocale en contexte mélodique. Cette expérience peut être schématisée par une variante du modèle idéal d'évaluation de McPherson et Schubert (2004), à laquelle ont été ajoutées les variables "performance" et "juge" (Figure 2). A l'aide de la méthode objective d'évaluation élaborée dans le cadre de la première étude, nous avons pu focaliser notre attention sur la variable "juge".



**Figure 2.** Schéma idéal d'évaluation d'une performance musicale de McPherson et Schubert (2004), avec ajout des variables "performance" et "juge" : évaluation subjective de la justesse vocale.

Autrement dit, nous avons comparé les méthodes objective et subjective actuellement utilisées afin d'observer le comportement des juges et donc de préciser les critères musicaux à quantifier. Pour ce faire, une base de données a été constituée. Elle comprend 166 "Joyeux anniversaire" produits par des chanteurs occasionnels (<http://sldr.org/sldr000774/en>). Ces productions ont été soumises à un jury d'écoute incluant 18 juges experts (logopèdes, experts en musique – étudiants et professionnels - et chanteurs). Parallèlement, les performances ont été analysées en suivant les règles musicales du système tonal occidental (respect des contours mélodiques, de la taille des intervalles et de la tonalité). Les résultats indiquent que les scores obtenus à l'aide des deux méthodes (subjective versus objective) sont fortement corrélés. Notons également que les juges évaluent les performances de manière similaire. En effet, les résultats montrent une corrélation significative et élevée ( $r = .85$ ) entre la déviation des intervalles et la note globale attribuée par les juges, lorsque l'échantillon des juges est réduit. Par ailleurs, les résultats indiquent que la variable relative au respect de la tonalité est à considérer comme une mesure pertinente de la justesse en contexte mélodique. En effet, ce critère n'a pour l'instant pas été quantifié dans les études employant une méthode objective (Dalla Bella & Berkowska, 2009 ; Dalla Bella et al, 2007 ; Lévêque et al., 2009 ; Pfordresher & Brown, 2007 ; Pfordresher et al., 2010) alors que différentes études ont montré que la notion de tonalité se développe progressivement (Hannon & Trainor, 2007 ; Miyamoto, 2007) pour être acquise à l'âge adulte (Trainor & Trehub, 1992). Cette

étude confirme donc qu'il est pertinent de quantifier les erreurs relatives au respect de la tonalité pour évaluer objectivement la justesse vocale de chanteurs occasionnels.

Alors que les deux premières études ont permis de mettre au point et de préciser notre méthode objective d'évaluation de la justesse vocale en contexte mélodique, les troisième et quatrième études visent à appliquer cette méthode en contexte expérimental afin d'en observer la pertinence et les limites.

Dans l'**étude 3**, nous avons examiné l'intérêt d'utiliser une performance vocale en contexte mélodique, c'est-à-dire une mélodie complète, pour évaluer la justesse vocale. Comme indiqué dans le chapitre II, les recherches menées sur la justesse vocale utilisent majoritairement des tâches de répétition de sons isolés ou de séquences mélodiques (Amir et al., 2003 ; Granot et al., in press ; Hutchins & Peretz, 2012 ; Moore et al., 2007, 2008 ; Nikjeh et al., 2009 ; Pfordresher & Brown, 2007, 2009 ; Pfordresher et al., 2010 ; Watts et al., 2002). Cependant, le principe d'une mélodie est l'organisation de sons dans une structure régie par un système musical (Cross, 2001 ; Lerdahl & Jackendoff, 1983). Peu d'études se sont pour l'instant intéressées à la justesse de performances en contexte mélodique (Dalla Bella & Berkowska, 2009 ; Dalla Bella et al., 2007 ; Wise & Sloboda, 2008) et aucune n'a, à ce jour, comparé de manière directe les scores obtenus lors de performances en contexte mélodique à des tâches de répétition (sons isolés ou séquences). Nous avons donc souhaité observer si une tâche de répétition vocale reflète la capacité à chanter une mélodie entière. Les résultats mettent en évidence que les scores moyens à la tâche de répétition et à la performance mélodique ne diffèrent pas. De plus, nous observons une corrélation positive entre les deux. Cependant, seul le critère relatif aux intervalles est corrélé au score obtenu à la tâche de répétition vocale. Autrement dit, les chanteurs ayant de bons scores à la tâche de répétition respectent également bien la taille des intervalles. Etant donné que la tâche de répétition de sons isolés ne donne pas d'information sur le respect de la tonalité et que ce critère est un indicateur de justesse (voir Etude 2), il semble que les informations obtenues à l'aide d'une tâche de répétition de sons isolés ne représentent pas complètement la capacité de justesse d'un chanteur occasionnel. En d'autres termes, si les tâches de répétition de sons isolés

proposées couramment ont une utilité certaine dans le but de comprendre les causes d'un trouble de justesse, les conclusions de ces études sont néanmoins à modérer étant donné que la composante relative à la tonalité n'est pas observée.

L'**étude 4** a été conçue pour examiner l'effet du contexte d'évaluation sur la justesse et plus spécifiquement l'impact du stress sur la justesse d'une performance mélodique. En effet, il a déjà été montré qu'une performance musicale en public génère du stress (Kenny, 2011) et que, sous stress, la qualité de la performance peut être altérée ou améliorée (Craske & Craig, 1984 ; Hamman & Sobaje, 1983 ; Yerkes & Dodson, 1908 ; Yoshie et al., 2008, 2009). Nous avons souhaité observer l'effet du stress sur les différents critères de justesse. Dans ce but, nous avons analysé les performances vocales de 31 élèves inscrits au conservatoire (musiciens mais non chanteurs), à l'aide de notre méthode d'évaluation. Les productions ont été récoltées en situation calme et en situation stressante, c'est-à-dire le jour de leur examen. Afin de contrôler l'enjeu de la situation, nous avons différencié deux groupes : les étudiants de première année, pour qui l'examen est formatif, et les étudiants de deuxième année, pour qui l'examen est certificatif. Les résultats indiquent que tous les élèves montrent un niveau d'anxiété supérieur le jour de l'examen (fréquence cardiaque plus rapide, symptômes somatiques et cognitifs ressentis comme plus intenses et plus "négatifs"), ce qui confirme que les deux situations d'évaluation sont bien contrastées. Il est intéressant de noter que l'effet du stress n'est pas le même selon l'enjeu de la situation. En effet, les élèves de première année respectent mieux la taille des intervalles lors de la situation stressante, tandis que les élèves de deuxième année ont des difficultés à respecter le centre tonal de la mélodie en situation d'examen. Ces résultats confirment donc que le contexte d'évaluation (situation stressante versus non stressante) mais également l'enjeu de l'évaluation (formatif versus certificatif) ont une influence sur la justesse des performances. L'intérêt de tenir compte du contexte de l'évaluation et de l'enjeu de la performance s'avère donc primordial. Par ailleurs, les critères de justesse peuvent évoluer différemment, il est donc intéressant de mesurer les deux afin d'avoir une vue d'ensemble de la justesse d'une performance mélodique.



Les trois dernières études s'adressent aux voix travaillées et visent à mieux comprendre les résultats surprenants de l'étude 1. En effet, la justesse des chanteurs professionnels utilisant une technique vocale lyrique ne se différencie pas de celle des chanteurs occasionnels pour la précision des intervalles alors que des chanteurs professionnels sont censés chanter plus juste que des chanteurs occasionnels (Amir et al., 2003 ; Dalla Bella et al., 2007 ; Murry, 1990 ; Watts et al., 2003). Sundberg et al. (1996) et Vurma et Ross (2006) avaient déjà observé que des chanteurs lyriques n'étaient pas toujours précis dans leurs performances vocales. Deux hypothèses ont été avancées pour expliquer le profil particulier des chanteurs lyriques. La première concerne le choix du matériel musical (variations de la fréquence fondamentale dans le but de donner un caractère "lyrique" à une mélodie populaire) et la seconde concerne la perception de ces voix particulières (influence de paramètres acoustiques sur l'évaluation de la justesse).

Dans l'**étude 5**, nous avons quantifié l'effet de la technique vocale sur deux critères (intervalles et tonalité) en demandant à 50 chanteurs de produire "Joyeux Anniversaire" avec et sans technique lyrique. Afin de tester l'hypothèse relative au choix du matériel musical, nous leur avons également proposé d'interpréter une mélodie de leur répertoire, dans les deux conditions précitées. L'analyse de la base de données ainsi constituée (<http://sldr.org/sldr000792/en>) a permis de confirmer l'effet de la technique vocale sur la précision des intervalles et d'écarter l'hypothèse liée au choix du matériel musical (chanson populaire versus mélodie de leur répertoire). Par ailleurs, les résultats relatifs au respect de la tonalité montrent que la production d'une mélodie de leur répertoire engendre davantage de changements de tonalité que la production d'un chant populaire tel que "Joyeux Anniversaire". Ce résultat n'est pas étonnant car que nous avons choisi d'analyser de courts extraits afin de préserver un nombre de notes similaire pour les deux mélodies et que les informations relatives au centre tonal de la mélodie sont alors limitées. En conclusion, les chanteurs professionnels respectent peu la taille des intervalles lorsqu'ils utilisent une technique lyrique, quelle que soit la mélodie interprétée.

Les deux dernières études visent à comprendre pourquoi les mesures objectives de la justesse ne semblent pas refléter ce qui est attendu chez des chanteurs professionnels.

Comme mentionné dans le chapitre IV, Garnier et al. (2007) ont mis en évidence l'importance du vibrato et du timbre dans l'appréciation de voix lyriques. Ces caractéristiques acoustiques sont développées à l'aide d'un entraînement (Brown et al., 2000 ; Lundy et al., 2000 ; Mürbe et al., 2007 ; Omori et al., 1996) et seraient donc particulièrement recherchées par les chanteurs professionnels. Cependant, la combinaison complexe de ces caractéristiques et la définition de la voix lyrique restent à clarifier.

**L'étude 6** vise à dresser un profil acoustique et musical des chanteurs lyriques en observant l'effet de la technique et de la mélodie sur les caractéristiques musicales (hauteur de la performance, tempo, niveau de pression sonore), acoustiques (balance spectrale, amplitude et fréquence du vibrato) et relatives à la perturbation du signal ( $SDF_0$ , SNR, Jitter, Shimmer). En effet, celles-ci sont, pour l'instant, peu examinées (Butte et al., 2009). Les résultats mettent en évidence que la majorité des paramètres observés ne varie pas en fonction de la mélodie chantée. L'utilisation d'une technique vocale lyrique, quant à elle, a un effet sur la perturbation du signal, sur le tempo mais également sur les deux caractéristiques acoustiques particulièrement recherchées, à savoir le vibrato et le timbre. Finalement, un premier modèle, qui mériterait d'être complété par l'analyse d'autres composantes vocales, permet de préciser les caractéristiques acoustiques et musicales définissant les voix lyriques.

**L'étude 7** a pour objectif de tester l'hypothèse relative à l'influence de paramètres acoustiques sur l'évaluation de la justesse. En effet, au vu des précédents résultats concernant la justesse des chanteurs lyriques (études 1 et 5) ainsi que la complexité de ces voix (étude 6), nous avons émis l'hypothèse que le critère de justesse ne serait peut-être pas lié à la relation entre les notes, comme l'est la justesse des voix de chanteurs occasionnels. Dans ce but, nous avons suivi le principe utilisé dans l'étude 2, c'est à dire la mise en parallèle des méthodes subjective (évaluation par des juges experts) et objective (mesures acoustiques) sur un même corpus de voix (<http://sldr.org/sldr000792/en>). Dans la base de données des voix lyriques, 14 performances ont été sélectionnées, la dernière phrase de la mélodie "Joyeux Anniversaire" a été extraite et analysée (précision des intervalles, balance spectrale, amplitude et fréquence du vibrato, tempo, hauteur de la performance).

Ces extraits ont été proposés à 22 juges experts à l'aide d'une tâche de comparaison par paires, effectuée à deux reprises (test et retest). Nous avons donc pu observer la concordance intra et inter-juges et comparer le classement des performances avec les mesures acoustiques des 14 échantillons vocaux. Les résultats mettent en évidence que seulement deux juges n'évaluent pas les performances de la même manière entre le test et le retest et que les 20 juges cohérents se basent visiblement sur les mêmes critères pour évaluer la justesse d'une performance vocale. Par ailleurs, toutes les variables observées (dont le juge lui-même) interviennent dans le modèle prédictif de l'évaluation de la justesse vocale, mais toujours en interaction avec une ou plusieurs variables. En d'autres termes, l'évaluation des juges ne peut pas être expliquée par une seule variable. Dans ce modèle, la précision des intervalles est associée à la hauteur de la performance et au tempo, c'est à dire qu'à justesse égale, la performance sera mieux notée lorsque la hauteur tonale est basse ou si le tempo est rapide. Les conclusions de cette étude suggèrent donc que des juges experts sont capables d'évaluer des voix lyriques et que la notion de justesse est partagée. Cependant, contrairement au cas des chanteurs occasionnels (Etude 2), la justesse vocale n'explique pas à elle seule l'évaluation donnée par les juges. Au contraire, les multiples interactions observées entre les variables confirment que les caractéristiques musicales et acoustiques du chanteur interviennent dans l'évaluation de la justesse des voix lyriques.

Les sept études réalisées dans ce travail de thèse permettent d'approfondir nos connaissances aux niveaux méthodologique et théorique. Outre les éléments de réponses apportés au sujet de l'évaluation objective de la justesse d'une performance vocale en contexte mélodique, de la pertinence de cette méthode et de la clarification de la notion de justesse, ce travail de thèse ouvre également la voie à de multiples pistes à explorer.

## **Evaluation objective de la justesse mélodique**

Le second chapitre de cette thèse met en évidence les limites des outils actuellement utilisés pour évaluer objectivement la justesse des performances vocales en contexte mélodique, que ce soit sur le plan

technique (analyses acoustiques) ou des critères (types d'erreur à quantifier). Afin de répondre à la question concernant l'évaluation objective de la justesse mélodique et envisager des perspectives de recherches à ce niveau, nous rappelons la méthode élaborée dans le cadre de cette thèse et décrivons les avantages ainsi que les limites des outils d'analyse utilisés et des critères d'évaluation sélectionnés.

### ***Méthode d'analyse de la justesse vocale en contexte mélodique***

La première étape de nos recherches consistait à trouver des programmes informatiques adéquats pour représenter, segmenter et extraire les informations fréquentielles permettant d'observer les variations de la fréquence fondamentale tout au long d'une mélodie. Nous avons opté pour les programmes créés au sein de l'Institut de Recherche et Coordination Acoustique/Musique (Ircam), initialement destinés à la recherche musicale. En effet, ceux-ci permettent une visualisation précise du signal et la segmentation manuelle des éléments choisis pour les analyses acoustiques. De plus, les options de type "time stretching", "masquage" ou encore "nettoyage" sont aisément accessibles. La première permet d'écouter le signal acoustique de manière ralentie ou statique, à l'aide d'un curseur, sans que la hauteur ou le timbre ne soient altérés. La seconde consiste à poser des filtres sur tout élément non désiré (bruits ou subharmoniques). La troisième permet d'isoler la voix d'un environnement bruyant. Il s'agit alors de sélectionner un échantillon de bruit, de l'analyser et de soustraire ces informations acoustiques de la séquence sonore.

Les réglages, recommandés par des experts en analyse acoustique et précisés à la suite de pré-tests, permettent d'obtenir une précision fréquentielle et temporelle élevée, adaptée aux voix de chanteurs occasionnels et professionnels. La suite de la procédure d'analyse est automatique. En effet, le logiciel AudioSculpt (Ircam, Paris, France) permet l'analyse des accords harmoniques de chaque segment. Le programme OpenMusic (Ircam, Paris, France) permet, quant à lui, d'extraire les informations relatives à la fréquence fondamentale et de sélectionner les segments à utiliser pour le calcul des erreurs, c'est à dire les notes et non les liaisons ou silences entre ces notes. Autrement

dit, le logiciel AudioSculpt permet d'analyser la composante spectrale tandis que le logiciel OpenMusic sélectionne la fréquence la plus basse, c'est à dire la fréquence fondamentale, et ce, même si les fréquences supérieures sont intenses. Une fois que la fréquence de chaque note est extraite, les fichiers .txt d'OpenMusic contenant la hauteur de chaque note de la mélodie, en cents, sont exportés dans un tableau Excel contenant les formules pour quantifier les erreurs. Quelle que soit la mélodie à analyser, ce tableau permet de comparer les données recueillies aux données de référence (les hauteurs relatives de chaque note de la mélodie étant préalablement encodées). Pour rappel, cette méthode vise à observer la justesse mélodique et se base donc sur la relation entre les notes et non sur une comparaison entre une cible et sa répétition, comme ce serait le cas pour une tâche de répétition de sons isolés. En d'autres termes, la tonalité de départ n'est pas prise en compte dans le calcul des erreurs, le principe étant de quantifier les erreurs mélodiques exposées au chapitre I, à savoir, le nombre d'erreurs de contour, le respect de la taille des intervalles et de la tonalité.

L'élaboration de cette méthode objective d'évaluation a été guidée par les règles du système tonal de la culture occidentale et les éléments musicaux perçus théoriquement par des auditeurs. Nous avons pu observer de nombreux avantages, tels que la rapidité, la précision, l'adaptabilité et la pertinence de l'évaluation, mais également des limites, concernant notamment le choix des critères musicaux à quantifier.

## ***Avantages et limites des outils d'analyse***

### **Rapidité de l'évaluation**

En ce qui concerne le temps nécessaire à l'analyse d'une performance vocale, nous n'avons pas de point de comparaison mais la méthode exposée nous a permis d'analyser deux bases de données conséquentes. En effet, l'étude 2 a nécessité l'analyse de 166 mélodies de chanteurs occasionnels (<http://sldr.org/sldr000774/en>) et l'étude 5, l'analyse de 200 mélodies de chanteurs professionnels (<http://sldr.org/sldr000792/en>). Au cours de la procédure d'analyse, seule la segmentation du signal se fait manuellement. Des programmes de segmentation existent (notamment par le logiciel Praat, Boersma,

2001 ; Boersma & Weenink, 2008) mais ils nécessitent la présence d'un silence ou d'une consonne entre deux sons. Ces conditions ne sont pas présentes dans des performances chantées avec des paroles. Par ailleurs, les options disponibles dans AudioSculpt (Ircam, Paris, France) facilitent la segmentation manuelle. En effet, outre la visualisation précise du signal, il est possible de zoomer ou encore de ralentir la vitesse de présentation (grâce à l'option "time stretching") tout en bénéficiant d'informations auditives pertinentes. La segmentation manuelle est donc une alternative permettant de limiter les erreurs de détection - et donc les multiples vérifications des données - sans être trop contraignante. Suite à la segmentation, les étapes d'analyse et de calcul des erreurs sont automatiques et permettent donc au chercheur de travailler sur un grand nombre d'échantillons vocaux. A titre informatif, la segmentation ne nécessite que 2-3 minutes pour des séquences de 20 secondes et les étapes suivantes sont quasi instantanées.

### **Efficacité de l'évaluation**

Au niveau du choix des logiciels d'analyse, le logiciel Praat (Boersma, 2001 ; Boersma & Weenink, 2008) est majoritairement utilisé dans les études relatives à la justesse vocale. Comme mentionné dans le chapitre II, plusieurs auteurs rapportent des erreurs de type "saut d'octave" (par exemple Dalla Bella et al., 2007 ; Granot et al., in press ; Pfordresher et al., 2010). Par ailleurs, une autre méthode utilisant l'algorithme de Yin (De Cheveigné & Kamahara, 2002) a jusqu'à présent été limitée à l'analyse de sons isolés (Hutchins & Peretz, 2012). Notre méthode utilise les logiciels proposés par l'Ircam et nous avons pu constater que les analyses effectuées (études 1, 2, 3, 4, 5 et 7) n'ont pas occasionné des erreurs telles que celles rencontrées par les utilisateurs du logiciel Praat concernant l'extraction de la fréquence fondamentale. Le choix de la fenêtre d'analyse mais également la possibilité de « masquer » des bruits ou harmoniques générées par l'environnement ou le matériel d'enregistrement limitent en effet les erreurs décrites dans de précédentes études sur la justesse vocale. De plus, AudioSculpt (Ircam, Paris, France) permet l'analyse des accords de chaque note et le programme créé dans OpenMusic (Ircam, Paris, France) repère

systématiquement la note la plus basse de l'accord, évitant ainsi de considérer des harmoniques particulièrement intenses.

### **Précision de l'évaluation**

Outre l'intérêt de segmenter manuellement la séquence mélodique dans le but d'éviter des erreurs de détection du signal, la segmentation manuelle a également l'avantage d'être précise. En effet, que ce soit pour des voix de chanteurs occasionnels où le signal acoustique n'est pas toujours stable (études 1, 2, 3 et 4) ou des voix lyriques où la présence de vibrato nécessite une découpe du signal respectant le cycle vibratoire (études 5 et 7), la segmentation manuelle proposée par AudioSculpt (Ircam, Paris, France) permet de déterminer précisément le segment à moyenniser pour chaque note. Les données recueillies sont disponibles en cents (un cent étant un centième de demi-ton) et procurent ainsi une unité de mesure fine et adaptée à l'analyse de performances vocales. En effet, comme mentionné dans le premier chapitre, l'instrument vocal permet de chanter de manière continue, contrairement aux instruments à claviers où le plus petit intervalle est le demi-ton. Notons que les études menées sur la justesse vocale ont utilisé des seuils de 50 cents, c'est à dire un quart de ton (Brattico et al., 2006 ; Hutchins & Peretz, 2012), ou de 100 cents, c'est à dire un demi-ton (Pfordresher & Brown, 2009 ; Pfordresher et al., 2010). Ces seuils constituent, pour ces auteurs, la limite à partir de laquelle l'intervalle est considéré comme faux. Alors que le seuil de 50 cents fait référence à la discrimination des hauteurs et que celui de 100 cents se base sur la composante musicale (il s'agit du plus petit intervalle), nous avons pris le parti de considérer la déviation en unité centésimale afin de garantir la précision des mesures.

Ces déviations sont alors calculées sur la base du tempérament égal (*cf.* Chapitre I). Bien que ce tempérament constitue une solution purement mathématique qui n'est pas congruente par rapport aux harmoniques naturelles, nous avons choisi cette référence, où tous les demi-tons sont égaux, pour différentes raisons. Premièrement, il s'agit d'un système utilisé depuis le XVIII<sup>ème</sup> siècle et largement répandu à l'heure actuelle. Deuxièmement, tout comme les règles musicales sont implicitement apprises par simple exposition (Trainor, 2005 ; Trainor et al., 2012), les auditeurs/chanteurs, habitués au tempérament égal, percevraient/produiraient les relations mélodiques sur base de cette

référence (Pfordresher & Brown, 2009 ; Pfordresher et al., 2010), d'autant plus lorsqu'ils ont bénéficié d'un entraînement musical (Zatara et al., 2012). En ce qui concerne les chanteurs professionnels, nous avons pris soin de sélectionner des participants ayant suivi une formation classique et se produisant régulièrement accompagnés (d'un ensemble instrumental de type orchestre classique ou d'un piano) afin d'éviter que les chanteurs ne se soient spécialisés dans le domaine de la musique ancienne où le tempérament égal n'est pas appliqué. Troisièmement, les performances analysées étaient soit une mélodie du répertoire populaire (études 1, 2, 3, 5 et 7), soit une mélodie non connue (études 4 et 5) mais composée sur le même modèle (centre tonal fort marqué, peu de modulations, cadence finale) et donc peu associé à de la musique ancienne. Notons qu'Howard (2007a, 2007b) a observé qu'en contexte de chœur, des chanteurs entraînés s'harmonisent en suivant un tempérament naturel. Dans le cadre de cette thèse, nous avons focalisé notre attention sur des performances en contexte solo, c'est à dire que les chanteurs occasionnels ou professionnels n'avaient pas à s'harmoniser avec d'autres voix. Au vu de ces arguments, l'utilisation d'un tempérament égal semble le plus approprié pour l'évaluation objective de la justesse vocale de chanteurs occasionnels ou professionnels. Cependant, comme nous le développons par la suite, il serait tout à fait envisageable de modifier les paramètres encodés dans le tableau d'analyse afin de définir une référence basée sur un autre tempérament.

### **Adaptabilité de la méthode**

En plus d'être rapide, efficace et précise, la méthode présentée a l'avantage d'être adaptable étant donné qu'elle permet d'analyser toute mélodie prédéfinie. En effet, la segmentation manuelle permet le découpage de toute mélodie et les logiciels AudioSculpt et OpenMusic (Ircam, Paris, France) fonctionnent quel que soit le matériel musical. De même, pour le calcul des erreurs, le tableau et ses formules sont aisément utilisables quelle que soit la mélodie de référence. En effet, il s'agit simplement d'encoder les données musicales de la partition de référence et les formules s'adaptent aux nouvelles données. Dans le cadre de cette thèse, nous avons ainsi pu évaluer objectivement, selon la même procédure, une mélodie populaire (études 1, 2, 3, 5 et 7) et



d'autres mélodies (études 4 et 5). Notons que l'utilisation de ces outils pourrait être élargie à l'analyse des performances mélodiques d'autres instruments que la voix, à la condition de modifier les paramètres d'analyse selon les caractéristiques acoustiques des instruments considérés. Il en est de même si le choix porte sur l'utilisation d'autres tempéraments ou, plus généralement, sur des règles appartenant à d'autres cultures. La méthode d'analyse précédemment décrite permet donc une analyse aisée et efficace du signal acoustique, que ce soit pour des voix de chanteurs occasionnels ou des échantillons vocaux plus complexes, comme des voix de chanteurs lyriques.

### ***Avantages et limites des critères d'évaluation***

En ce qui concerne les critères musicaux observés, le chapitre II mentionne l'utilisation quasi exclusive par la communauté scientifique de la déviation des intervalles par rapport à la mélodie de référence (Dalla Bella et al., 2007, 2009 ; Pfordresher et al., 2010). À notre connaissance, Dalla Bella et al. (2007, 2009) sont les seuls à mentionner la prise en compte de modulations. Cependant, leur mesure ne s'applique que lors de la répétition d'une séquence mélodique au sein d'une chanson et se trouve donc limitée. En effet, toute mélodie n'a pas de « refrain » et le centre tonal peut être dévié de manière régulière tout au long de la performance. D'un point de vue méthodologique, les résultats de l'étude 2 suggèrent la nécessité de considérer la variable relative au respect de la tonalité afin que des outils informatiques reflètent l'évaluation de juges experts. De plus, nous avons pu observer, à l'occasion de l'étude 4, que les critères de justesse (intervalles, tonalité) peuvent évoluer différemment. En effet, les étudiants de première année, pour lesquels la condition stressante avait un caractère formatif, étaient plus précis en ce qui concerne les intervalles, tandis que le centre tonal n'était pas plus dévié en situation de stress qu'en situation calme. Le profil inverse apparaît chez les étudiants de deuxième année, pour qui l'enjeu de la situation stressante était particulièrement important. Pour ces étudiants, la précision des intervalles n'était pas significativement différente entre la situation calme et stressante, alors que le centre tonal était davantage dévié en situation stressante qu'en situation calme. Il semble donc intéressant de

distinguer ces deux types d'erreur pour l'évaluation de la justesse mélodique de chanteurs occasionnels.

L'étude 2 met également en évidence que les erreurs de contour n'entrent pas en jeu dans l'évaluation de juges experts. Il est vrai que la mélodie qui a été analysée objectivement et évaluée par le groupe d'experts (Joyeux Anniversaire) n'induisait que peu d'erreurs de ce type et qu'il serait intéressant de proposer une mélodie plus complexe. Cependant, la création d'un large corpus nécessite que de nombreux chanteurs occasionnels soient en mesure de chanter la même mélodie. Le choix du matériel musical est alors ardu. Par ailleurs, l'observation du développement de la justesse chez l'enfant montre que le respect du contour mélodique apparaît très tôt (Dowling, 1999 ; Flowers & Dunne-Sousa, 1990 ; Ostwald, 1973 ; Welch, 2006) ce qui pourrait expliquer le peu d'erreurs à ce niveau. Cependant, il nous paraît intéressant de considérer le nombre d'erreurs de contour, non comme critère de justesse, mais à titre informatif, c'est à dire pour nous assurer que la représentation du schéma mélodique est correcte.

Nous avons donc focalisé notre attention sur la déviation des intervalles (déjà exploitée par Dalla Bella et al., 2007, 2009 ; Pfordresher et al., 2010) et sur le critère relatif à la tonalité. Ces deux critères sont importants dans le système tonal occidental et sont pris en compte lorsque l'évaluation est effectuée par des juges experts (étude 2). Comme nous le développons dans la partie consacrée à la perception de la justesse, le cas des voix lyriques est discutable. En effet, les données de l'étude 1 confirment le manque de justesse des chanteurs lyriques (Sundberg et al., 1996 ; Vurma & Ross, 2006) et ce, quelle que soit la mélodie interprétée (étude 5). Cependant, l'objectif d'une méthode d'évaluation est de tenir compte des critères saillants pour les auditeurs. Dans le cas des voix lyriques, le signal acoustique est particulièrement riche (étude 6) et les auditeurs se servent de multiples informations acoustiques pour évaluer la justesse d'une performance (étude 7). Alors que notre méthode objective d'évaluation de la justesse vocale en contexte mélodique reste à affiner pour évaluer la justesse de voix lyriques, elle semble tout à fait judicieuse pour l'évaluation de la justesse des performances de chanteurs occasionnels.

### ***Piste à explorer : seuils de justesse***

Pour notre méthode, nous avons choisi de mesurer la justesse vocale par la valeur moyenne de la déviation comparée à la partition musicale de référence. Ce choix s'est fait pour deux raisons. D'une part, l'instrument vocal permet de chanter de manière continue et les erreurs peuvent se situer entre les notes, contrairement aux instruments à clavier. D'autre part, comme mentionné dans le chapitre II, le choix d'un seuil à partir duquel la performance est considérée comme "fausse" est problématique. En effet, si la norme est basée sur un seuil statistique, il va dépendre de la population sur laquelle porte l'analyse. Par ailleurs, l'utilisation d'un seuil d'un quart de ton (Brattico et al., 2006 ; Hutchins & Peretz, 2012), ou d'un demi-ton (Pfordresher & Brown, 2009 ; Pfordresher et al., 2010) se fonde sur des tâches de discrimination en contexte isolé, de prise en compte, en contexte mélodique, de la seule dernière note (par exemple, Hutchins et al., 2012 ; Warrier & Zatorre, 2002) ou sur la valeur minimale d'un écart entre deux notes de musique. Néanmoins, ces choix sont discutables (*cf.* Chapitre II) et l'introduction de ces seuils – ou d'autres – dans notre méthode nécessite des études complémentaires. Ces études pourraient permettre la détermination de seuils objectifs de justesse, dans un contexte donné. Notons que notre méthode laisse la possibilité d'intégrer ces seuils une fois qu'ils auront été définis. Par ailleurs, ces études, comportementales, pourraient être complétées par des mesures au niveau neurophysiologique avec, par exemple, l'observation de l'activité cérébrale par électro-encéphalographie, lors de détection d'erreurs de justesse en contexte mélodique. En effet, comme l'ont remarqué François et Schön (2010) au sujet de l'apprentissage implicite de séquences langagières et musicales, les réponses électro-physiologiques montrent une meilleure sensibilité que les études comportementales.

### **Pertinence d'une méthode objective**

L'évaluation de la justesse vocale en contexte mélodique, peu étudiée en tant que telle, a de multiples implications. En effet, que ce soit dans un contexte pédagogique, pour évaluer de manière adéquate une performance vocale en évitant les biais liés aux juges (Chapitre II), ou dans un contexte clinique, pour cibler une difficulté (Chapitre III), des

outils d'évaluation pertinents sont nécessaires. Avant de détailler les applications possibles de la méthode élaborée dans le cadre de cette thèse, nous discutons de sa pertinence.

### ***Pertinence de notre méthode d'évaluation***

Pour être applicables, des outils d'évaluation doivent remplir certaines conditions relatives à leur manipulation. Sur le plan technique, nous avons constaté, à l'occasion des études 1, 2, 3, 4, 5 et 7, que les logiciels proposés par l'Ircam sont particulièrement pratiques et appropriés pour évaluer les mélodies chantées par des chanteurs occasionnels et professionnels. De plus, cette méthode est adaptable et permet ainsi l'évaluation de différents matériels mélodiques.

Par ailleurs, la pertinence d'une méthode d'évaluation tient également aux informations qu'elle procure. Avec l'étude 1, nous avons vérifié la pertinence de ces outils de manière globale en confirmant l'effet de l'entraînement vocal sur la justesse d'une performance mélodique (Amir et al., 2003 ; Dalla Bella et al., 2007 ; Murry, 1990 ; Watts et al., 2003) ainsi que la relation attendue entre justesse et tempo chez les chanteurs occasionnels (Dalla Bella et al., 2007, 2009). De plus, vous avons pu observer que l'évaluation d'un matériel mélodique a des avantages qu'une tâche de répétition ne peut offrir. Nous avons mentionné dans l'introduction théorique que nous répétons rarement des sons isolés dans nos activités quotidiennes alors que chanter des séquences mélodiques apparaît très tôt (Dowling, 1999 ; Ostwald, 1973 ; Welch, 2006) et est une activité particulièrement appréciée (Chong, 2010). Par ailleurs, pour évaluer la justesse d'une performance vocale, les résultats de l'étude de 2 suggèrent qu'il est important de tenir compte de deux des règles musicales mentionnées au chapitre I, à savoir, les erreurs relatives à la précision des intervalles et à la tonalité. De plus, l'étude 4, destinée à observer l'effet du stress sur la justesse d'une performance vocale, confirme l'intérêt de tenir compte de ces deux types d'erreur étant donné qu'elles peuvent évoluer différemment en fonction du contexte de l'évaluation et de l'enjeu de la performance. Cependant, ces deux types d'erreur ne peuvent être observés qu'en contexte mélodique puisqu'il s'agit d'observer la relation entre les notes. Ces arguments (aspect naturel d'une performance vocale et utilité

d'observer le respect des intervalles et de la tonalité) sont donc en faveur de l'utilisation de mélodies complètes pour évaluer la capacité à chanter juste. Alors que peu d'auteurs ont objectivement analysé les mélodies complètes de chanteurs occasionnels (Dalla Bella & Berkowska, 2009 ; Dalla Bella et al., 2007), la majorité des études qui évaluent la justesse de performance vocales se basent sur des tâches de répétition de sons isolés ou de courtes séquences mélodiques (Amir et al., 2003 ; Granot et al., in press ; Hutchins & Peretz, 2012 ; Moore et al., 2007, 2008 ; Nikjeh et al., 2009 ; Pfordresher & Brown, 2007, 2009 ; Pfordresher et al., 2010 ; Watts et al., 2002). Nous avons souhaité observer, dans l'étude 3, si les scores obtenus à une tâche de répétition permettent d'estimer la capacité d'un individu à respecter la précision des intervalles et de la tonalité. Autrement dit, nous avons comparé les scores obtenus à la fois lors d'une tâche de répétition et lors de performances mélodiques. Les résultats mettent en évidence que seul le critère relatif aux intervalles est corrélé au score obtenu à la tâche de répétition vocale (les chanteurs répétant précisément des sons isolés respectaient mieux la taille des intervalles en contexte mélodique). Etant donné que nous n'avons pas observé de relation entre le respect de la tonalité et la capacité à répéter adéquatement un son isolé, nous déduisons que l'évaluation d'une performance mélodique procure des informations plus complètes quant à la capacité de justesse d'un chanteur occasionnel. Comme indiqué dans le chapitre III, les études visant à décrire le développement de la justesse chez l'enfant se basent majoritairement sur des tâches de répétition et nous pensons que notre méthode d'évaluation permettrait une meilleure compréhension du développement de la justesse vocale.

### ***Applications en pédagogie musicale***

Si l'évaluation des productions vocales à l'occasion de fêtes d'anniversaire peut avoir un intérêt sociologique, les résultats d'une évaluation peuvent également avoir un enjeu au niveau professionnel. En effet, dans le cadre de la formation des enseignants et des futurs musiciens professionnels, la capacité à chanter juste est évaluée régulièrement, dans un but formatif ou certificatif. Or, de nombreux facteurs non musicaux influencent l'évaluation d'une performance musicale (Chapitre II) et la réussite ou l'échec aux épreuves musicales

ne peuvent être exclusivement expliqués par la qualité de la performance.

Parmi les facteurs non musicaux influençant l'évaluation effectuée par des juges, les éléments relatifs au musicien (par exemple, l'apparence ou la gestuelle) ont été largement décrits (Behne & Wöllner, 2011 ; Davidson & Edgar, 2003 ; Elliott, 1996 ; Griffiths, 2008, 2010 ; Howard, 2012 ; Juchniewicz, 2008 ; Kurosawa & Davidson, 2005 ; Livingstone et al., 2009 ; Ryan, & Costa-Giomi, 2004 ; Wapnick et al., 1997, 1998, 2000). Afin d'éviter l'influence de ces facteurs, l'évaluation d'une performance musicale à partir d'un enregistrement semble alors judicieuse (comme illustré Figure 1, page 212). Par ailleurs, même lorsque l'évaluation est effectuée sur la base d'enregistrements, c'est à dire sans la présence physique du musicien, d'autres facteurs interviennent dans l'évaluation d'une performance musicale. En effet, nous avons détaillé dans le chapitre II que les juges sont sensibles au matériel musical présenté (par exemple, Flores & Ginsburgh, 1996 ; Glejser & Heyndel, 2001 ; Hutchins et al., 2012 ; Kinney, 2009 ; Wapnick et al., 2005 ; Warrier & Zatorre, 2002) que ce soit par leurs préférences ou du fait de leur perception mélodique (effet des intervalles, du contexte, du timbre, etc.). De plus, les jugements vont dépendre de leurs attentes quant au niveau (débutant versus confirmé) des participants à évaluer (Cavitt, 1997 ; Duerksen, 1972).

Dans un contexte pédagogique, l'évaluation objective des performances mélodiques aurait un intérêt certain pour éviter les facteurs précités (comme illustré Figure 5, page 38). De plus, l'étude 2 met en évidence que trois juges experts suffiraient à évaluer la précision des intervalles sur la base d'un enregistrement mais une évaluation globale ne permet pas de distinguer les erreurs relatives aux intervalles et à la tonalité qui pourtant peuvent évoluer différemment (étude 4). L'évaluation globale effectuée par les jurys d'examen pourrait donc être avantageusement secondée par l'utilisation de notre méthode d'évaluation. En effet, elle permettrait de quantifier de manière distincte différents types d'erreur et de guider ainsi les enseignants dans la formation des élèves.

## ***Applications en clinique***

La recherche des causes d'un déficit de justesse s'est largement développée ces dernières années. Parmi les troubles décrits, nous avons détaillé les problèmes de perception, de mémoire, les troubles sensorimoteur et moteur (*cf.* Chapitre III). Cependant, diagnostiquer un trouble nécessite une évaluation précise. Comme exposé au chapitre II, l'évaluation subjective de performances vocales est encore utilisée (par exemple, Alcock et al., 2000a, 2000b ; Hébert et al., 2003 ; Kinsella et al., 1988 ; Lévêque et al., 2012 ; Prior et al., 1990 ; Racette et al., 2006 ; Schön et al., 2004 ; Wise & Sloboda, 2008). Pour catégoriser un chanteur occasionnel en "bon" ou "mauvais", les juges de ces études utilisent des grilles d'évaluation, suivent des consignes quant aux erreurs à repérer ou encore remettent un score global. Comme l'indique l'étude 2, des juges experts tiennent compte, dans leur évaluation, de la précision des intervalles et du respect de la tonalité. Cependant, la catégorisation du chanteur en "bon" ou "mauvais" ne permet pas de cibler et de quantifier le problème qui, comme le montre l'étude 4, peut concerner un des deux types d'erreur. Par ailleurs, lorsque les chercheurs utilisent une méthode objective pour définir la justesse d'une performance, elle est principalement appliquée à des tâches de répétition. Or, comme nous avons pu le constater dans l'étude 3, les informations recueillies sont partielles étant donné que le respect de la tonalité n'est pas renseigné. La solution semble donc d'évaluer des mélodies complètes, comme l'ont proposé par exemple Dalla Bella et al. (2007). Cependant, leurs mesures objectives ne tiennent pas compte des erreurs relatives à la tonalité tout au long de la mélodie. Au vu de ces études et des limites soulevées, nous estimons que notre méthode objective d'évaluation a une utilité dans le domaine clinique. En effet, elle permettrait de cibler précisément un trouble de justesse. Ce type d'évaluation permettrait également d'objectiver les bénéfices d'une prise en charge d'un déficit de justesse. Ajoutons cependant que le contexte de l'évaluation influence la qualité de la production (étude 4) et qu'il conviendrait donc de prendre en compte l'état émotionnel du sujet pour interpréter correctement les données.

La méthode objective d'évaluation élaborée dans le cadre de cette thèse semble donc tout à fait pertinente et recommandée tant pour

obtenir des scores appropriés dans un contexte pédagogique que pour cibler les problèmes de justesse rencontrés par certains chanteurs occasionnels et ainsi pouvoir y remédier.

### ***Piste à explorer : facteurs influençant l'évaluation***

Comme exposé dans le chapitre relatif à l'évaluation de la justesse vocale, une performance musicale est un événement qui fait intervenir de multiples éléments ayant une incidence sur l'évaluation de la performance (comme illustré Figure 5, page 38). Les facteurs peuvent être relatifs au musicien, à la performance ou encore au juge (Godlovitch, 1998 ; Landy & Farr, 1980 ; McPherson & Schubert ; 2004 ; McPherson & Thompson, 1998). Depuis ces modèles et ces revues de littérature, de nombreuses études ont porté sur l'influence de ces différents facteurs sur l'évaluation de la performance globale. En ce qui concerne l'évaluation de la justesse vocale, aucune étude, à notre connaissance, n'a permis d'observer l'influence de tels facteurs. Avec l'étude 2, nous avons pu observer le comportement des juges face à une performance dont nous connaissions la qualité, étant donné que les critères étaient prédéfinis (d'après les règles du système tonal occidental) et que nous étions en mesure de quantifier (à l'aide de nos outils) le respect de ces critères au cours des performances. Autrement dit, nous avons résolu une équation à une inconnue puisque les variables "performance" et "critères" étaient connues. Maintenant que les variables "performance", "critères" et "juge" sont mesurées, il ne s'agirait plus que de faire intervenir différents facteurs - relatifs aux juges, à la performance ou encore au musicien - et d'apprécier l'évolution de l'évaluation effectuée par les juges. La poursuite de telles recherches, à l'aide de notre méthode objective d'évaluation, permettrait donc de quantifier les facteurs en jeu dans l'évaluation de la justesse vocale et ainsi de construire un modèle d'évaluation de la justesse.



## Vers une définition de la justesse vocale

Ce travail de thèse débute par une citation de Jean-Philippe Rameau (1683-1764). Ce compositeur du XVIII<sup>ème</sup> siècle a décrit, dans son traité sur l'harmonie musicale, l'intérêt d'utiliser les mathématiques pour étudier les principes qui dictent les règles musicales. Cette citation pourrait être transposée au thème de la justesse. Dans le cadre de nos recherches, nous avons utilisé les mathématiques dans le but de mieux comprendre les règles d'évaluation de la justesse, que les juges soient confrontés à des voix de chanteurs occasionnels ou lyriques. En effet, nous avons quantifié différents critères de justesse grâce à l'analyse du signal acoustique des voix, afin de comparer ces données aux évaluations effectuées par des juges et de déduire les règles qui les guident. Par ailleurs, comme indiqué dans le chapitre II, de nombreux facteurs non musicaux influencent l'évaluation d'une performance musicale (Godlovitch, 1998 ; Landy & Farr, 1980 ; McPherson & Schubert, 2004 ; McPherson & Thompson, 1998). De ce fait, une situation d'évaluation en face à face, normale dans la majorité des cas, complexifie l'observation du comportement des juges lorsqu'ils évaluent la justesse de performances vocales. Nous avons donc contrôlé l'intervention de ces facteurs, dans les études 2 et 7, en présentant des versions enregistrées des performances à évaluer.

Qu'une mélodie soit chantée par un chanteur occasionnel ou un chanteur lyrique, il s'agit toujours d'une succession de notes obéissant aux conventions et contraintes culturelles (Ringer, 2002). Dans le système tonal occidental, l'organisation des sons suit des règles précises (Cross, 2001 ; Lerdahl & Jackendoff, 1983) et la succession des notes se fait donc suivant une certaine logique (voir Figure 2, page 21) engendrant des attentes de la part de l'auditeur (Marmel et al., 2008). Ces attentes semblent partagées par les auditeurs d'une même culture étant donné le développement de la perception mélodique et la capacité humaine à apprendre implicitement un système de règles musicales (Jonaitis & Saffran, 2009 ; Loui et al., 2010). De ce fait, un auditeur intègre et utilise les informations relatives au contour mélodique (Dowling & Fujitani, 1970 ; Edworthy, 1985), à la tonalité (Trainor & Trehub, 1992) et aux intervalles (Stalinski et al., 2008) sans apprentissage explicite. Cependant, le fait que des auditeurs perçoivent

des erreurs ne nous renseigne pas sur deux points : le partage de la notion de justesse entre les auditeurs et l'importance des erreurs de justesse dans l'évaluation d'une performance. En effet, les attentes musicales sont partagées par des auditeurs mais les éventuelles variations interindividuelles quant à la définition de la justesse vocale n'ont pour l'instant pas été observées. Autrement dit, nous ne savons pas si des auditeurs perçoivent les erreurs de justesse de la même manière. De plus, l'importance des erreurs n'a pas été examinée et nous ne disposons pas de données relatives à la quantification des différentes erreurs observables dans le système tonal occidental.

### ***Capacité à évaluer la justesse mélodique***

En ce qui concerne l'évaluation des performances produites par des chanteurs occasionnels, l'étude 2 nous indique que des sous-groupes de trois juges suffisent à conserver un coefficient de corrélation élevé entre les mesures objective et subjective. Outre l'intérêt au niveau de la relation entre la déviation des intervalles et la note globale des juges, ce résultat suggère que l'ensemble des juges évalue de manière similaire la justesse des performances de chanteurs occasionnels.

En ce qui concerne l'évaluation des performances produites par des chanteurs lyriques, nous avons opté pour l'utilisation d'un paradigme de comparaison par paires. En effet, plusieurs études ont pointé la difficulté d'obtenir un consensus entre les juges lorsqu'il s'agit d'évaluer ce type de voix (Elkholm et al., 1998 ; Garnier et al., 2007 ; Howes et al., 2004). La raison majeure réside dans la complexité de ces voix, comme l'illustre l'étude 6. La modélisation de cette technique est certes à compléter mais cette étude confirme que toutes les caractéristiques musicales et acoustiques observées contribuent à la définition d'une voix lyrique. Alors que Kacha et al. (2005) ont observé que la comparaison par paires facilite le jugement des voix dysphoniques, nous avons donc souhaité appliquer cette méthode pour favoriser l'accord entre les juges experts de l'étude 7. Les méthodes utilisées dans les études 2 (voix occasionnelles) et 7 (voix lyriques) sont donc différentes et, de ce fait, non directement comparables. Cependant, les résultats vont dans le même sens. En effet, dans l'étude 7, seuls deux des 22 juges donnent des classements différents d'une évaluation à

l'autre (test et retest) et les 20 autres juges évaluent de manière similaire la justesse des voix lyriques proposées.

Plusieurs études avaient mis en évidence un manque de concordance entre les juges lors de l'évaluation de performances musicales (Kinsella et al, 1988 ; Prior et al., 1990 ; Thompson & Williamon, 2003). En ce qui concerne la justesse vocale, nos données confirment la convergence entre les juges, observée par Racette et al. (2006). Comme mentionné dans le chapitre II, une évaluation peut être influencée par des facteurs relatifs aux juges et à leur perception de la performance. Dans notre cas, les juges devaient évaluer la mélodie "Joyeux Anniversaire" interprétée par 166 chanteurs occasionnels ou comparer 14 extraits de cette mélodie. Bien que cette mélodie soit fréquemment déformée, les résultats suggèrent que les juges en ont une représentation similaire.

Par ailleurs, nous avons distingué différents types d'entraînement (expertise vocale versus instrumentale) dans nos critères de recrutement. Alors que différentes études ont mis en évidence que le type d'expertise a une influence sur l'apprentissage statistique ou encore sur la perception (Larrouy-Maestri et al., 2013 ; Seppänen et al., 2007) et donc sur les connaissances implicites de la culture occidentale et sur la perception des erreurs mélodiques, nous pouvions imaginer qu'une expertise spécifique au domaine musical permettrait une meilleure concordance. Par exemple, concernant la spécificité de l'expertise, Sofranki et Prosek (2012) ont observé la concordance de logopèdes, de professeurs de chant et de sujets non experts dans une tâche de jugement de la qualité vocale. Ils ont remarqué des différences entre les groupes, avec une meilleure concordance pour les experts en voix dysphoniques (logopèdes), une concordance modérée pour les professeurs de chant mais toutefois plus élevée que celle des non experts. En ce qui concerne l'évaluation de la justesse vocale, les études 2 et 7 suggèrent donc que le type ou le niveau d'expertise n'influence pas le jugement global de performances chantées par des chanteurs occasionnels ou lyriques, lorsqu'elles sont présentées en modalité auditive seule, via des enregistrements.

Il apparaît donc que l'évaluation de performances vocales n'est ni dépendante du juge ni de son type d'expertise. Notons que la piste

proposée précédemment, au sujet de la recherche des facteurs influençant l'évaluation de la justesse, pourrait être instructive à ce niveau. En effet, des études, incluant progressivement des facteurs non musicaux afin d'observer leur effet sur l'évaluation effectuée par des juges, permettraient également de clarifier les éléments favorisant ou non l'accord inter-juges.

### ***Notion de justesse mélodique***

L'accord entre les juges, observé lors de l'évaluation de voix occasionnelles et lyriques, suggère que la notion de justesse est partagée par les juges experts. Nous avons dès lors souhaité observer les critères utilisés par l'ensemble des juges. Comme indiqué au début de cette discussion, le point de départ de nos recherches a été l'élaboration d'une méthode d'évaluation objective qui tienne compte de critères pertinents, tant au niveau des règles musicales dictées par le système tonal occidental que de la perception mélodique qu'un auditeur en a. Le chapitre I détaille les éléments à la base de notre choix et la prise en compte des erreurs relatives aux contours mélodiques, à la taille des intervalles et à la tonalité, dans la conception de notre méthode. Notons que pour les voix lyriques, étant donné les contraintes techniques liées à l'utilisation d'un paradigme de comparaison par paires, nous avons sélectionné de courtes séquences mélodiques (dernière phrase de Joyeux Anniversaire). En effet, cette méthode nécessite d'associer chaque séquence mélodique avec toutes les autres, ce qui allonge considérablement la durée de la tâche. Pour cette même raison, nous avons dû limiter la durée des séquences mélodiques et sélectionner le critère de justesse le plus pertinent dans ce cas, c'est à dire la précision des intervalles (la mesure des erreurs de modulation nécessite de plus longues séquences). Parallèlement au critère de précision, nous avons analysé les caractéristiques acoustiques et musicales qui se sont avérées importantes pour décrire le profil d'une voix lyrique (étude 6).

Les résultats de l'étude 2 montrent l'importance du respect de la taille des intervalles et de la tonalité lors de l'évaluation des voix des chanteurs occasionnels par les juges. Alors que le critère relatif à la taille des intervalles est régulièrement utilisé (Dalla Bella et al., 2007, 2009 ; Pfordresher et al., 2010), cette étude confirme que le critère relatif à la

tonalité est bien perçue (Trainor & Trehub, 1992) et a tout son sens (Flowers & Dunne-Sousa, 1990 ; Price, 2000) pour l'évaluation d'une mélodie complète. Notons que les performances à évaluer comportaient peu d'erreurs de contours, ce qui pourrait expliquer que les juges n'en tiennent pas particulièrement compte, même s'ils les perçoivent. Autrement dit, si une erreur de contour apparaît de manière isolée, elle n'affecterait pas la note finale. Ce phénomène n'a pas été décrit dans la littérature mais il est aisé d'imaginer que l'auditeur serait plus tolérant pour une seule erreur "flagrante", plutôt que pour de nombreuses "petites" erreurs au cours de la performance musicale. Ces conclusions, à confirmer à l'aide d'une étude spécifique comparant les erreurs de contours et les erreurs relatives aux intervalles et à la tonalité, pourraient encourager les chanteurs occasionnels à minimiser les conséquences d'erreurs de contour et à se concentrer sur la précision de la ligne mélodique à interpréter. Par ailleurs, l'étude 2 indique que 19 % de la variance des juges n'est pas expliquée par les erreurs de justesse objectivement mesurées. Malgré leur expertise, il semble donc que les juges ont des difficultés à se focaliser essentiellement sur la justesse vocale. D'autres facteurs, tels que la qualité vocale (Butte et al., 2009) ou encore les caractéristiques musicales (Prince, 2011), qu'il conviendrait de définir, pourraient intervenir de manière isolée ou en interaction (comme pour l'étude 7) dans la définition de la justesse. Cependant, ce pourcentage reste limité et nous observons que deux critères (précision des intervalles et respect de la tonalité) expliquent à eux seuls 81 % de la note globale donnée par les juges.

La définition de la justesse est plus complexe quand il s'agit d'évaluer des voix lyriques. Comme détaillé au chapitre IV, une voix lyrique est une combinaison complexe de différents paramètres acoustiques et l'étude 6 confirme que des caractéristiques musicales (hauteur de la performance, tempo, niveau de pression sonore) et vocales (balance spectrale, amplitude et fréquence du vibrato), entrent en jeu dans la définition de cette technique vocale. Ces caractéristiques peuvent affecter la perception de la fréquence fondamentale, comme par exemple le vibrato (Hutchins et al., 2012 ; van Besouw & Howard, 2008) ou la balance spectrale (Hutchins & Peretz, 2012 ; Russo & Thompson, 2005 ; Vurma et al., 2010 ; Warrier & Zatorre, 2002). Ces éléments rendent donc la perception du signal particulièrement complexe.

Comme stipulé précédemment, nous avons focalisé notre attention sur la précision des intervalles au sein de courtes séquences pour analyser de manière objective la justesse vocale. Notons que ce critère était particulièrement affecté par l'utilisation de la technique lyrique (étude 1), quel que soit le matériel musical analysé (étude 5). Les paramètres acoustiques précités ont également été observés et les résultats indiquent qu'ils contribuent tous à l'explication du classement des performances évaluées par les juges. Cependant, ces éléments apparaissent en interaction les uns avec les autres pour expliquer 78.8 % du modèle prédictif et sont donc difficilement dissociables. Cette étude utilise un matériel "naturel" étant donné que nous avons sélectionné des séquences vocales de chanteurs professionnels (base de données : <http://sldr.org/sldr000792/en>). Ce point constitue un frein à la compréhension de l'importance de chaque paramètre. Dans ce but, nous suggérons que de futures recherches se focalisent sur un matériel "naturel" mais dont les caractéristiques seraient manipulées isolément à l'aide d'outils de synthèse sonore, comme par exemple avec le logiciel AudioSculpt (Ircam, Paris, France), afin de contrôler leur ampleur et d'observer leurs effets sur la détection d'erreurs de justesse.

Outre ces informations quant à la perception globale des voix lyriques, cette étude remet en question la définition même de la justesse pour ces voix particulières. En effet, les résultats relatifs à la concordance des juges nous indiquent que la notion de justesse est partagée mais la relation entre les mesures objective et subjective n'est pas aussi claire que pour les voix de chanteurs occasionnels. Il semblerait donc que le terme "justesse" ne recouvre pas la même notion quand il s'agit de voix de chanteurs occasionnels et de voix de chanteurs lyriques. Rappelons que les éléments qui caractérisent une "bonne voix" (Garnier et al., 2007), à savoir le vibrato et le timbre, entrent en jeu dans le modèle explicatif du classement des chanteurs alors que nous avons clairement demandé, à l'occasion de chaque présentation de paires, de choisir la voix la plus juste. Une récente étude (Sundberg, La, & Himonides, in press) ouvre la voie à une nouvelle interprétation des erreurs de justesse en ce qui concerne les voix lyriques. En effet, ces auteurs ont proposé à des juges experts d'évaluer l'expressivité d'extraits mélodiques originaux (non modifiés) et de ces mêmes extraits dont ils ont modifié les variations de la fréquence fondamentale afin de

les rendre “justes”. Les résultats indiquent que les juges experts trouvent les extraits originaux plus expressifs que les extraits “justes”. Cependant, les auteurs remarquent que la manipulation du signal a pour conséquence une modification du timbre de la performance. Or, ce paramètre est particulièrement travaillé lors de la formation des chanteurs professionnels classiques (Brown et al., 2000 ; Lundy et al., 2000 ; Omori et al., 1996) et contribue à définir une voix lyrique (étude 6), par définition expressive. Il se pourrait donc qu’une modification du signal altère le timbre et donc la composante expressive de la voix du chanteur analysée. Les conclusions de Sundberg et al. (in press) nous indiquent donc la possibilité que des erreurs de justesse servent l’expressivité musicale mais cette hypothèse demande à être testée sur différentes voix et surtout avec un meilleur contrôle des caractéristiques acoustiques de la voix à évaluer.

En suivant la démarche de Jean-Philippe Rameau (1683-1764), nous avons donc pu explorer les règles d’évaluation de la justesse, que les juges soient confrontés à des voix de chanteurs occasionnels ou lyriques. Ces connaissances permettent de mieux comprendre les éléments acoustiques en jeu dans l’évaluation de la justesse vocale, de préciser la définition de la justesse et donc d’améliorer notre méthode objective d’évaluation de la justesse en contexte mélodique.

### ***Piste à explorer : expertise musicale***

La comparaison des méthodes objective et subjective nous a permis d’examiner le comportement de juges experts lorsqu’ils évaluent la justesse vocale de chanteurs occasionnels et professionnels. Cette population de juges a été choisie pour différentes raisons. D’une part, l’ensemble des études portant sur l’évaluation a examiné les jugements d’experts du domaine concerné (Chapitre II), le but étant de mieux comprendre les processus d’évaluation utilisés au cours de situations formatives et certificatives. D’autre part, les auditeurs experts semblent particulièrement à même d’évaluer la justesse d’une performance vocale. Nos études mettent bien en évidence que la notion de justesse est globalement partagée par les juges experts. Notons que, contrairement aux conclusions de Seppänen et al. (2007) ou encore Sofranko et Prosek (2012), le type d’expertise n’entre visiblement pas

en jeu dans l'évaluation de la justesse vocale en contexte mélodique et ce, quel que soit le type de voix (chanteurs occasionnels versus lyriques). En ce qui concerne le niveau d'expertise, l'étude 2 montre que l'évaluation de la justesse par des étudiants de Conservatoire est cohérente par rapport à celle de juges plus expérimentés. Alors que Kinney (2009) ou encore Morrison et al. (2004) observent un effet d'expertise sur l'évaluation de performances musicales, nos résultats soutiennent que des étudiants de Conservatoire sont à considérer comme des experts en musique même s'ils n'ont pas autant d'années d'expérience que les musiciens professionnels recrutés dans le cadre de nos recherches.

Cependant, nous ne sommes pas tous des experts en musique alors que nous sommes tous des juges potentiels. En effet, chanter est une activité particulièrement répandue et appréciée (Chong, 2010) et les occasions de porter un jugement sur une performance sont innombrables.

Les effets d'un entraînement musical sur le traitement d'un matériel musical et plus spécifiquement sur la perception mélodique sont nombreux (Chapitre I) : le bénéfice de l'expertise musicale se retrouve dans des tâches de discrimination de sons purs et complexes (Micheyl et al., 2006 ; Tervaniemi et al., 2005), dans les tâches de perception de hauteur de sons isolés (Hutchins & Peretz, 2012), d'estimation d'un intervalle (Russo & Thompson, 2005), de comparaison de sons de timbres différents en contexte isolé ou d'intervalles (Hutchins et al., 2012 ; Vurma et al., 2010 ; Zarate et al., 2012), de reconnaissance de mélodies (Orsmond & Miller, 1999), ou encore de détection d'erreurs (Fujiroka et al., 2004 ; Hutchins et al., 2012 ; Warrior & Zatorre, 2002). Si nous avons ici des réponses quant à l'évaluation effectuée par des juges experts, il est légitime de se demander si des individus n'ayant pas suivi d'entraînement musical seraient à même d'évaluer la justesse d'une performance vocale. A ce sujet, le chapitre I permet de conclure que l'exposition d'un individu à la culture environnante le rend expert de sa culture (Bigand & Delbé, 2010). Que ce soit par un apprentissage implicite (pour une revue, voir Patel, 2008) ou par l'activité même de chanter (Welch, 2005), l'enfant devient rapidement sensible aux règles du système musical dans lequel il évolue. Par ailleurs, les différentes études proposées chez l'adulte non expert



confirment que les informations relatives au contour mélodique, à la tonalité et aux intervalles sont acquises sans apprentissage explicite (Dowling & Fujitani, 1970 ; Edworthy, 1985 ; Stalinski et al., 2008 ; Trainor & Trehub, 1992). Sur la base de nos travaux, nous envisageons à présent d'élargir notre champ d'investigation à des auditeurs non experts. En effet, la comparaison de nos données avec celles recueillies auprès d'enfants et d'adultes non musiciens de notre culture et d'une culture dont le système musical est différent, permettrait d'examiner le processus d'enculturation, le développement de la perception mélodique et l'éventuel effet d'une expertise musicale sur l'évaluation de la justesse vocale en contexte mélodique.



# CONCLUSIONS

---

Les études menées dans le cadre de cette thèse nous ont permis de développer une méthode d'évaluation objective de la justesse d'une performance vocale en contexte mélodique ainsi que d'en mesurer la pertinence. Au niveau méthodologique, ce travail de thèse permet de nombreuses applications pratiques à destination des chercheurs, pédagogues et cliniciens intéressés par la justesse vocale en contexte mélodique. Par ailleurs, l'utilisation d'une méthode objective nous a permis d'aborder les aspects théoriques de l'évaluation de performances vocales. En effet, en contrôlant le contexte d'évaluation et en quantifiant la justesse des performances vocales à évaluer, nous avons pu observer que la notion de justesse est partagée par des juges experts, que les performances soient produites par des chanteurs occasionnels ou des chanteurs lyriques. Cependant, alors que deux critères sont importants pour évaluer la justesse de chanteurs occasionnels, la définition de la justesse vocale en contexte mélodique est plus complexe quand il s'agit de voix lyriques.

Ce travail de recherche ouvre donc la voie à des perspectives multiples, tant par l'application de cette méthode dans des contextes pédagogiques ou cliniques, qu'en recherche fondamentale. En effet, sur la base de notre méthode, les prochaines étapes permettront ainsi d'explorer le développement de la notion de justesse, l'influence de facteurs non musicaux sur sa perception mais également les processus sous-jacents à l'évaluation de performances musicales.







# **Références bibliographiques**





- Abel, J. L., & Larkin, K. T. (1990). Anticipations of performance among musicians: physiological arousal, confidence, and state-anxiety. *Psychology of Music, 18*, 171-182.
- Adachi, M. (2011). Effects of interactions with young children on Japanese women's interpretation of musical babbling. Proceedings of the Society for Music Perception and Cognition Meeting, Rochester, NY.
- Agresti, A. (1996). *An introduction to categorical data analysis*. New York: Wiley, Springer.
- Akiva-Kabiri, L., Vecchi, T., Granot, G., Basso, D., & Schön, D. (2009). Memory for tonal pitches – A music-length effect hypothesis. *Annals of the New York Academy of Sciences, 1169*, 266-269.
- Alcock, K. J., Passingham, R. E., Watkins, K., & Vargha-Khadem, F. (2000a). Pitch and timing abilities in inherited speech and language impairment. *Brain and Language, 75*(1), 34-46.
- Alcock, K. J., Wade, D., Anslow, P., & Passingham, R. E. (2000b). Pitch and timing abilities in adult left-hemisphere-dysphasic and right-hemisphere-damaged subjects. *Brain and Language, 75*(1), 47-65.
- Amir, O., Amir, N., & Kishon-Rabin, L. (2003). The effect of superior auditory skills on vocal accuracy. *The Journal of the Acoustical Society of America, 113*(2), 1102-1108.
- Amir, O., Amir, N., & Michaeli, O. (2005). Evaluating the influence of warmup on singing voice quality using acoustic measures. *Journal of Voice, 19*(2), 252-260.
- ANSI. (1994). *American national standard acoustical terminology*. New York: American National Standards Institute.

- Apfelstadt, H. (1984). Effects of melodic perception instruction on pitch discrimination and vocal accuracy of kindergarten children. *Journal of Research in Music Education*, 32(1), 15–24.
- Awan, S. N. (1991). Phonetographic profiles and F0-SPL characteristics of untrained versus trained vocal groups. *Journal of voice*, 5(1), 41-50.
- Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain*, 125(2), 238-251.
- Bailey, B. A., & Davidson, J. (2005). Effects of group singing and performance for marginalised and middle-class singers. *Psychology of Music*, 33(3), 269-303.
- Baken, R. J. (1987). *Clinical Measurement of Speech and Voice*. Boston, UK: Little, Brown and Company.
- Barlow, C., & Lovetri, J. (2010). Closed quotient and spectral measures of female adolescent singers in different singing styles. *Journal of Voice* 24, 314-318.
- Barlow, C., Lovetri, J., and Howard, D. (2007). Voice source and acoustic measures of girls singing “classical” and “contemporary commercial” styles. In A. Williamon, & D. Coimbra (Eds.). *Proceeding of the International Symposium of Performance Science* (pp. 195-200). Utrecht, The Netherlands.
- Barnes, J. J., Davis, P., Oates, J., & Chapman, J. (2004). The relationship between professional operatic soprano voice and high range spectral energy. *Journal of the Acoustical Society of America*, 116(1), 530-538.

- Behne, K. E., & Wöllner, C. (2011). Seeing or hearing the pianists? A synopsis of an early audiovisual perception experiment and a replication. *Musicae Scientiae*, 15(3), 324-342.
- Bergeson, T. R., & Trehub, S. E. (2002). Absolute pitch and tempo in mothers' speech and song to infants. *Psychological Science*, 13, 72-75.
- Berkowska, M., & Dalla Bella, S. (2009). Reducing linguistic information enhances singing proficiency in occasional singers. *Annals of the New York Academy of Sciences*, 1169, 108-111.
- Berkowska, M., & Dalla Bella, S. (2009). Acquired and congenital disorders of sung performance: A review. *Advances in cognitive psychology*, 5, 69-83.
- Bermudez de Alvear, R. M., Baron-Lopez, F. J., Alguacil, M. D., Dawid-Milner, M.S. (in press). Interactions between voice fundamental frequency and cardiovascular parameters. Preliminary results and physiological mechanisms. *Logopedics Phoniatics Vocology*.
- Besson, M., Chobert, J., & Marie, C. (2011). Transfer of Training between Music and Speech: Common Processing, Attention, and Memory. *Frontiers in Psychology*, 2, 94.
- Besson, M., & Faïta, F. (1995). An Event-Related Potential (ERP) study of musical expectancy: Comparison of musicians with nonmusicians. *Journal of Experimental Psychology: Human Perception and Performance*, 21(6), 1278-1296.
- Bigand, E. (1994). Contributions de la musique aux recherches sur la cognition auditive humaine. In S. McAdams & E. Bigand (Eds.), *Penser les sons : La psychologie cognitive de l'audition* (pp. 249-298). Paris : PUF.

- Bigand, E., & Delbé, C. (2010). L'apprentissage implicite de la musique occidentale. In R. Kolinsky, J. Morais, & I. Peretz (Eds), *Musique, Langage, Emotion : Approche neuro-cognitive* (pp. 35-47). Rennes, France : Presses Universitaires de Rennes.
- Björkner, E. (2008). Musical theater and opera singing - why so different? A study of subglottal pressure, voice source, and formant frequency characteristics, *Journal of Voice*, 22, 533-540.
- Bloothoof, G., & Plomp, R. (1985). Spectral analysis of sung vowels. The effect of fundamental frequency on vowel spectra. *Journal of the Acoustical Society of America*, 77(4), 1580-1588.
- Bloothoof, G. & Plomp, R. (1988). The timbre of sung vowels. *Journal of the Acoustical Society of America*, 84, 847-860.
- Boersma, M. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ration of a sampled sound. *Proceedings of the Institute of Phonetics Sciences, University of Amsterdam*, 17, 97-110.
- Boersma, P. (2001). Praat: a system for doing phonetics by computer. *Glott International*, 5, 341-345.
- Boersma, P., & Weenink, D. (2008). Praat: Doing phonetics by computer (Software Version 5.0.25). Available from: <http://praat.org/>
- Boersma, P., & Weenink, D. (2012). Praat: doing phonetics by computer [Computer program]. Version 5.1.44, retrieved 22 October 2010 from <http://www.praat.org/>
- Bond, V. L. (2012). Music's representation in early childhood education journal. A literature review. *Update: Applications of Research in Music Education*, 31(1), 34-43.

- Borch, D. Z., & Sundberg, J. (2011). Some phonatory and resonatory characteristics of the rock, pop, soul, and Swedish dance band styles of singing. *Journal of Voice*, *25*, 532-537.
- Bourne, T., & Garnier, M. (2012). Physiological and acoustic characteristics of the female music theatre voice. *Journal of the Acoustical Society of America*, *131*(2), 1586-1594.
- Bradshaw, E., & McHenry, M. (2005). Pitch discrimination and pitch matching abilities of adults who sing inaccurately. *Journal of Voice*, *19*(3), 431-439.
- Brattico, E., Tervaniemi, M., Näätänen, R., & Peretz, I. (2006). Musical scale properties are automatically processed in the human auditory cortex. *Brain Research*, *1117*(1), 162-174.
- Bretos, J., & Sundberg, J. (2003). Vibrato extent and intonation in professional Western lyric singing. *Journal of Voice*, *17*, 343-352.
- Brodsky, W. (1996). Music performance anxiety re-conceptualized: a critique of current research practices and findings. *Medical Problems of Performing Artists*, *11*(3), 88-98.
- Brotons, M. (1994). Effects of performing conditions on music performance anxiety and performance quality. *Journal of Music Therapy*, *31*, 63-81.
- Brown, W. S., Rothman, H. B., & Sapienza, C. M. (2000). Perceptual and acoustic study of professionally trained versus untrained voices. *Journal of Voice*, *14*(3), 301-309.
- Bunch, M., & Chapman, J. (2000). Taxonomy of singers used as subjects in scientific research. *Journal of Voice*, *14*, 363-369.
- Burns, E. M., & Ward, W. D. (1978). Categorical perception - phenomenon or epiphenomenon: Evidence from experiments in

- the perception of musical intervals. *Journal of the Acoustical Society of America*, 63, 456–468.
- Butte, C. J., Zhang, Y., Song, H., & Jiang, J. J. (2009). Perturbation and nonlinear dynamic analysis of different singing styles. *Journal of Voice*, 23, 647-652.
- Casals, A, Ayats, J., & Vilar, M. (2010). Improvised song in schools: breaking away from the perception of traditional song as infantile by introducing a traditional adult practice. *Oral Tradition*, 25(2), 365-380.
- Cavitt, M. E. (1997). Effects of expectations on evaluators' judgments of music performance. In R. A. Duke, & J. C. Henninger (Eds.). *Proceedings of the Annual Meeting of the Texas Music Educators Association*. San Antonio, Texas.
- Chang, H. W., & Trehub, S. E. (1977). Auditory processing of relational information by Young infants. *Journal of Experimental Child Psychology*, 24, 324-331.
- Chong, H. J. (2010). Do we all enjoy singing? A content analysis of non-vocalists' attitudes toward singing. *The Arts in Psychotherapy*, 37(2), 120-124.
- Cleall, C. (1979). Vocal range in young children. Paper presented at the course in Development of Young Children's Music Skills, University of Reading. Cited in Goetze, M., Cooper, N., & Brown, C.J. (1990). Singing in the general music classroom. *Bulletin of the Council for Research in Music Education*, 104, 16–37.
- Cleveland, T. F., Sundberg, J., & Stone, R. E. (2001). Long-Term-Average Spectrum characteristics of country singers during speaking and singing. *Journal of Voice*, 15(1), 54-60.

- Clift S., & Hancox G. (2001). The perceived benefits of singing: Findings from preliminary surveys of a university college choral society. *Journal of the Royal Society for the Promotion of Health, 121*, 248-256.
- Clift, S., Hancox, G., Morrison, I., Hess, B., Kreutz, G., & Stewart, D. (2007). Choral singing and psychological wellbeing: Findings from English choirs in a crossnational survey using the WHOQOL-BREF. In A. Williamon & D. Coimbra (Eds). Proceedings of the *International Symposium on Performance Science* (pp. 201-207). Porto, Portugal.
- Clift, S., Hancox, G., Morrison, I., Hess, B., Kreutz, G., & Stewart, D. (2009). What do singers say about the effects of choral singing on physical health ? Findings from a survey of choristers in Australia, England and Germany. In J. Louhivuori, T. Eerola, S. Saarikallio, T. Himberg, & P. -S. Eerola (Eds). Proceedings of the *7th Triennial Conference of European Society for the Cognitive Sciences of Music*. Jyväskylä, Finland.
- Collyer, S., David, P. J., Thorpe, C. W., & Callaghan, J. (2007). Sound pressure level and spectral balance linearity and symmetry in the  *messa di voce*  of female classical singers. *Journal of the Acoustical Society of America, 121*, 1728-1736.
- Collyer, S., David, P. J., Thorpe, C. W., & Callaghan, J. (2009). Fundamental frequency influences the relationship between sound pressure level and spectral balance in female classically trained singers. *Journal of the Acoustical Society of America, 126*, 396-406.
- Collyer, S., Kenny, D. T., & Archer, M. (2011). Listener perception of the effect of abdominal kinematic directives on respiratory behavior in female classical singing, *Journal of Voice, 25*, 15-24.
- Connell, L., Gai, Z. G., & Holler, J. (2013). Do you see what I'm singing ? Visuospatial movement biases pitch perception. *Brain and Cognition, 81*, 124-130.

- Cox, R. H., Martens, M. P., & Russell, W. D. (2003). Measuring anxiety in athletics: the revisited competitive state anxiety inventory-2. *Journal of Sport and Exercise Psychology, 25*, 519-533.
- Craske, M. G., & Craig, K. D. (1984). Musical performance anxiety: The three systems-model and self-efficacy-theory. *Behaviour Research and Therapy, 22*, 267-280.
- Cross, I. (2001). Music, cognition, culture, and evolution. *Annals of the New York Academy of Science, 930*, 28-42.
- Cuddy, L. L., Balkwill, L. L., Peretz, I., & Holden, R. R. (2005). Musical difficulties are rare: a study of "tone deafness" among university students. *Annals of the New York Academy of Sciences, 1060*, 311-324.
- Curtis, M.E., & Bharucha, J.J. (2009). Memory and musical expectations for tones in cultural context. *Music Perception, 26*(4), 365-375.
- Dalla Bella, S., & Berkowska, M. (2009). Singing Proficiency in the Majority. *Annals of the New York Academy of Sciences, 1169*(1), 99-107.
- Dalla Bella, S., Berkowska, M., & Sowinski, J. (2011). Disorders of pitch production in tone deafness. *Frontiers in Psychology, 2*, 164.
- Dalla Bella, S., Giguère, J. -F., & Peretz, I. (2007). Singing proficiency in the general population. *The Journal of the Acoustical Society of America, 121*(2), 1182-1189.
- Dalla Bella, S., Giguère, J. -F., & Peretz, I. (2009). Singing in congenital amusia. *The Journal of the Acoustical Society of America, 126*(1), 414-424.
- Davidson, J. W., & Edgar, R. (2003). Gender and race bias in the judgment of Western art music performance. *Music Education Research, 5*(2), 169-181.



- Davidson, J. W. (2011). Musical participation: expectations, experiences, and outcomes. In I. Deliège and J. W. Davidson (Eds), *Music and the mind* (pp. 65-87). Oxford : Oxford University Press.
- Davidson, L. (1994). Songsinging by young and old: a developmental approach to music. In R. Aiello and J. Sloboda (Eds.), *Musical Perceptions* (pp. 99-130). New York : Oxford University Press.
- de Almeida Bezerra, A., Cukier-Blaj, S., Duprat, A., Camargo, Z., & Granato, L. (2009). The characterization of the vibrato in lyric and sertanejo singing styles: acoustic and perceptual auditory aspects. *Journal of Voice*, *23*, 666-670.
- De Cheveigné, A. & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America*, *111*, 1917-1930.
- Dejonckere, P. H., Giordano, A., Schoentgen, J., Fraj, S., Bocchi, L., & Manfredi, C. (2012). To what degree of voice perturbation are jitter measurements valid? A novel approach with synthesized vowels and visuo-perceptual pattern recognition, *Biomedical Signal Processing and Control*, *7*, 37-42.
- Deliège, I., & Sloboda, J. (2003). *Musical beginnings: Origins and development of musical competence*. New York : Oxford University Press.
- Demorest, S. M., & Clements, A. (2007). Factors influencing the pitch-matching of junior high boys. *Journal of Research in Music Education*, *55*, 190-203.
- Dowling, W. J. (1999). The development of music perception and cognition. In D. Deutsch (Ed.), *The psychology of music* (pp. 603-625). San Diego, CA: Academic Press.

- Dowling, W. J., & Fujitani, D. S. (1970). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America*, *49*, 524-531.
- Duerksen, G. L. (1972). Some effects of expectation on evaluation of recorded musical performance. *Journal of Research in Music Education*, *20*, 268-272.
- Edworthy, J. (1985). Interval and contour in melody processing. *Music Perception*, *2*, 375-388.
- Ekholm, E., Papagiannis, G. C., & Chagnon, F. P. (1998). Relating objective measurements to expert evaluation of voice quality in Western classical singing: Critical perceptual parameters. *Journal of Voice*, *12*, 182-196.
- Elliott, C. (1996). Race and gender as factors in judgments of musical performance. *Bulletin of the Council for Research in Music Education*, *127*, 50-55.
- Elmer, S. S., & Elmer, F. J. (2000). A new method for analysing and representing singing. *Psychology of Music*, *28*(1), 23-42.
- Elwood, L. S., Wolitzky-Taylor, K., & Olatunji, B. O. (2012). Measurement of anxious traits: A contemporary review and synthesis. *Anxiety, Stress, and Coping*, *25*(6), 647-666.
- Ericsson, K. E., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *American Psychological Association*, *100*, 363-406.
- Estill, J. (1988). Belting and classic voice quality: Some physiological differences. *Medical Problems of Performing Artists*, *3*, 37-43.
- Estis, J. M., Coblentz, J. K., & Moore, R. E. (2009). Effects of increasing time delays on pitch-matching accuracy in trained singers and untrained individuals. *Journal of Voice*, *23*(4), 439-445.

- Estis, J. M., Dean-Claytor, A., Moore, R. E., & Rowell, T. L. (2011). Pitch-matching accuracy in trained singers and untrained individuals: the impact of musical interference and noise. *Journal of Voice, 25*(2), 173-180.
- Evans, M., & Howard, D. M. (1993). Larynx closed quotient in female belt and opera qualities: A case study. *Journal of Voice, 2*, 7-14.
- Fant, G. (1960). *Acoustic theory of speech production*. Netherlands, The Hague : Mouton.
- Fehm L, & Schmidt K. (2006). Performance anxiety in gifted adolescent musicians. *Journal of Anxiety Disorders, 20*(1), 98-109.
- Ferguson, S., Kenny, D. T., & Cabrera, D. (2010). Effects of training on time-varying spectral energy and sound pressure level in nine male classical singers. *Journal of Voice, 24*(1), 39-46.
- Ferland, M. B., & Mendelson, M. J. (1989). Infants' categorization of melodic contour. *Infant Behaviour Development, 12*, 341-355.
- Ferrante, I. (2011). Vibrato rate and extent in soprano voice: a survey on one century of singing. *Journal of the Acoustical Society of America, 130*, 1683-1688.
- Fiske, H. E. (1975). Judge-group differences in the rating of secondary school trumpet performances. *Journal of Research in Music Education, 23*, 186-196.
- Fiske, H. E. (1977). Relationship of selected factors in trumpet performance adjudication. *Journal of Research in Music Education, 25*, 256-263.
- Fiske, H. E. (1978). *The effect of a training procedure in musical performance evaluation on judge reliability*. Canada, Toronto : Ontario Educational Research Council Report.

- Flores, R., & Ginsburgh, V. (1996). The Queen Elisabeth musical competition, how fair is the final ranking. *Journal of the Royal Statistical Society*, *45*, 97-104.
- Flowers, P. J., & Dunne-Sousa, D. (1990). Pitch-pattern accuracy, tonality, and vocal range in preschool children's singing. *Journal of Research in Music Education*, *38*, 102-114.
- Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., & Griffiths, T. D. (2004). Characterization of deficits in pitch perception underlying 'tone deafness'. *Brain*, *127*, 801-810.
- Foxton, J. M., Nandy, R. K., & Griffiths, T. D. (2006). Rhythm deficit in "tone deafness". *Brain and Cognition*, *62*, 24-29.
- Francois, C., & Schön, D. (2010). Learning of musical and linguistic structures: comparing event-related potentials and behavior. *NeuroReport*, *21*(14), 928-932.
- François, C., & Schön, D. (2011). Musical expertise boosts implicit learning of both musical and linguistic structures. *Cerebral Cortex*, *21*(10), 2357-2365.
- François, C., Tillmann, B., & Schön, D. (2012). Cognitive and methodological considerations on the effects of musical expertise on speech segmentation. *Annals of the New York Academy of Sciences*, *1252*, 108-115.
- Franklin, M. S., Sledge Moore, K., Yip, C. Y., Jonides, J., Rattray, K., & Moher, J. (2008). The effects of musical training on verbal memory. *Psychology of Music*, *36*(3), 353-365.
- Fredrikson, M., & Gunnarsson, R. (1992). Psychobiology of stage fright: the effect of public performance on neuroendocrine, cardiovascular and subjective reactions. *Biological Psychology*, *33*, 51-61.

- Fujioka, T., Trainor L. J., Ross, B., Kakigi R., Pantev, C. (2004). Musical training enhances automatic encoding of melodic contour and interval structure. *Journal of Cognitive Neuroscience*, 16, 1010-1021.
- Gable, F. K. (1992). Some observations concerning Baroque and Modern vibrato. *Performance Practice Review*, 5(1), 90-102.
- Garnier, M., Henrich, N., Castellengo, M., Sotiropoulos, D., & Dubois, D. (2007). Characterisation of voice quality in Western lyrical singing: from teachers' judgements to acoustic descriptions. *Journal of interdisciplinary music studies*, 1, 62-91.
- Garnier, M., Henrich, N., Smith, J., & Wolfe, J. (2010) Vocal tract adjustments in the high soprano range. *Journal of the Acoustical Society of America*, 127(6), 3771-3780.
- George, E. M., & Coch, D. (2011). Music training and working memory: an ERP study. *Neuropsychologia*, 49(5), 1083-1094.
- Geringer, J. M. (1983). The Relationship of pitch-matching and pitch-discrimination abilities of preschool and fourth-grade students. *Journal of Research in Music Education*, 31(2), 93-100.
- Giddens, C. L., Barron, K. W., Byrd-Craven, J., Clark, K. F., & Winter, A. S. (2013). Vocal indices of stress: A review. *Journal of Voice*, 27, e21-e29.
- Glejser, H., & Heyndel, B. (2001). Efficiency and inefficiency in the ranking in competitions: The case of the Queen Elisabeth music contest. *Journal of Cultural Economics*, 25, 109-129.
- Godlovitch, S. (1998). *Musical Performance: A Philosophical Study*. London: Routledge.

- Goetze, M., Cooper, N., & Brown, C. J. (1990). Recent research on singing in the general music classroom. *Bulletin of the Council for Research in Music Education, 104*, 16-37.
- Goetze, M., & Horii, Y (1989). A comparison of the pitch accuracy of group and individual singing in young children. *Bulletin of the Council for Research in Music Education, 99*, 57-73.
- Gomez, E., Klapuri, A., & Meudic, B. (2003). Melody description and extraction in the context of music content processing. *Journal of New Music Research, 32*(1), 23-40.
- Gooding, L., & Standley, J. M. (2011). Musical development and learning characteristics of students: A compilation of key points from the research literature organized by age. *National Association for Music Education, 30*(1), 32-45.
- Granot, R. Y., Israel-Kolatt, R., Gilboa, A., & Kolatt, T. (in press). Accuracy of pitch matching significantly improved by live voice model. *Journal of Voice*.
- Green, G. A. (1989). The effect of vocal modeling on pitch-matching accuracy of elementary schoolchildren. *Journal of Research in Music Education, 38*, 105-114.
- Greenberg, M. (1979). *Your children need music*. Englewood, NJ: Prentice-Hall.
- Griffiths, N. K. (2008). The effects of concert dress and physical appearance on perceptions of female solo performers. *Musicae Scientiae, 12*(2), 273-290.
- Griffiths, N. K. (2010). 'Posh music should equal posh dress': an investigation into the concert dress and physical appearance of female soloists. *Psychology of Music, 38*(2), 159-177.

- Hagenaars, M. A., & van Minnen, A. (2005). The effect of fear on paralinguistic aspects of speech in patients with panic disorder with agoraphobia. *Journal of Anxiety Disorders, 19*, 521–537.
- Halpern, A. R. (1988). Perceived and imagined tempos of familiar songs. *Music Perception, 6*, 193-202.
- Halpern, A. R., & Zatorre, R. J. (1979). Identification, discrimination, and selective adaptation of simultaneous musical intervals. *Perception and Psychophysics, 26*, 384–395.
- Hamann, D. L., & Sobaje, M. (1983). Anxiety and the college musician: a study of performance conditions and subject variables. *Psychology of Music, 11*, 37–50.
- Hanning, B. R. (2011). *Concise history of Western music* (fourth edition). New York: Norton & Company.
- Hannon, E. E., & Trainor, L. J. (2007). Music acquisition : effects of enculturation and formal training on development. *Trends in Cognitive Sciences, 11*(11), 466-472.
- Hargreaves, D. J. (1996). The development of artistic and musical competence. In I. Deliège & J. Sloboda (Eds.), *Musical Beginnings* (pp. 145-170). Oxford: Oxford University Press.
- Hash, P. M. (in press). An analysis of middle/high school band and orchestra festival ratings. *Journal of Band Research*.
- Hébert, S., Racette, A., Gagnon, L., & Peretz, I. (2003). Revisiting the dissociation between singing and speaking in expressive aphasia. *Brain, 126*(8), 1838-1850.
- Hedden, D. (2012). An overview of existing research about children's singing and the implications for teaching children to sing. *Update: Application of Research in Music Education, 30*(2), 52-62.

- Henrich, N. (2012). Physiologie de la voix chantée: vibration laryngée et adaptations phono-résonantielles. 40èmes Entretiens de Médecine physique et de réadaptation, Montpellier, France.
- Henrich, N., Kiek, M., Smith, J., & Wolfe, J. (2007). Resonance strategies used in Bulgarian women's singing style: a pilot study. *Logopedics Phoniatrics Vocology*, 32(4), 171-177.
- Henrich, N., Smith, J., & Wolfe, J. (2011). Vocal tract resonances in singing: Strategies used by sopranos, altos, tenors, and baritones. *Journal of the Acoustical Society of America*, 129(2), 1024-1035.
- Henry, M. J., & McAuley, J. D. (2010). On the prevalence of congenital amusia. *Music Perception*, 27(5), 413-418.
- Hess, W. (1983). *Pitch determination of speech signals : algorithms and methods*. New York : Springer-Verlag.
- Hirano, M., Hibi, S., & Hagino, S. (1995). Physiological aspects of vibrato. in P. Dejonckere, M. Hirano, & J. Sundberg (Eds.), *Vibrato* (pp. 9-34). San Diego, CA: Singular Publishing Group.
- Howard, D. M. (2007a). Equal or non-equal temperament in a capella SATB singing. *Logopedics Phoniatrics Vocology*, 32(2), 87-94.
- Howard, D. M. (2007b). Intonation drift in a capella soprano, alto, tenor, bass quartet singing with key modulation. *Journal of Voice*, 21(3), 300-315.
- Howard, S. A. (2012). The effect of selected nonmusical factors on adjudicators' ratings of high school solo vocal performances. *Journal of Research in Music Education*, 60(2), 166-185.
- Howard, D. M., & Angus, J. (2009). *Acoustics and Psychoacoustics*. Oxford : Focal Press.



- Howard, D. M., & Welch, G. F. (1989). Microcomputer-based singing ability assessment and development. *Applied Acoustics*, *27*, 89-102.
- Howes, P., Callaghan, J., Davis, P., Kenny, D., & Thorpe, W. (2004). The relationship between measured vibrato characteristics and perception in Western operatic singing. *Journal of Voice*, *18*, 216-230.
- Hunter, E. J., Svec, J. G., & Titze, I. R. (2006). Comparison of the produced and perceived voice range profiles in untrained and trained classical singers. *Journal of Voice*, *20*(4), 513-526.
- Hutchins, S., Zarate, J. M., Zatorre, R. J., & Peretz, I. (2010). An acoustical study of vocal pitch matching in congenital amusia. *The Journal of the Acoustical Society of America*, *127*(1), 504-512.
- Hutchins, S., & Peretz, I. (2012). A frog in your throat or in your ear? Searching for the causes of poor singing. *Journal of Experimental Psychology: General*, *141*, 76-97.
- Hutchins, S., Roquet, C., & Peretz, I. (2012). The vocal generosity effect: How bad can your singing be? *Music Perception*, *30*(2), 147-159.
- Hutchins, S., & Peretz, I. (2011). Perception and action in singing. *Progress in Brain Research*, *191*, 103-118.
- Hyde, K., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science*, *15*(5), 356-360.
- Hyde, K., Zatorre, R. J., Griffiths, T. D., Lerch, J. P., & Peretz, I. (2006). Morphometry of the amusic brain: A two-side study. *Brain*, *129*, 2562-2570.
- Hyde, K., Zatorre, R. J., & Peretz, I. (2011). Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. *Cerebral Cortex*, *21*, 292-299.

- Iusca, D., & Dafinoiu, I. (2012). Performance anxiety and musical level of undergraduate students in exam situations: The role of gender and musical instrument. *Procedia - Social Behavioral Sciences*, 33, 448-452.
- Jakobson, L., Lewycky, S., Kilgour, A., & Stoesz, B. (2008). Memory for verbal and visual material in highly trained musicians. *Music Perception*, 26 (1), 41-55.
- Jancke, L. (2008). Music, memory and emotion. *Journal of Biology*, 7, 21.
- Jancke, L. (2009). The plastic human brain. *Restorative Neurology and Neuroscience*, 27, 521-538.
- Jonaitis, E. M. M., & Saffran, J. R. (2009). Learning harmony: The role of serial statistics. *Cognitive Science*, 33, 951-968.
- Joyner, D. R. (1969). The monotone problem. *Journal of Research in Music Education*, 31, 115-124.
- Juchniewicz, J. (2008). The influence of physical movement on the perception of musical performance. *Psychology of Music*, 36(4), 417-427.
- Kacha, A., Grenez, F., & Schoentgen, J. (2005). Voice quality assessment by means of comparative judgments of speech tokens. Proceedings of 9th European Conference on Speech Communication and Technology, *Interspeech* (pp. 1733-1736). Lisbon, Portugal.
- Kenny, D. (2011). *The psychology of music performance anxiety*. Oxford: Oxford University Press.
- Kim, S. G., Kim, J. S., & Chung, C. K. (2011). The effect of conditional probability of chord progression on brain response: An MEG study. *PLoS ONE*, 6(2), e17337.

- King, J. B., & Horii, Y. (1993). Vocal matching of frequency modulation in synthesized vowels. *Journal of Voice*, 7, 151-157.
- Kinney, D. W. (2009). Internal consistency of performance evaluations as a function of music expertise and excerpt familiarity. *Journal of Research in Music Education*, 56(4), 322-337.
- Kinsella, G., Prior, M. R., & Murray, G. (1988). Singing ability after right and left sided brain damage. A research note. *Cortex*, 24, 165-169.
- Kisilevsky, B. S., & Low, J. A. (1998). Human fetal behavior : 100 years of study. *Developmental Review*, 18, 1-29.
- Koelsch, S. (2011). Toward a neural basis of music perception – A review and updated model. *Frontiers in Psychology*, 2, 1-20.
- Koelsch, S., Jentschke, S., Sammler, D., & Mietchen, D. (2007). Untangling syntactic and sensory processing: An ERP study of music perception. *Psychophysiology*, 44(3), 476-490.
- Kokotsaki, D., & Davidson, J. W. (2003). Investigating musical performance anxiety among music college singing students: A quantitative analysis. *Music Education Research*, 5(1), 45-59.
- Krampe, R. T., & Ericsson, K. A. (1996). Maintaining excellence: Deliberate practice and elite performance in younger and older pianists. *Journal of Experimental Psychology: General*, 125(4), 331-359.
- Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Neuroscience*, 11, 599-605.
- Kreutz, G., Bongard, S., Rohrman, S., Hodapp, V., & Grebe, D. (2004). Effects of choir singing or listening on secretory immunoglobulin A, cortisol and emotional state. *Journal of Behavioral Medicine*, 27, 623-635.

- Krumhansl, C. (1990). *Cognitive foundations of musical pitch*. Oxford: Oxford University Press.
- Krumhansl, C. L., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 739-751.
- Kurosawa, K., & Davidson, J. W. (2005). Nonverbal behaviours in popular music performance : A case study of The Corrs. *Musicae Scientiae*, *19*(1), 111-136.
- Kveraga, K., Boshyan, J., & Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *Journal of Neuroscience*, *27*, 13232-13240.
- Lam Tang, J. A., Boliek, C. A., and Rieger, J. M. (2008). Laryngeal and respiratory behavior during pitch change in professional singers. *Journal of Voice*, *22*, 622-633.
- Lamarche, A., Ternström, S., & Pabon, P. (2010). The singer's voice range profile: Female professional opera soloists. *Journal of Voice*, *24*, 410-426.
- Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological bulletin*, *87*(1), 72-107.
- Larrouy-Maestri, P., Lévêque, Y., Schön, D., Giovanni, A., & Morsomme, D. (2013). The evaluation of singing voice accuracy: A comparison between subjective and objective methods. *Journal of voice*, *27*(2), 259.e1-e5.
- Larrouy-Maestri, P., Leybaert, J., & Kolinsky, R. (2013). The benefit of musical and linguistic expertise on language acquisition in sung material. *Musicae Scientiae*, *17*(2), 217-228.

- Larrouy-Maestri, P., Magis, D., & Morsomme, D. (in press). The effect of melody and technique on the singing voice accuracy of trained singers. *Logopedics Phoniatics Vocology*.
- Larrouy-Maestri, P., & Morsomme, D. (in press). Criteria and tools for objectively analysing the vocal accuracy of a popular song. *Logopedics Phoniatics, Vocology*.
- Leanderson, R., Sundberg, J., & van Euler, C. (1987). Breathing muscle activity and subglottal pressure dynamics in singing and speech. *Journal of Voice, 1*, 258-261.
- LeBlanc A, Chang Jin Y, Obert M, & Siivola C. (1997). Effect of audience on music performance anxiety. *Journal of Research in Music Education, 45*(3), 480-490.
- LeBorgne, W. D., Lee, L., Stemple, J. C., & Bush, H. (2010). Perceptual findings on the Broadway belt voice. *Journal of Voice, 24*, 678-689.
- LeBorgne, W. D., & Weinrich, B. D. (2002). Phonetogram changes for trained singers over a nine-month period of vocal training. *Journal of voice, 16*(1), 37-43.
- Lebowitz, A., & Baken, R. J. (2011). Correlates of the belt voice: A broader examination. *Journal of Voice, 25*, 159-165.
- Leighton, G., & Lamont, A. (2006). Exploring children's singing development: Do experiences in early schooling help or hinder? *Music Education Research, 8*(3), 311-330.
- Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA : MIT Press.
- Lévêque, Y., Amy de la Bretèque, B., Giovanni, A., & Schön, D. (2009). Les défauts de justesse de la voix chantée: compétences et déficits tonaux. *Revue de Laryngologie Otologie Rhinologie, 130*(1), 23-38.

- Lévêque, Y., Giovanni, A., & Schön, D. (2012). Pitch-matching in poor singers : Human model advantage. *Journal of voice*, 26(3), 293-298.
- Levin, T., & Edgerton, M. (1999). The throat singers of Tuva. *Scientific American*, 281(3), 70-77.
- Levinowitz, L. M. (1989). An investigation of preschool children's comparative capability to sing songs with and without words. *Bulletin of the Council for Research in Music Education*, 100, 14-19.
- Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception & Psychophysics*, 56, 414-423.
- Levitin, D. J., & Cook, P. R. (1996). Memory for musical tempo: Additional evidence that auditory memory is absolute. *Perception & Psychophysics*, 58, 927-935.
- Lieberman, P. (1961). Perturbations in vocal pitch. *Journal of the Acoustical Society of America*, 33, 597-603.
- Lieberman, P. (1963). Some acoustic measures of the fundamental periodicity of normal and pathological larynges. *Journal of the Acoustical Society of America*, 35, 344-353.
- Liégeois-Chauvel, C., Peretz, I., Babai, M., Laguitton, V., & Chauvel, P. (1998). Contribution of different cortical areas in the temporal lobes to music processing. *Brain*, 121, 1853-1867.
- Livingstone, S. R., Thompson, W. F., & Russo, F. A. (2009). Facial expressions and emotional dinging : A study of perception and production with motion capture and electromyography. *Music Perception*, 26(5), 475-488.
- London, J. (2012). *Hearing the time : Psychological aspects of musical meter*. Oxford : Oxford University Press.

- Louhivuori, J., Salminen, V. M., & Lebaka, E. (2005). 'Singing together': A cross cultural approach to the meaning of choirs as a community. In: Campbell, Patricia Shehan (Editor). *Cultural Diversity in Music Education* (pp. 81-92). Bowen Hills, Qld.: Australian Academic Press.
- Loui, P., Guenther, F., Mathys, C., & Schlaug, G. (2008). Action-perception mismatch in tone-deafness. *Current Biology*, 18(8), 331-332.
- Loui, P., & Wessel, D. (2008). Learning and linking an artificial musical system: Effect of set size and repeated exposure. *Musicae Scientiae*, 12(2), 207-230.
- Loui, P., Wessel, D. L., & Kam, C. L. H. (2010). Humans rapidly learn grammatical structure in a new musical scale. *Music Perception*, 27(5), 377-388.
- Lundy, D. S., Roy, S., Casiano, R. R., Xue, J. W., & Evans, J. (2000). Acoustic analysis of the singing and speaking voice in singing students. *Journal of Voice*, 14, 400-493.
- MacDonald, R. A. R., Kreutz, G., & Mitchell, L. (2012). *Music, health, and wellbeing*. Oxford: Oxford University Press.
- Mampe, B., Friederici, A.D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology*, 19(23), 1994-1997.
- Mantell, J. T., & Pfordresher, P. Q. (2013). Vocal imitation of song and speech. *Cognition*, 127(2), 177-202.
- Marmel, F., Tillmann, B., & Dowling, W. J. (2008). Tonal expectations influence pitch perception. *Perception & Psychophysics*, 70(5), 841-852.
- Martinet, G., Ferrand, C., Guillet, E., & Gauthier, S. (2010). Validation of the French version of the Competitive State Anxiety Inventory – 2

- Revised (CSAI-2R) including frequency and direction scales. *Psychology of Sport and Exercise*, 11(1), 51-57.
- McAuley, J. D., Henry, M. J., & Tuft, S. (2011). Musician advantages in music perception: An issue of motivation, not just ability. *Music Perception*, 28(5), 505-518.
- McDermott, J. H., Keebler, M. V., Micheyl, C., & Oxenham, A. J. (2010). Musical intervals and relative pitch: frequency resolution, not interval resolution, is special. *Journal of the Acoustical Society of America*, 128(4), 1943-1951.
- McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2008). Music perception, pitch, and the auditory system. *Current opinion in neurobiology*, 18(4), 452-463.
- McPherson, G. E. & Schubert, E. (2004). Measuring performance enhancement in music. In A. Williamon (Ed), *Musical excellence: Strategies and techniques to enhance performance* (pp. 61-82). Oxford : Oxford University Press.
- McPherson, G. E., & Thompson, W. F. (1998). Assessing music performance: Issues and influences. *Research Studies in Music Education*, 10, 12-24.
- Melara, R. D., & Marks, L. E. (1990a). HARD and SOFT interacting dimensions: Differential effects of dual context on classification. *Perception & Psychophysics*, 47, 307-325.
- Melara, R. D., & Marks, L. E. (1990b). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception & Psychophysics*, 48, 169-178.
- Melara, R. D., & Marks, L. E. (1990c). Perceptual primacy of dimensions: Support for a model of dimensional interaction. *Journal of*



Experimental Psychology. *Human Perception and Performance*, 16, 398-414.

Mendes, A. P., Rothman, H. B., Sapienza, C., & Brown, W. S. (2003). Effects of vocal training on the acoustic parameters of the singing voice. *Journal of Voice*, 17(4), 529-543.

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, 219(1-2), 36-47.

Mitchell, H. F., & Kenny, D. T. (2008). The tertiary singing audition: Perceptual and acoustical differences between successful and unsuccessful candidates. *Journal of Interdisciplinary Music Studies*, 2(1), 1-16.

Miyamoto, K. A. (2007). Musical characteristics of preschool-age students : A review of literature. *National Association for Music Education*, 26, 26-40.

Moog, H. (1976). *The musical experience of the pre-school child*. London: Schott.

Moorcroft, L., & Kenny, D. T. (2012). Vocal warm-up produces acoustic change in singers' vibrato rate. *Journal of Voice*, 26, 13-18.

Moore, B. C. J. (1973). Frequency difference limens for short-duration tones. *Journal of the Acoustical Society of America*, 54, 610-619.

Moore, R. E., Estis, J., Gordon-Hickey, S., & Watts, C. (2008). Pitch discrimination and pitch matching abilities with vocal and nonvocal stimuli. *Journal of Voice*, 22(4), 399-407.

Moore, R. E., Keaton, C., & Watts, C. (2007). The role of pitch memory in pitch discrimination and pitch matching. *Journal of Voice*, 21, 560-567.

- Moran, H., & Pratt, C. C. (1926). Variability of judgments of musical intervals. *Journal of Experimental Psychology*, 9, 492–500.
- Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., & Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychological Science*, 22(11), 1425-1433.
- Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2008). Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cerebral Cortex*, 19(3), 712-723.
- Morrison, S. J., Montemayer, M., & Wiltshire, E. S. (2004). The effect of a recorded model on band students' performance self-evaluations, achievement and attitude. *Journal of Research in Music Education*, 52, 116–129.
- Murayama, J., Kashiwagi, T., Kashiwagi, A., & Mimura, M. (2004). Impaired pitch production and preserved rhythm production in a right brain-damaged patient with amusia. *Brain and Cognition*, 56, 36-42.
- Murry, T. (1990). Pitch-matching accuracy in singers and nonsingers. *Journal of Voice*, 4(4), 317-321.
- Mürbe, D., Sundberg, J., Iwarsson, J., Parst, F., & Hofmann, G. (1999). Longitudinal study of solo singer education effects on maximum SPL and level in the singers' formant range. *Logopedics Phoniatrics Vocology*, 24, 178-186.
- Mürbe, D., Zahnert, T., Kuhlisch, E., & Sundberg, J. (2007). Effects of professional singing education on vocal vibrato: A longitudinal study. *Journal of Voice*, 21, 683-688.

- Nakata, T., Trehub, S. E., Mitani, C., & Kanda, Y. (2006). Pitch and timing in the sings of deaf children with cochlear implants. *Music Perception, 24*(2), 147-154.
- Nettl, B. (2000). An ethnomusicologist contemplates universals in musical sound and musical culture. In: Wallin, N.L., Merker, B., and Brown, S. (Eds.), *The origins of music* (pp. 463-472). Cambridge, MA: MIT Press.
- Nikjeh, D. A., Lister, J. J., & Frisch, S. A (2009). The relationship between pitch discrimination and vocal production : Comparison of vocal and instrumental musicians. *Journal of the Acoustical Society of America, 125*(1), 328-338.
- Norris, C. E., & Borst, J. D. (2007). An examination of the reliabilities of two choral festival adjudication forms. *Journal of Research in Music Education, 55*(3), 237-251.
- Odena, O. (2007). Music as a way to address social inclusion and respect for diversity in early childhood. Study Paper for the Bernard van Leer Foundation. Belfast: NFER at Queen's.
- Omori, K., Kacker, A., Carroll, L. M., Riley, W. D., & Blaugrund, S. M. (1996). Singing power ratio: Quantitative evaluation of singing voice quality. *Journal of Voice, 10*, 228-235.
- Orsmond, G. I., & Miller, L. K. (1999). Cognitive, musical and environmental correlates of early music instruction. *Psychology of Music, 27*, 18-37.
- Ostwald, P. F. (1973). Musical behavior in early childhood. *Developmental Medicine and Child Neurology, 15*, 367-375.
- Papageorgi, I., Creech, A., & Welch, G. (2013). Perceived performance anxiety in advances musicians specializing in different musical genres. *Psychology of Music, 41*(1), 18-41.

- Papageorgi, I., Hallam, S., & Welch, G. F. (2007). A conceptual framework for understanding musical performance anxiety. *Research Studies in Music Education*, 28(1), 83-107.
- Papoušek, H. (1996). Musicality in infancy research: Biological and cultural origins of early musicality. In I. Deliège & J. Sloboda (Eds.), *Musical beginnings* (pp. 37-55). Oxford: Oxford University Press.
- Patel, A. D. (2008). *Music, language, and the brain*. Oxford : Oxford University Press.
- Patel, A. D. (2010). Music, biological evolution, and the brain. In M. Bailar (Ed.), *Emerging Disciplines* (pp. 91-144). Houston, TX : Rice University Press.
- Peretz, I. (1993). Auditory atonalia for melodies. *Cognitive Neuropsychology*, 10, 21-56.
- Peretz, I., (2008). Musical disorders: From behavior to genes. *Current Directions in Psychological Science*, 17(5), 329-333.
- Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhure, V. B., & Jutras, B. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron*, 33, 185-191.
- Peretz, I., Gosselin, N., Tillmann, B., Cuddy, L. L., Gagnon, B., Trimmer, C. G., Paquette, S., & Bouchard, B. (2008). On-line identification of congenital amusia. *Music perception*, 25, 331-343.
- Perruchet, P. (2008). Implicit learning. In J. Byrne (Ed.). *Cognitive psychology of memory* (volume 2, pp. 597-621). Oxford : Elsevier.
- Petzold, R. G. (1969). Auditory perception by children. *Journal of Research in Music Education*, 812-787.

- Pfordresher, P. Q. (2012). Musical training and the role of auditory feedback during performance. *Annals of the New York Academy of Sciences, 1252*, 171-178.
- Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of "tone deafness". *Music Perception, 25*(2), 95-115.
- Pfordresher, P. Q., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention, Perception & Psychophysics, 71*(6), 1385-1398.
- Pfordresher, P. Q., Brown, S., Meier, K. M., Belyk, M., & Liotti, M. (2010). Imprecise singing is widespread. *The Journal of the Acoustical Society of America, 128*(4), 2182-2190.
- Pfordresher, P. Q., & Mantell, J. T. (2009). Singing as a Form of Vocal Imitation: Mechanisms and Deficits. In J. Louhivuori, T. Eerola, S. Saarikallio, T. Himberg, & P. -S. Eerola (Eds.) *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music* (pp. 425-430). Jyväskylä, Finland.
- Pierce, J. (1984). *Le son musical : musique, acoustique et informatique* (p. 126). Paris : Belin.
- Pitt, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 976-986.
- Plantinga, J., & Trainor, L. (2005). Memory for melody: infants use a relative pitch code. *Cognition, 98*(1), 1-11.
- Portowitz, A., Lichtenstien, O., Egorov, L., & Brand, E. (2009). Underlying mechanisms linking music education and cognitive modifiability. *Research Studies in Music Education, 31*(2), 107-128.
- Prame, E. (1994). Measurements of the vibrato rate of ten singers, *Journal of the Acoustical Society of America, 94*, 1979-1984.

- Prame, E. (1997). Vibrato extent and intonation in professional Western lyric singing. *Journal of the Acoustical Society of America*, *102*, 616-621.
- Price, H. E. (2000). Interval matching by undergraduate nonmusic majors. *Journal of Research in Music Education*, *48*(4), 360-372.
- Prince, J. B. (2011). The integration of stimulus dimensions in the perception of music. *Quarterly Journal of Experimental Psychology*, *64*(11), 2125-2152.
- Prior, M., Kinsella, G., & Giese, J. (1990). Assessment of musical processing in brain-damaged patients: implications for laterality of music. *Journal of Clinical and Experimental Neuropsychology*, *12*, 301-312.
- Racette, A., Bard, C., & Peretz, I. (2006). Making non-fluent aphasics speak: sing along! *Brain*, *129*, 2571-2584.
- Rakowski, A. (1976). Tuning of isolated musical intervals. *Journal of the Acoustical Society of America*, *59*, S50.
- Rakowski, A. (1990). Intonation variants of musical intervals in isolation and in musical contexts. *Psychology of Music*, *18*, 60-72.
- Rakowski, A., & Miskiewicz, A. (1985). Deviations from equal temperament in tuning isolated musical intervals. *Archives of Acoustics*, *10*, 95-104.
- Reid, K., Davis, P., Oates, J., Cabrera, D., Ternström, S., Black, M., and Chapman, J. (2007). The acoustic characteristics of professional opera singers performing in chorus versus solo mode. *Journal of Voice*, *21*(1), 35-45.
- Ringer, A. L. (2002). Melody : Definition and origins. In *The New Grove Dictionary of Music Online*, L. Macy Editor, Macmillan Online Publishing, <http://www.grovemusic.com>.

- Rothman, H. B., Rullman, J. F., & Arroyo, A. A. (1990). Inter-and intrasubject changes in vibrato: Perceptual and acoustic aspects. *Journal of Voice, 4*, 309-316.
- Rossi, V., & Pourtois, G. (2012). Transient state-dependent fluctuations in anxiety measured using STAI, POMS, PANAS or VAS: A comparative review. *Anxiety, Stress and Coping, 25*(6), 603-645.
- Roubeau, B., Castellengo, M., Bodin, P., & Ragot, M. (2004). Phonétogramme par registre laryngé. *Folia Phoniatica et Logopedica, 56*, 321-333.
- Ruiz, R., Absil, E., Harmegnies, B., Legros, C., & Poch, D. (1996). Time- and spectrum-related variabilities in stressed speech under laboratory and real conditions. *Speech Communication, 20*, 111-129.
- Russo, F. A., & Thompson, W. F. (2005). An interval size illusion: The influence of timbre on the perceived size of melodic intervals. *Perception & Psychophysics, 67*(4), 559-568.
- Russo, F. A., & Thompson, W. F. (2005). The subjective size of melodic intervals over a two-octave range. *Psychonomic Bulletin & Review, 12*(6), 1068-1075.
- Rutkowski, J. (1990). The measurement and evaluation of children's singing voice development. *The Quarterly, 1*(1-2), 81-95.
- Rutkowski, J. (1996). The effectiveness of individual/small-group singing activities on kindergartners' use of singing voice and developmental music aptitude. *Journal of Research in Music Education, 44*, 353-368.
- Rutkowski, J. (1997). The nature of children's singing voices: Characteristics and assessment. In B. A. Roberts (Ed.), *The*

- Phenomenon of Singing* (pp. 201-209). St. John's, NF: Memorial University Press.
- Rutkowski, J., & Snell Miller, M. (2003). A longitudinal study of elementary children's acquisition of their singing voices. *Update: Applications of Research in Music Education*, 22(1), 5-14.
- Rutkowski, J., & Snell Miller, M. (2003). The effectiveness of frequency of instruction and individual/small group singing activities on first graders' use of singing voice and developmental music aptitude. *Contributions to Music Education*, 30(1), 23-28.
- Ryan, C., & Costa-Giomi, E. (2004). Attractiveness bias in the evaluation of young pianists' performances. *Journal of Research in Music Education*, 52(2), 141-154.
- Saffran, J., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926-1928.
- Saffran, J., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70, 27-52.
- Saffran, J., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35(4), 606-621.
- Salmon, P. (1990). A psychological perspective on musical performance anxiety: A review of the literature. *Medical Problems of Performing Artists*, 5, 2-11.
- Samson, J. (2001). Romanticism. In S. Sadie & J. Tyrrell (Eds.), *The New Grove Dictionary of Music and Musicians* (second edition). London : Macmillan Publishers.
- Sandgren, M. (2009). Evidence from strong immediate well-being effects of choral singing – with more enjoyment for women than for men.



In J. Louhivuori, T. Eerola, S. Saarikallio, T. Himberg, & P. -S. Eerola (Eds). *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music*, Jyväskylä, Finland.

Sataloff, R. T. (2005). *Professional voice: the science and art of clinical care*. San Diego, CA: Plural Publishing.

Schellenberg, E. G. (2001). Asymmetries in the discrimination of musical intervals: Going out-of-tune is more noticeable than going in-tune. *Music Perception*, 19, 223–248.

Schellenberg, E. G. (2009). Music Training and Nonmusical Abilities: Commentary on Stoesz, Jakobson, Kilgour, and Lewycky (2007) and Jakobson, Lewycky, Kilgour, and Stoesz (2008). *Music Perception*, 27(2), 139-143.

Schellenberg, E. G., & Moreno, S. (2010). Music lessons, pitch processing, and g. *Psychology of Music*, 38(2), 209-221.

Scherer, K. R. (1977). Effect of stress on fundamental frequency of the voice. *Journal of the Acoustical Society of America*, 62, S25.

Schlaug, G. (2011). Music, musicians, and brain plasticity, in S. Hallam, I. Cross, and M. H. Thaut (Eds.). *The Oxford Handbook of Music Psychology* (pp. 197–207). Oxford: Oxford University Press.

Schlaug, G., Norton, A., Overy, K., & Winner, E., (2005). Effects of music training on the child's brain and cognitive development. *Annals of the New York Academy of Sciences*, 1060, 219–230.

Schmuckler, M. A. (2009). Components of melodic processing. In: Hallam (Ed.). *The Oxford Handbook of Music Psychology* (pp. 93-106). Oxford: Oxford University Press.

Schneider, P., & Wengenroth, M. (2009). The neural basis of individual holistic and spectral sound perception. *Contemporary music Review*, 28(3), 315-328.

- Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition*, *106*(2), 975-983.
- Schön, D., Lorber, B., Spacal, M., & Semenza, C. (2004). A selective deficit in the production of exact musical intervals following right-hemisphere damage. *Cognitive Neuropsychology*, *21*(7), 773-784.
- Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, *41*(3), 341-349.
- Schuppert, M., Munte, T. M., Wieringa, B. M., & Altenmüller, E. (2000). Receptive amusia: evidence for cross-hemispheric neural networks underlying music processing strategies. *Brain*, *123*, 546-559.
- Schutte, H. K., & Miller, D. G. (1993). Belting and pop, nonclassical approaches to the female middle voice: Some preliminary conclusions. *Journal of Voice*, *7*, 142-150.
- Seashore, C. E. (1938). *Psychology of Music* (pp. 33-52). New York: McGraw Hill.
- Seddon, F. A. (2011). Musical encounters of the temporary kind. In I. Deliège and J. W. Davidson (Eds), *Music and the mind* (pp. 189-203). Oxford : Oxford University Press.
- Seidner, W., Schutte, H., Wendler, J., & Rauhut, A. (1983). Dependence of the High singing formant on pitch and vowel in different voice types. In A. Askenfelt, S. Felicetti, E. Jansson, & J. Sundberg (Eds). *Proceedings of Stockholm Music Acoustics Conference*, vol. 46 (pp. 261-268). Royal Swedish Academy of Music, Stockholm.

- Seppänen, M., Brattico, E., & Tervaniemi, M. (2007). Practice strategies of musicians modulate neural processing and the learning of sound-patterns. *Neurobiology of Learning and Memory*, 87(2), 236-247.
- Sheldon, D. A. (1994). The effects of competitive versus noncompetitive performance goals on music students' ratings of band performance. *Bulletin of the Council for Research in Music Education*, 121, 29-41.
- Shepard, R. N. (1964). Circularity in judgments of relative pitch. *Journal of the Acoustical Society of America*, 36(12), 2346-2353.
- Shipp, T., Leanderson, R., & Sundberg, J. (1980). Some acoustic characteristics of vocal vibrato. *Journal of Research in Singing*, 4, 18-25.
- Shonle, J. I., & Horan, K. E. (1980). The pitch of vibrato tones. *Journal of the Acoustical Society of America*, 67(1), 246-252.
- Siegel, J. A., & Siegel, W. (1977). Categorical perception of tonal intervals: Musicians can't tell sharp from flat. *Perception & Psychophysics*, 21, 399-407.
- Sims, W. L., Moore, R. S., & Kuhn, T. L. (1982). Effects of female and male vocal stimuli, tonal pattern length, and age on vocal pitch-matching abilities of young children from England and the United States. *Psychology of Music*, 104-108.
- Sloboda, J. (2005). *Exploring the musical mind: cognition, emotion, ability, function* (p. 269). Oxford : Oxford University Press.
- Sloboda, J. A., Wise, K. J. & Peretz, I. (2005). Quantifying tone deafness in the general population. *Annals of the New York Academy of Sciences*, 1060, 255-261.

- Small, A. R., & McCachern, F. L. (1983). The effect of male and female vocal modeling on pitch-matching accuracy of first-grade children. *Journal of Research in Music Education, 31*, 227-233.
- Sofranko, J. L., & Prosek, R. A. (2012). The effect of experience on classification of voice quality. *Journal of voice, 26*(3), 299-303.
- Stalinski, S. M., & Schellenberg, E. G. (2010). Shifting perceptions : Developmental changes in judgments of melodic similarity. *Developmental Psychology, 46*, 1799-1803.
- Stalinski, S. M., & Schellenberg, E. G. (2012). Music cognition: A developmental perspective. *Topics in Cognitive Science, 4*(4), 485-497.
- Stalinski, S. M., Schellenberg, E. G., & Trehub, S. E. (2008). Developmental changes in the perception of pitch contour. Distinguishing up from down. *Journal of the Acoustical Society of America, 124*, 1759-1763.
- Stone, R. E., Cleveland, T. F., Sundberg, J., & Prokop, J. (2002). Aerodynamic and acoustical measures of speech, operatic, and Broadway vocal styles in a professional female singer. *Quarterly Progress and Status Report, 43*, 17-29.
- Strait, D. L., Kraus, N., Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research, 261*(1-2), 22-29.
- Sulter, A. M., Schutte, H. K., & Miller, D. G. (1995). Differences in phonetogram features between male and female subjects with and without vocal training. *Journal of Voice, 9*(4), 363-377.
- Sundberg, J. (1978). Effects of the vibrato and the «singing formant» on pitch. *Journal of Research in Singing, 5*, 5-17.

- Sundberg, J. (1979). Perception of singing. *Quarterly Progress and Status Report, 20*, 1–48.
- Sundberg, J. (1982). In tune or not? A study of fundamental frequency in music practice. *Quarterly Progress and Status Report, 23*, 49–78.
- Sundberg J. (1987). *The Science of the Singing Voice*. DeKalb: North Illinois University Press.
- Sundberg, J. (1994). Acoustic and psychoacoustic aspects of vocal vibrato. *Quarterly Progress and Status Report, 35*, 45-68.
- Sundberg, J. (1994). Perceptual aspects of singing. *Journal of Voice, 8*, 106–122.
- Sundberg, J. (1995). The singer's formant revisited. *Quarterly Progress and Status Report, 36*, 83-96.
- Sundberg, J. (1999). The perception of singing. In D. Deutsch D (Ed.). *Psychology of Music* (pp.171-214). San Diego, CA: Academic Press.
- Sundberg, J. (2001). Level and center frequency of the singer's formant. *Journal of Voice, 15*, 176-186.
- Sundberg, J. (2009). Articulatory configuration and pitch in a classically trained soprano singer. *Journal of Voice, 23*, 546-551.
- Sundberg, J. (2013). Perception of Singing. In D. Deutsch (Ed.), *The psychology of music* (pp. 69-105). San Diego, CA: Academic Press.
- Sundberg, J., & Bauer-Huppmann, J. (2007). When does a sung tone start? *Journal of Voice, 21*, 285–293.
- Sundberg, J., Gramming, P., & Lovetri, J. (1993). Comparisons of pharynx, source, formant, and pressure characteristics in Operatic and Musical Theatre singing. *Journal of Voice, 7*, 301-310.

- Sundberg, J., Gu, L., Huang, Q., & Huang, P. (2012). Acoustical study of classical Peking Opera singing. *Journal of voice*, 26(2), 137-143.
- Sundberg, J., La, F. M., & Gill, B. P. (2013). Formant tuning strategies in professional male opera singers. *Journal of Voice*, 27, 278-288.
- Sundberg, J., La, F. M., & Himonides, E. (in press). Intonation and Expressivity: A Single Case Study of Classical Western Singing. *Journal of voice*.
- Sundberg, J., Leanderson, R., & Von Euler, C. (1989). Activity Relationship between diaphragm and cricothyroid muscles. *Journal of Voice*, 3, 225-232.
- Sundberg, J., & Nordenberg, M. (2006). Effects of vocal loudness variation on spectrum balance as reflected by the alpha measure of long-term average spectra of speech. *Journal of the Acoustic Society of America*, 120, 453-457.
- Sundberg, J., Prame, E., & Iwarsson, J. (1996). Replicability and accuracy of pitch patterns in professional singers. In P. J. Davis, & N. H. Fletcher (Eds.), *Vocal fold physiology, controlling complexity and chaos* (pp. 291-306). San Diego, CA: Singular Publishing Group.
- Sundberg, J., & Romedahl, C. (2009). Text intelligibility and the singer's formant – a relationship? *Journal of Voice*, 23, 539-545.
- Sundberg, J., & Thalén, M. (2010). What is “Twang”? *Journal of Voice*, 24, 654-660.
- Sundberg, J., Thalén, M., & Popeil, L. (2012). Substyles of belting: phonatory and resonatory characteristics. *Journal of Voice*, 26, 44-50.
- Tafari, J. (2009). *Infant Musicality*. Farnham, UK: Ashate Publishing.

- Tan, S. -L., Pfordresher, P. Q., & Harré, R. (2010). *Emergence of sound and music perception Psychology of Music: From sound to significance* (pp. 133-154). United Kingdom: Taylor & Francis Group.
- Terao, Y., Mizuno, T., Shindoh, M., Sakurai, Y., Ugawa, Y., Kobayashi, S., Ngai, C., Furubayashi, T., Arai, N., Okabe, S., Mochizuki, H., Hanajima, R., & Tsuji, S. (2006). Vocal amusia in a professional tango singer due to a right superior temporal cortex infarction. *Neuropsychology*, 44(3), 479-488.
- Ternström, S., Sundberg, J., & Colldén, A. (1988). Articulatory F<sub>0</sub> perturbations and auditory feedback. *Journal of Speech, Language, and Hearing Research*, 31, 187-192.
- Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schröger, E. (2005). Pitch-discrimination accuracy in musicians vs. nonmusicians - an event-related potential and behavioral study. *Experimental Brain Research*, 161, 1-10.
- Thalén, M., & Sundberg, J. (2001). Describing different styles of singing. A comparison of a female singer's voice source in "Classical", "Pop", "Jazz" and "Blue". *Logopedics Phoniatics Vocology*, 26, 82-93.
- Thompson, W. F., Graham, P., & Russo, F. A. (2005). Seeing music performance: Visual influences on perception and experience. *Semiotica*, 156(1/4), 203-227.
- Thompson, W. F., Peter, V., Olsen, K. N., & Stevens, C. J. (2012). The effect of intensity on relative pitch. *The Quarterly Journal of Experimental Psychology*, 65(10), 2054-2072.
- Thompson, W. F., & Russo, F. A. (2007). Facing the music. *Psychological Science*, 18, 756-757.

- Thompson, S., & Williamon, A. (2003). Evaluating evaluation: Musical performance assessment as a research tool. *Music Perception, 21*, 21-41.
- Thorpe, C. W., Cala, S. J., Chapman, J., & Davis, P. J. (2001). Patterns of breath support in projection of the singing voice. *Journal of Voice, 15*, 86-104.
- Titze, I. R. (1989). On the relation between subglottal pressure and fundamental frequency in phonation. *Journal of the Acoustical Society of America, 85*(2), 901-906.
- Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, NJ: Prentice-Hall.
- Titze, I. R. (2000). *Principles of voice production* (2<sup>nd</sup> edition). Iowa City, IA: National Center for Voice and Speech.
- Titze, I. R., & Sundberg, J. (1992). Vocal intensity in speakers and singers. *Journal of the Acoustical Society of America, 91*(5), 2936-2946.
- Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit Learning of tonality : A self-organizing approach. *Psychological review, 107*(4), 885-913.
- Trainor, L. J. (2005). Are there critical periods for musical development ? *Developmental Psychobiology, 46*(3), 262-278.
- Trainor, L. J., & Hannon, E. E. (2012). Musical development. In D. Deutsch (Ed.), *The Psychology of Music* (pp. 423-497). San Diego, CA: Academic Press.
- Trainor, L. J., Marie, C., Gerry, D., Whiskin, E., & Unrau, A. (2012). Becoming musically enculturated: effects of music classes for infants on brain and behavior. *Annals of the New York Academy of Sciences, 1252*, 25-36.



- Trainor, L. J., & Trehub, S. E. (1992). A comparison of infants' and adults' sensitivity to Western musical structure. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 394–402.
- Trehub, S. E., Bull, D., & Thorpe, L. A. (1984). Infants' perception of melodies: The role of melodic contour. *Child Development*, *55*, 821–830.
- Trehub, S. E., Thorpe, L. A., & Morrongiello, B. A. (1987). Organizational processes in infants' perception of auditory patterns. *Child Development*, *58*, 741–749.
- Tremblay-Champoux, A., Dalla Bella, S., Phillips-Silver, J., Lebrun, M., & Peretz, I. (2010). Singing proficiency in congenital amusia: Imitation helps. *Cognitive Neuropsychology*, *27*(6), 463–476.
- van Besouw, R. M., Brereton, J. S., & Howard, D. M. (2008). Range of tuning for tones with and without vibrato. *Music Perception*, *26*, 145–155.
- van Besouw, R. M., & Howard, D. M. (2008). Effects of carrier and phase on the pitch of long-duration vibrato tone. *Musicae Scientiae*, *8*(1), 139–161.
- Vurma, A., Raju, M., & Kuuda, A. (2010). Does timbre affect pitch?: Estimations by musicians and non-musicians. *Psychology of Music*, *39*(3), 291–306.
- Vurma, A., & Ross, J. (2006). Production and perception of musical intervals. *Music Perception*, *23*, 331–344.
- Wapnick, J., Darrow, A. A., Kovacs, J., & Dalrymple, L. (1997). Effects of physical attractiveness on evaluation of vocal performance. *Journal of Research in Music Education*, *45*, 470–479.

- Wapnick, J., Flowers, P. J., Alegant, M., & Jasinskas, L. (1993). Consistency in piano performance evaluation. *Journal of Research in Music Education, 41*, 282–292.
- Wapnick, J., Mazza, J., & Darrow, A. A. (1998). Effects of performer attractiveness, stage behavior, and dress on violin performance evaluation. *Journal of Research in Music Education, 46*, 510–521.
- Wapnick, J., Mazza, J., & Darrow, A. A. (2000). Effects of performer attractiveness, stage behavior, and dress on evaluation of children's piano performances. *Journal of Research in Music Education, 48*, 323–336.
- Wapnick, J., Ryan, C., Campbell, L., Deek, P., Lemire, R., & Darrow, A. (2005). Effects of excerpt tempo and duration on musicians' ratings of high-level piano performances. *Journal of Research in Music Education, 53*, 162–176.
- Ward, W. D. (1954). Subjective musical pitch. *Journal of the Acoustical Society of America, 26*, 369–380.
- Warrier, C. M., & Zatorre, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. *Perception & Psychophysics, 64*(2), 198–207.
- Watts, C., Barnes-Burroughs, K., Andrianopoulos, M., & Carr, M. (2003). Potential factors related to untrained singing talent: a survey of singing pedagogues. *Journal of Voice, 17*(3), 298–307.
- Watts, C., Barnes-Burroughs, K., Estis, J., & Blanton, D. (2006). The singing power ratio as an objective measure of singing voice quality in untrained talented and nontalented singers. *Journal of Voice, 20*(1), 82–88.
- Watts, C. R., & Hall, M. D. (2008). Timbral influences on vocal pitch-matching accuracy. *Logopedics Phoniatrics Vocology, 33*, 74–82.

- Watts, C., Moore, R., & McCaghren, K. (2005). The relationship between vocal pitch matching skills and pitch discrimination skills in untrained accurate and inaccurate singers. *Journal of Voice, 19*(4), 534-543.
- Watts, C., Murphy, J., & Barnes-Burroughs, K. (2003). Pitch matching accuracy of trained singers, untrained subjects with talented singing voices, and untrained subjects with nontalented singing voices in conditions of varying feedback. *Journal of Voice, 17*(2), 185-194.
- Weill-Chounlamountry, A., Soyez-Gayout, L., Tessier, C., & Pradat-Diehl, P. (2008). Vers une rééducation cognitive de l'amusie. *Annales de réadaptation et de médecine physique, 51*(5), 332-341.
- Welch, G. F. (1979). Poor pitch singing: A review of the literature. *Psychology of Music, 7*, 50-58.
- Welch, G. F. (1979). Vocal range and poor pitch singing. *Psychology of Music, 7*, 13-31.
- Welch, G. F. (1986). A developmental view of children's singing. *British Journal of Music Education, 3*(3), 295-303.
- Welch, G. F. (1994). The assessment of singing. *Psychology of Music, 22*, 3-19.
- Welch, G. F. (1998). Early childhood musical development. *Research Studies in Music Education, 11*, 27-41.
- Welch, G. F. (2002). Early Childhood Musical Development. In L. Bresler & C. Thompson (Eds.), *The Arts in Children's Lives: Context, Culture and Curriculum* (pp. 113-128). Dordrecht, NL: Kluwer.
- Welch, G. F. (2005). Singing as communication. In D. Miell, R. MacDonald & D. Hargreaves (Eds.), *Musical communication* (pp. 239-259). Oxford: Oxford University Press.

- Welch, G. F. (2006). Singing and vocal development. In G. McPherson (Ed.), *The child as musician: A handbook of musical development* (pp. 311–329). Oxford: Oxford University Press.
- Welch, G. F. (2009). Evidence of the development of vocal pitch matching ability in children. *Japanese Journal of Music Education Research*, 39(1), 38-47.
- Welch, G. F., Rush, C., & Howard, D. M. (1991). A developmental continuum of singing ability: Evidence from a study of five-year-old developing singers. *Early Child Development and Care*, 69, 107-119.
- Welch, G. F., Sergeant, D. C., & White, P. (1996). The singing competences of five-year old developing singers. *Bulletin of the Council for Research in Music Education*, 127, 155-162.
- Welch, G. F., Sergeant, D. C. & White, P. (1997). Age, sex and vocal task as factors in singing 'in-tune' during the first years of schooling. *Bulletin of the Council for Research in Music Education*, 133, 153-160.
- Welch, G. F., Sergeant, D. C., & White, P. (1998). The role of linguistic dominance in the acquisition of song. *Research Studies in Music Education*, 10, 67-74.
- Welch, G. F., & Sundberg, J. (2002). Solo voice, in R. Parncutt & G. E. McPherson (Eds). *The science and psychology of music performance: creative strategies for teaching and learning* (pp. 253-268). Oxford : Oxford University Press.
- Wilsonarboleda, B., & Frederick, A. (2008). Considerations for maintenance of postural alignment for voice production. *Journal of Voice*, 22(1), 90-99.

- Wise, K. J., & Sloboda, J. A. (2008). Establishing an empirical profile of self-defined "tone deafness": Perception, singing performance and self-assessment. *Musicae Scientiae*, 12(1), 3-26.
- Wise, K. J., Sloboda, J. A., & Peretz, I. (2007). Progress in understanding "tone deafness". *British Academy Review*, 10, 52-54.
- Wong, P. C. M., Ciocca, V., Chan, A. H. D., Ha, L. Y. Y., Tan, L. -H., & Peretz, I. (2012). Effects of culture on musical pitch perception. *PLoS One*, 7(4), e33424.
- Yarbrough, C., Bowers, J., & Benson, W. (1992). The effect of vibrato on the pitch-matching accuracy of certain and uncertain singers. *Journal of Research in Music Education*, 40, 30-38.
- Yerkes, R. M., & Dodson, J. D. (1908). The relationship of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology and Psychology*, 18, 459-482.
- Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.
- Yoshie, M., Kudo, K., Murakoshi, T., & Ohtsuki, T. (2009). Music performance anxiety in skilled pianists: Effects of social-evaluative performance situation on subjective, autonomic, and electromyographic reactions. *Experimental Brain Research*, 199(2), 117-126.
- Yoshie, M., Kudo, K., & Ohtsuki, T. (2008). Effects of psychological stress on state anxiety, electromyographic activity, and arpeggio performance in pianists. *Medical Problems of Performing Artists*, 23, 120-132.
- Yoshie, M., Shigemasu, K., Kudo, K., & Ohtsuki, T. (2008). Effects of state anxiety on music performance: Relationship between the Revised Competitive State Anxiety inventory-2 subscales and piano performance. *Musicae Scientiae*, 13, 55-84.

- Zangger Borch, D., Sundberg, J., Lindestad, P. -A., & Thalén, M. (2003). Vocal fold vibration and voice source aperiodicity in phonatorily distorted singing. *TMH-QPSR*, 45(1), 087-091.
- Zarate, J. M., Ritson, C. R., & Poeppel, D. (2012). Pitch-interval discrimination and musical expertise: is the semitone a perceptual boundary? *Journal of the Acoustical Society of America*, 132(2), 984-993.
- Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: auditory-motor interactions in music perception and production. *Nature Reviews Neuroscience*, 8(7), 547-558.
- Zhang, Y., Jiang, J. J., & Wallace, S. M. (2005). Comparison of nonlinear dynamic methods and perturbation methods for voice analysis. *Journal of the Acoustical Society of America*, 118, 2551-2560.