Published in : Logopedics Phoniatrics Vocology (2019)
DOI: 10.1080/14015439.2019.1659410
Status: Postprint (Author's version)

IMITATING DYSPHONIC VOICE: A SUITABLE TECHNIQUE TO CREATE SPEECH STIMULI FOR SPOKEN LANGUAGE PROCESSING TASKS?

Isabel S. Schiller[a], Angélique Remacle[ab] and Dominique Morsomme[a]

[a]*Faculté de Psychologie, Logopédie et Sciences de L'Éducation, Université de Liège, Liège, Belgium*
[b]*Fund for Scientific Research - F.R.S. - FNRS, Brussels, Belgium*

ABSTRACT

Purpose: The aim of this study was to assess the suitability of imitated dysphonic voice samples for their application in listening tasks investigating the impact of speakers' voice quality on spoken language processing. *Methods*: A female voice expert recorded speech samples (sustained vowels and connected speech) in her normal voice and while imitating a dysphonic voice. Voice characteristics, authenticity, and consistency of the two voice qualities were evaluated by means of acoustic measurements (Acoustic Voice Quality Index [AVQI], jitter, shimmer, harmonics-to-noise ratio [HNR]) and perceptual evaluation (GRBAS scale, consistency, and authenticity rated by five speech-language pathologists). *Results*: Based on acoustic and perceptual assessments, the degree of voice impairment for the imitated dysphonic voice was found to be moderate to severe. Roughness and asthenia were the predominant perceptual features. The perceptual rating indicated a high consistency and acceptable authenticity of the imitated dysphonic voice. *Conclusions:* Results suggest that an imitation of dysphonic voice quality may resemble the voice characteristics typically found in dysphonic patients. Implications: The voice samples validated here shall be applied in future listening tasks and may promote our understanding of how dysphonic speech is processed.

KEYWORDS: Imitation of dysphonia; mimicking dysphonic voice; voice quality; voice assessment; dysphonic voice; voice authenticity; voice naturalness; speech samples; listening tasks; speech perception

Published in : Logopedics Phoniatrics Vocology (2019)
DOI: 10.1080/14015439.2019.1659410
Status: Postprint (Author's version)

Introduction

During the past decade, an increasing amount of studies have investigated the effect of speaker's impaired voice quality on spoken language processing [1-4]. In listening experiments, participants are presented with speech samples in a normal and/or dysphonic voice and perform a linguistic task, such as sentence-picture matching. The impact of voice quality on spoken language processing is then assessed in terms of answer accuracy or response time. To draw conclusions applying to real-life listening situations, the voice qualities used in such tasks should be selected carefully. This study evaluates the ecologic suitability of an imitated dysphonic voice for assessing spoken language processing in listening tasks.

Spoken language processing refers to a listener's ability to process acoustic information and map it onto linguistic representations which can then be manipulated and memorized to allow the understanding of speech [5]. Research suggests that listening to impaired voice may impede spoken language processing at different stages ranging from low-level speech perception to high-level listening comprehension [1-4,6-12]. This may be due to the increased noise components characterizing dysphonic voices [13]. Compensation for such signal degradations is assumed to increase the cognitive load, thereby leaving less capacity available for understanding speech [14].

The impact of impaired voice on the listening experience has received particular attention in children. Children's language development could be negatively affected in case their mother's or father's voice was dysphonic. In therapeutic situations, for example psychotherapy or SLT sessions, children might feel disturbed if their therapist's voice was impaired. In the educational context, past studies repeatedly addressed the question whether children are still able to listen effectively and recall oral information if their teacher is dysphonic [1-3,7]. Effects of normal versus impaired voice on children's spoken language processing have been investigated in sentence-picture matching tasks [3,11], passage comprehension tasks [1,2,4,6,7,9,10], and phoneme discrimination tasks [1]. To date, there is no consensus on how dysphonic voice samples should best be obtained. Three methods were applied in the past: (1) recording a real dysphonic patient, (2) provoking a dysphonic voice by means of a vocal loading task, or (3) mimicking dysphonic voice. A fourth option would be the use of synthesized dysphonic voice. Advantages and drawbacks of these four methods are discussed in the following paragraphs.

Unquestionably, using speech samples from real dysphonic speakers is favorable in terms of ecologic validity. This method was applied by Morton and Watson [6] who were the first to investigate the effect of impaired voice on spoken language processing in children. Children were presented with text passages read by a vocally healthy speaker versus a speaker diagnosed with moderate to severe dysphoria. Children's ability to recall words was significantly better under the normal voice condition than the dysphonic voice condition. However, despite the authors' claim of having controlled for normal articulation, neutral accent and comparable speaking rates, the impact of speaker-dependent variables may not be ruled out. Distinct prosodic features, voice characteristics such as timber, or articulatory

Published in : Logopedics Phoniatrics Vocology (2019)
DOI: 10.1080/14015439.2019.1659410
Status: Postprint (Author's version)

differences between the speakers shape the listening experience and may thus affect listening task results.

Such potential confounding factors may be overcome when speech samples are produced by the same person. Dysphonic voice in otherwise voice-healthy speakers may be provoked through vocal loading tasks - a method repeatedly applied by Lyberg-Åhlander and associated researchers [2-4,8,9,12], who followed the procedure described in Whitling, Rydell, and Lyberg-Åhlander [15]. In this vocal loading task, which is claimed to temporarily provoke impaired voice [8], speakers read out a text against background noise leveled at 85 dB SPL until the point when they "feel a discomfort in their throat" (maximum time limit of 30 minutes) [15, p.261.el5]. Provoking impaired voice with this technique is interesting as it reflects real classroom situations, in which high background noise levels may increase teachers' vocal effort. The problem is that some speakers may be resistant to vocal loading [16]. Note also that only mild to moderate degrees of dysphonia were achieved in the above-mentioned studies [2-4,8,9,12], ranging from 4 to 5 on an 11-point scale. While hyperfunctional voice quality was provoked, hoarseness was not. In fact, results from another study suggests that some voice quality parameters, such as breathiness, might even improve as a result of vocal loading [16]. Considering these findings, the effectivity of vocal loading tasks provoking dysphonic voice for listening tasks remains therefore questionable.

Imitation of dysphonic voice is another technique for obtaining different voice qualities from the same speaker. Several researchers have used imitated dysphonic voice for listening tasks [1,7,10,11]. Again, the benefit is that speaker-dependent vocal-, prosodic-, or articulatory features are controlled for. As opposed to provoking impaired voice through vocal loading, this method allows speakers to simulate different degrees of dysphonia or emphasize particular voice characteristics such as roughness, breathiness, or hoarseness. However, not every speaker may authentically mimic an impaired voice. It is also challenging to maintain a consistent impaired voice quality throughout the recording. Impersonators could perform this task, but while they are able to mimic the voice of another speaker [17], they might not necessarily be able to modify their own voice. Mimicking another person's speaking style is different to making one's own voice sound dysphonic yet natural. This may be one of the reasons why past studies made recordings of voice experts with profound knowledge of dysphonia [1,7,10,11].

Finally, speech synthesis could be a way to generate different voice qualities for listening tasks. Compared to the three methods presented above, speech synthesis would offer the highest control of voice parameters over time. Distinct voice characteristics could be manipulated to obtain the voice quality of interest for the listening experiment. In the context of dysphonic voice creation, speech synthesis has primarily been performed on sustained vowels [18-26] or vowel combinations [22,27]. To our knowledge, synthesis of dysphonic voice in connected speech has only been performed by Yiu and colleagues [28,29], who addressed the problem of limited naturalness of the samples [29]. To assess the effect of dysphonic voice on spoken language processing, researchers require dysphonic samples of connected speech which sound natural. It seems that speech synthesis technology cannot yet respond to that need.

In order to evaluate dysphonia, voice assessment involves a combination of multiple approaches. In clinical practice, ENTs or speech-language pathologists diagnose voice

Published in : Logopedics Phoniatrics Vocology (2019)
DOI: 10.1080/14015439.2019.1659410
Status: Postprint (Author's version)

disorders using laryngoscopy, aerodynamic measurements, self-evaluation, perceptual assessment and acoustic measurements. Here, we are interested in the description of voice as related to the listening experience. In the following, we will therefore address the issue of voice quality evaluation on a perceptual and acoustic level.

As voice is above all a perceptual phenomenon [30], perceptual assessment has a high clinical relevance and is often considered the gold standard for evaluating voice quality [30-32]. A wide range of standardized and non-standardized rating instruments for perceptual voice assessment exist. One of the most common perceptual rating tools used in clinics is the GRBAS scale [33]. The GRBAS scale encompasses five voice quality parameters, namely grade (G), roughness (R), breathiness (B), asthenia (A), and strain (S), which are rated on a 4-point ordinal scale (0 = no pathology, 1 = mild pathology, 2 = moderate pathology, and 3 = severe pathology).

Acoustic measurements represent an objective supplement to perceptual voice quality assessment. Common acoustic measures used in clinical practice are jitter (i.e. frequency perturbations), shimmer (i.e. intensity perturbations) and harmonics-to-noise ratio (HNR), which is the relation of harmonic parts of the spectrum compared to non-harmonic parts. While these measures are calculated from sustained vowels, the Acoustic Voice Quality Index (AVQI) [34] allows acoustic voice analyses based on vowels in combination with connected speech. The strength of the AVQI is thus its high ecologic validity [34]. Combining different acoustic markers from the domains time, frequency, and quefrency, the AVQI provides a score that predicts the degree of dysphonia severity [35]. AVQI scores range between 0 and 10 (the higher the scores, the more severe the degree of dysphonia). Cut-off values for the distinction between normal and pathologic voice reside around 3 (3.05 for French) [36]. A paper on an updated version of the French AVQI is currently in press [37].

Based on perceptual voice assessments and acoustic measurements, the present study assessed the suitability of imitated dysphonic voice quality for listening tasks investigating spoken language processing. The imitated dysphonic voice samples shall subsequently be used in two studies: a laboratory experiment and a field experiment conducted in a real classroom. We recorded speech material (vowels and connected speech) of a female speech-language pathologist using her normal versus imitated dysphonic voice. This material was then evaluated in terms of authenticity, consistency, and voice quality characteristics. Three purposes were served: (1) validating speech material for future experiments, (2) sharing this speech material with other researchers, and (3) providing recommendations for the creation of imitated dysphonic voice samples.

Methods

RECORDING OF NORMAL AND DYSPHONIC SPEECH SAMPLES

For the recording procedure, we followed the recommendations provided in Barsties and De Bodt [38]. Recordings were made in a quiet room with a background noise level of 30dB(A) (as measured with a PCE-353 sound level meter, PCE Holding GmbH, Germany). The speaker wore a head-mounted condenser microphone (C 544 L, AKG Acoustics GmbH,

Published in : Logopedics Phoniatrics Vocology (2019)
DOI: 10.1080/14015439.2019.1659410
Status: Postprint (Author's version)

Austria), which was connected to a Lenovo laptop (IdeaPad, U430p, Lenovo, China) via an external soundcard (iTrackSolo, Focusrite Audio Engineering Ltd., China). Recordings were digitalized at a sampling frequency of 44.1 kHz sampling frequency and 16 bit resolution using audacity software (http://audacityteam.org/). Speech material consisted of the following:

- The first sentence of the phonetically balanced text "La bise et le soleil"
- Two randomly selected sentences from the Epreuve du Langage Orale (ELO) (Oral Language Assessment) subtest C2 [39]
- Six randomly selected pseudo-words from the Epreuve Lilloise de Discrimination Phonologique (ELDP) [40]
- The sustained vowel /a:/

The speaker was a 51-year-old vocologist with an experience of 26 years in diagnostics and treatment of voice disorders. She was a native speaker of French and grew up in the Wallonian Region of Belgium. For the first recording, the speaker used her normal voice at an intensity typically used in conversations. For the second recording she imitated a dysphonic voice while trying to maintain a comparable intensity. Before recording, she practiced the imitation of a dysphonic voice based on a previous audio file of her own voice during a severe laryngitis. During the recording, another voice specialist provided feedback to the speaker regarding the quality, authenticity, and consistency of her voice production.

ACOUSTIC ANALYSIS

Acoustic analyses were performed using the speech processing software Praat version 6.0.29 [41]. Analyses were based on two audio files per voice quality, one of the 3-second mid-vowel portion of /a:/, one of connected speech (i.e. "La bise et le soleil se disputaient, chacun assurant qu'il était le plus fort").

AVQI scores were computed using the script provided by Maryn et al. [34], based on the concatenation of the sustained vowel and connected speech. Complementary acoustic analyses were run on the sustained vowel to compare both voice qualities in terms of periodicity (jitter [local] and shimmer [local]) and harmonicity (HNR) measures.

PERCEPTUAL ANALYSIS

A questionnaire was created for the perceptual analysis of the two voice qualities. The goal was to confirm whether the normal voice was perceived as healthy and the imitated impaired voice as authentic and consistently dysphonic across the two respective samples.

Five independent female raters performed the perceptual voice assessment. They were all speech-language pathologists, native speakers of French, and blind to the aim of the study and the identity of the speaker. Average work experience in the field of speech-language pathology was 6 years (range = 1-22 years). Three out of five raters reported to treat dysphonic patients on a weekly basis.

Raters were instructed to listen to and evaluate four audio samples in a strict sequence. Sample 1 should be assessed before going on to sample 2, sample 3, and sample 4. Audio files are publicly available in the NODYS database [42] which was established in the context

Published in : Logopedics Phoniatrics Vocology (2019)
DOI: 10.1080/14015439.2019.1659410
Status: Postprint (Author's version)

of this study. Table 1 provides details on content, voice quality, and duration of the audio samples. Perceptual analysis for each sample included the GRBAS scale [33] and an evaluation of authenticity and consistency of voice quality. Authenticity was assessed by asking the rater to indicate how natural the respective voice sounded based on a 4-point scale (natural, rather natural, rather unnatural, unnatural). Consistency of each voice quality across stimuli was assessed by asking the rater to indicate how similar the dysphonic samples (i.e. Sample 1 and 2) and the normophonic samples (i.e. Sample 3 and 4) sounded to one another. Again, raters provided their answer using a 4-point scale (similar, rather similar, rather different, and different).

Sample	Voice quality	Linguistic content	Duration(sec)
1	Dysphonic	"La bise et le soleil se disputaient, chacun assurant qu'il était le plus fort" + /a:/	11
2	Dysphonic	/itãʀʏ/ /kopitʏl/ /mydƐzo/ + "La petite fille est lavée par le garçon"	9
3	Normal	"La bise et le soleil se disputaient, chacun assurant qu'il était le plus fort" + /a:/	10
4	Normal	/kafiʃygR/ /kopityn/ /bYRʃolã/+ "Je mange les cerises que maman cueille"	9

Table 1. *Details on speech samples used for perceptual assessment*.

STATISTICAL ANALYSIS

Acoustic analysis was performed descriptively. For the perceptual assessment, we calculated the modes of the GRBAS scale and the authenticity and consistency rating. Inter-rater reliability was measured using Light's kappa [43], which is suitable for fully-crossed designs with categorical variables and multiple raters [44]. First, we computed kappa values for pairwise comparisons based on evaluations made by each rater (i.e. GRBAS, authenticity, and consistency rating). Then the arithmetic mean of all kappa values was calculated as a measure of inter-rater reliability. Cohen's kappa [45] was used to calculate intra-rater reliability. In the context of this study, intra-rater reliability refers to the degree to which each rater was consistent with her own scoring of the dysphonic samples (i.e. comparison between a rater's responses for Sample 1 and 2) and likewise the normal voice samples (i.e. comparison between a rater's responses for Sample 3 and 4).

Results

ACOUSTIC ANALYSIS

The results from the acoustic analysis are presented in Table 2, in which acoustic parameters of the normal and imitated dysphonic voice are compared. The difference in AVQI scores for the normal voice and the imitated dysphonic was 4.36, with a higher score for the impaired voice.

Published in : Logopedics Phoniatrics Vocology (2019)
DOI: 10.1080/14015439.2019.1659410
Status: Postprint (Author's version)

We obtained higher jitter and shimmer values, and lower HNR values for the dysphonic voice. In other words the imitated dysphonic voice showed a higher degree of aperiodicity and a lower degree of harmonicity.

Parameter	Normal voice	Imitated dysphonic voice
AVQI	2.53	6.89
Jitter (local)	0.314%	2.772 %
Shimmer (local)	1.386 %	9.177 %
HNR	25.26 dB	10.84dB

Table 2. *Results from the acoustic analysis.*

PERCEPTUAL ANALYSIS

Taking into account the global results from the perceptual analysis, a moderate degree of inter-rater reliability was found, with a kappa coefficient of κ = 0.52. Table 3 lists the κ-statistics for each combination of raters. For the normal voice quality, a perfect intra-rater agreement was found (i.e. each rater gave identical scores for the two samples). For the imitated dysphonic voice, the kappa coefficient of κ = 0.95 indicates almost perfect intra-rater agreement. Details are provided in Table 4.

	Rater 2 *κ*	Rater 3 *κ*	Rater 4 *κ*	Rater 5 *κ*
Rater 1	0.53	0.47	0.46	0.65
Rater 2		0.69	0.30	0.66
Rater 3			0.52	0.50
Rater 4				0.40

Table 3. *Inter-rater reliability for the perceptual evaluation based on GRBAS, authenticity, and consistency rating for each pair of raters.*

Rater	κ
Rater 1	1.0
Rater 2	1.0
Rater 3	0.75
Rater 4	0.50
Rater 5	1.0

Table 4*. Intra-rater reliability for the perceptual evaluation based on the comparison of GBRAS ratings for Sample 1 and 2.*

GRBAS results indicated that the normal voice was perceived as healthy and the imitated impaired voice as pathologic. All raters scored all GRBAS parameters of the normal voice with 0. GRBAS results for the imitated impaired voice are presented in Figure 1. Scores are

Published in : Logopedics Phoniatrics Vocology (2019)
DOI: 10.1080/14015439.2019.1659410
Status: Postprint (Author's version)

based on mode values. Grade (G), roughness (R), and asthenia (A) were mostly rated with a score of 3, breathiness (B) with a score of 2 and strain (S) with a score of 1.

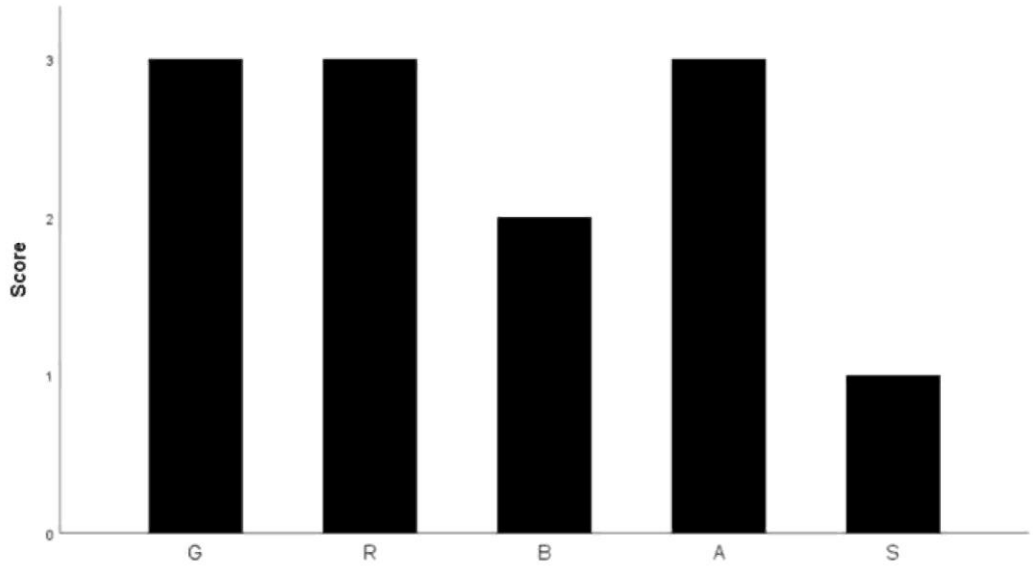

Figure 1. *GRBAS rating for the imitated impaired voice. GRBAS parameters are shown on the x-axis. The y-axis shows the mode score attributed to each parameter, ranging from 0 (no pathology) to 3 (severe pathology).*

Authenticity was assessed by asking raters how natural the voice quality sounded to them. All raters consistently provided the same responses for the two samples of imitated impaired voice and normal voice respectively. Four raters evaluated the normal voice as natural, one rater as rather unnatural. The imitated dysphonic voice was evaluated as natural by 2/5 raters, rather natural by 1/5 raters, and rather unnatural by 2/5 raters.

Consistency of each voice quality across speech stimuli was assessed based on raters' evaluation of similarity. Similarity was evaluated for the comparison of sample 1 and 2 (both imitated dysphonic voice) and the comparison of sample 3 and 4 (both normal voice quality). For both comparisons, 4/5 raters evaluated the respective voice quality as similar to the one in the previous sample. One rater evaluated both voice quality comparisons as rather similar.

Discussion

Several studies have assessed the effect of speaker's dysphonic voice on spoken language processing [1-4,6-12]. This study investigated if imitated dysphonic voice samples are suitable for this purpose. A speaker's normal voice was compared to her imitation of a dysphonic voice using perceptual and acoustic measures. Results suggest that the speaker succeeded in imitating a moderately to severely impaired voice of an authentic and consistent quality.

Published in : Logopedics Phoniatrics Vocology (2019)
DOI: 10.1080/14015439.2019.1659410
Status: Postprint (Author's version)

ACOUSTIC ANALYSIS

AVQI results for the speaker's normal voice were within the non-pathologic range (i.e. a score of 2.53 on a scale from 0-10 where 10 indicates the highest degree of dysphonia), while results for the imitated dysphonic voice indicated a moderate to severe voice pathology (i.e. a score of 6.89) [34-36]. Interpretations are based on the French cut-off value of 3.05, which was established in a cross-linguistic study comparing AVQIs performance in English, French, Dutch, and German [36]. AVQIs cross-linguistic criterion-related concurrent validity was lowest for French, which might relate to the fact that all speakers were Dutch native speakers. For future research, it is therefore advisable to use a modified version of the French AVQI [37].

The perturbation and harmonicity measures obtained in the present study point in the same direction as our AVQI results. Praat does not provide jitter or shimmer thresholds for the discrimination between normal and dysphonic voice and refers to thresholds proposed by the MDVP (Kay Elemetrics Corporation, Lincoln Park, NJ, USA): 1.040% for jitter and 3.20% for shimmer [41]. Compared with these values, our jitter- and shimmer values for the normal voice were within the healthy range and values for the imitated dysphonic voice were within the pathological range. Slight variations in jitter- and shimmer values according to software and algorithms used for calculation cannot be ruled out [38]. HNR values also indicated that normal voice was non-pathologic, and the imitated dysphonic voice pathologic [41]. Considering all values, the acoustic parameters extracted from the sustained /a:/ suggest a high degree of roughness for the imitated impaired voice quality.

PERCEPTUAL ANALYSIS

Findings of the perceptual assessment are in line with the acoustic results. Statistical results indicated moderate inter-rater reliability, which complies with past research [46,47], and almost perfect intra-rater reliability. All five raters perceived the normal voice as healthy and the imitated impaired voice as pathologic. While the normal voice was consistently rated with 0 points regarding the overall grade of dysphonia, the imitated dysphonic voice received scores from 2 to 3, indicating a moderate to severe pathology. Results from the GRBAS scale [33] also showed that most raters perceived the imitated impaired voice as severely rough (R) and asthenic (A). Moreover, the majority of raters indicated moderate breathiness (B) and mild strain (S) for that voice. When compared to the perceptual ratings of provoked impaired voices used in past studies [2-4,8,9,12], our findings suggest that hoarseness might be generated more successfully through imitation than vocal loading tasks. On the contrary, provoking dysphonic voice through vocal loading tasks might be a more effective technique when the aim is to generate vocal hyperfunction [15,16].

In addition to the GRBAS rating, we were interested if the speaker's voice quality was perceived as authentic and consistent across speech stimuli. First, raters evaluated how natural the two voice qualities sounded to them. The normal voice was perceived as natural and the impaired voice as rather natural. The latter results were interpreted as an indication of acceptable authenticity of the imitated impaired voice. Second, raters listened to two samples of each voice quality and evaluated how similar they sounded to one another. The two samples of each voice quality were perceived to be similar to one another, with a high

Published in : Logopedics Phoniatrics Vocology (2019)
DOI: 10.1080/14015439.2019.1659410
Status: Postprint (Author's version)

degree of consistency. Seemingly, the speaker was capable of maintaining the same degree of dysphonia across samples. We argue that this was thanks to her expertise in voice disorders, her prior practicing based on an audio sample of her real dysphonic voice, and the feedback of another voice specialist during sample recording.

LIMITATIONS, RECOMMENDATIONS, AND FUTURE DIRECTIONS

In this paper, we presented a first approach to determine the suitability of creating imitated impaired voice samples for listening tasks. The evaluation of our imitated impaired voice quality was based on reference values associated with healthy or impaired voice qualities (i.e. AVQI scores, jitter-, shimmer-, and HNR values, as well as GRBAS scores), as well as authenticity and consistency ratings. No direct comparison with real dysphonic voice samples was drawn. Each dysphonic voice has unique voice characteristics and there is no direct link between perceptual voice quality and underlying source of dysphonia. For example, voice patients with the same source of pathology (e.g. nodules) may show different degrees of hoarseness [48]. The informative value of a comparison between our imitated dysphonic voice and a real dysphonic voice would, therefore, be questionable.

Our study bears five methodological limitations that we address in the following. First, perceptual ratings were based on only two audio samples per voice quality. There is a remaining uncertainly as to whether these samples were truly representative for voice quality consistency throughout the entire recording. Second, we analyzed recordings from a single speaker to control for speaker-dependent confounding factors, such as F0 differences, prosodie aspects, or articulatory differences. General validity of our results is thus restricted. Third, only one dysphonic voice quality (i.e. a moderately to severely dysphonic voice quality, predominantly perceived as rough) was recorded. For the future, it would be interesting to compare different imitations of dysphonic voice. Fourth, we did not ask the raters to perform an authenticity rating on a real dysphonic voice. This might have been useful to confirm they actually assessed the intended underlying concept. Finally, test-retest reliability of the perceptual assessment was not determined. Nevertheless, raters assessed two different audio samples per voice quality, which gives a good indication of how consistent they were in their responses.

For researchers who aim to investigate effects of imitated dysphonic voice in listening tasks, we offer the following recommendations: First, we propose that an expertise with dysphonic patients may help to imitate a dysphonic voice. Second, the recording session should be preceded by a practice session in the presence of another voice expert, who provides feedback to enhance authenticity and consistency of the dysphonic voice imitation. For this purpose, a previous recording of the speaker's voice during an episode of dysphonia may provide a valuable orientation. Third, recordings should be made in compliance with the guidelines published by Barsties and De Bodt [38]. Fourth, a perceptual voice assessment of the normal and imitated voice qualities should be performed by several independent voice experts. We also recommend that perceptual assessment should be based on a sample selection that will later be used in the listening task. Finally, performing an acoustic analysis is advisable to objectively describe the voice characteristics and relate them to perceptual results or, in a next step, results obtained in listening experiments.

Published in : Logopedics Phoniatrics Vocology (2019)
DOI: 10.1080/14015439.2019.1659410
Status: Postprint (Author's version)

We hope that in the future it will also be possible to use synthesized dysphonic speech samples for listening tasks. To date, impaired voice quality has mainly been synthesized in vowels [18-27]. Advancements in the synthesis of dysphonic voice qualities in connected speech would have important applications in research. The use of synthesized speech would allow researchers to emphasize or attenuate specific voice parameters to study them respectively. Real speakers may be less precise in performing such manipulations. For the purpose of this study, we created the NODYS database [42] containing normophonic and imitated dysphonic speech samples. The aim is to complement the NODYS database with synthesized normophonic and dysphonic speech samples in the future, as well as further imitated dysphonic voice samples. This may help us to investigate how distinct voice characteristics shape listeners' perception or attitude towards the speaker, how they might impede or promote spoken language processing, or whether they have an impact on memory functions in listeners.

Moreover, future research should take into account other factors which may impede children's listening ability in addition to voice quality. Examples are signal-to-noise ratio or speech rate. Their potential interactions with impaired voice quality are still underdetermined.

Conclusion

Dysphonic voice samples may be created following these four methods: (1) recording dysphonic speakers, (2) provoking dysphonic voice in voice-healthy speakers, (3) asking voice-healthy speakers to imitate dysphonic voice and (4) synthesizing dysphonic voice. We assessed the suitability of imitated dysphonic voice samples for the use in listening tasks.

Acoustical and perceptual evaluation of the imitated dysphonic voice indicated a moderate to severe degree of voice pathology, a high voice quality consistency across speech samples, and an acceptable degree of authenticity. These results suggest that listeners may perceive an imitation of a dysphonic voice as realistic. We argue that the voice samples evaluated here, represent suitable material for upcoming listening experiments, which will assess voice quality effects on spoken language processing.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Université de Liège under grant number RD/DIR-vdu/2016.7166, and by the National Fund for Scientific Research - F.R.S. - FNRS. (Brussels, Belgium).

Published in : Logopedics Phoniatrics Vocology (2019)
DOI: 10.1080/14015439.2019.1659410
Status: Postprint (Author's version)

Notes on contributors

Isabel S. Schiller is a PhD student in the Faculty of Psychologie, Speech and Language Therapy, and Educational Sciences at the University of Liège, Belgium. She is attached to the Unit of Psychology and Neurosciences of Cognition. Her PhD is funded by the University of Liège and focusses on the effects of dysphonic voice and noise on children's spoken language processing.

Angélique Remacle is a researcher funded by the Fund for Scientific Research - F.R.S. - FNRS. Her research takes place at the Department of Speech Therapy, Université de Liège, Belgium. She is a visiting professor in Speech and Language Therapy at the Université libre de Bruxelles, Belgium.

Dominique Morsomme is a professor in the Faculty of Psychology, Speech and Language Therapy, and Educational Sciences at the University of Liège, Belgium. She is a lecturer in the Department of Speech and Language Therapy and heads the Voice Unit. She also works as a vocologist at the Hospital of Liège. Her clinical activities focus on the evaluation of voice disorders and on voice feminization.

ORCID

Isabel S. Schiller https://orcid.org/0000-0003-2387-7625

Angélique Remacle https://orcid.org/0000-0001-9338-977X

Dominique Morsomme https://orcid.org/0000-0002-7697-0498

Data availability statement

Audio files of Sample 1, 2, 3, 4, and of pseudo-word lists in healthy versus dysphonic voice are openly available from the NODYS database [42] in Mendeley data.

References

[1] Morsomme D, Minel L, Verduyckt I. Impact of teachers' voice quality on children's language processing skills. VOCOLOGIE: Stem Stemstoornissen. 2011;9-15.

[2] Rudner M, Lyberg-Åhlander V, Brännstrôm J, et al. Listening comprehension and listening effort in the primary school classroom. Front Psychol. 2018;9:1193.

[3] Sahlén B, Haake M, von Lochow H, et al. Is children's listening effort in background noise influenced by the speaker's voice quality? Logoped Phoniatr Vocol. 2018;43:47-55.

[4] von Lochow H, Lyberg-Åhlander V, Sahlén B, et al. The effect of voice quality and competing speakers in a passage comprehension task: performance in relation to cognitive functioning in children with normal hearing. Logoped Phoniatr Vocol. 2018;43:32-41.

[5] Medwetsky L. Spoken language processing model: bridging auditory and language processing to guide assessment and intervention. LSHSS. 2011;42:286-296.

Published in : Logopedics Phoniatrics Vocology (2019)
DOI: 10.1080/14015439.2019.1659410
Status: Postprint (Author's version)

[6] Morton V, Watson DR. The impact of impaired vocal quality on children's ability to process spoken language. Logoped Phoniatr Vocol. 2001;26:17-25.

[7] Rogerson J, Dodd B. Is there an effect of dysphonic teachers' voices on children's processing of spoken language? J Voice. 2005;19:47-60.

[8] Brännström KJ, Kastberg T, von Lochow H, et al. The influence of voice quality on sentence processing and recall performance in school-age children with normal hearing. J Speech Lang Hear Res. 2018;21:1-9.

[9] Brännström KJ, von Lochow H, Lyberg-Åhlander V, et al. Immediate passage comprehension and encoding of information into long-term memory in children with normal hearing: the effect of voice quality and multitalker babble in noise. Am J Audiol. 2018;27:231-237.

[10] Chui JCH, Ma E. The Impact of dysphonic voices on children's comprehension of spoken language. J Voice. 2018;pii: S0892-1997:30487-30488.

[11] Lyberg-Åhlander V, Haake M, Brännström J, et al. Does the speaker's voice quality influence children's performance on a language comprehension test? Int J Speech Lang Pathol. 2015; 17:63-73.

[12] Lyberg-Åhlander V, Holm L, Kastberg T, et al. Are children with stronger cognitive capacity more or less disturbed by classroom noise and dysphonic teachers? Int J Speech Lang Pathol. 2015;17:577-588.

[13] Yanagihara N. Significance of harmonic changes and noise components in hoarseness. J Speech Hear Res. 1967;10: 531-541.

[14] Mattys SL, Davis MH, Bradlow AR, et al. Speech recognition in adverse conditions: a review. Lang Cognit Process. 2012;27: 953-978.

[15] Whitling S, Rydell R, Lyberg-Åhlander V. Design of a clinical vocal loading test with long-time measurement of voice. J Voice. 2015;29:261.el3-261.e27.

[16] Remacle A, Schoentgen J, Finck C, et al. Impact of vocal load on breathiness: perceptual evaluation. Logoped Phoniatr Vocol. 2014;39:139-146.

[17] Delvaux V, Caucheteux L, Huet K, et al. Voice disguise vs. impersonation: acoustic and perceptual measurements of vocal flexibility in non experts. Proceedings of the 18th Annual Conference of the International Speech Communication Association (Interspeech); 2017 August 20-24; Stockholm, Sweden: ISCA; 2018. 3777-3781.

[18] Lucero JC, Schoentgen J, Behlau M. Physics-based synthesis of disordered voices. In: Bimbot F, editor. Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech); 2013 August 25-29; Lyon, France: Isca; 2014. 587-591.

[19] Sofranko JL, Prosek RA. The Effect of experience on response time when judging synthesized voice quality. J Voice. 2014;28:24-35.

[20] Gerratt BR, Kreiman J. Measuring vocal quality with speech synthesis. J Acoust Soc Am. 2001;110:2560-2566.

[21] Natour YS, Saleem AF. The performance of the time-frequency analysis software (TF32) in the acoustic analysis of the synthesized pathological voice. J Voice. 2009;23:414-424.

[22] Schoentgen J, Fraj S, Lucero JC. Testing the reliability of grade, roughness and breathiness scores by means of synthetic speech stimuli. Logoped Phoniatr Vocol. 2015;40:5-13.

Published in : Logopedics Phoniatrics Vocology (2019)
DOI: 10.1080/14015439.2019.1659410
Status: Postprint (Author's version)

[23] Murphy PJ, McGuigan KG, Walsh M, et al. Investigation of a glottal related harmonics-to-noise ratio and spectral tilt as indicators of glottal noise in synthesized and human voice signals. J Acoust Soc Am. 2008;123:1642-1652.

[24] Englert M, Madazio G, Gielow I, et al. Perceptual error analysis of human and synthesized voices. J Voice. 2017;31:516-5e5.

[25] Fraj S, Schoentgen J, Grenez F. Development and perceptual assessment of a synthesizer of disordered voices. J Acoust Soc Am. 2012;132:2603-2615.

[26] Kreiman J, Antonanzas-Barroso N, Gerratt BR. Integrated software for analysis and synthesis of voice quality. Behav Res Methods. 2010;42:1030-1041.

[27] Bergan CC, Titze IR, Story B. The perception of two vocal qualities in a synthesized vocal utterance: ring and pressed voice. J Voice. 2004;18:305-317.

[28] Yiu EM, Murdoch B, Hird K, et al. Perception of synthesized voice quality in connected speech by Cantonese speakers. J Acoust Soc Am. 2002;112:1091-1101.

[29] Yiu EML, Murdoch B, Hird K, et al. Cultural and language differences in voice quality perception: a preliminary investigation using synthesized signals. Folia Phoniatr Logop. 2008;60:107-119.

[30] Oates J. Auditory-perceptual evaluation of disordered voice quality. Folia Phoniatr Logop. 2009;61:49-56.

[31] De Bodt MS, Van de Heyning PH, Wuyts FL, et al. The perceptual evaluation of voice disorders. Acta Otorhinolaryngol Belg. 1996;50:283-291.

[32] Wuyts FL, De Bodt MS, Van de Heyning PH. Is the reliability of a visual analog scale higher than an ordinal scale? An experiment with the GRBAS scale for the perceptual evaluation of dysphonia. J Voice. 1999;13:508-517.

[33] Hirano M. Clinical examination of voice. Disord Hum Commun. 1981;5:1-99.

[34] Maryn Y, Corthals P, Van Cauwenberge P, et al. Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels. J Voice. 2010;24:540-555.

[35] Maryn Y, Weenink D. Objective dysphonia measures in the program Praat: smoothed cepstral peak prominence and acoustic voice quality index. J Voice. 2015;29:35-43.

[36] Maryn Y, De Bodt M, Barsties B, et al. The value of the acoustic voice quality index as a measure of dysphonia severity in subjects speaking different languages. Eur Arch Otorhinolaryngol. 2014;271:1609-1619.

[37] Pommée T, Maryn Y, Finck C, et al. Validation of the Acoustic Voice Quality Index, version 03.01, in French. J Voice. 2019;pii: S0892-1997:30517-30514.

[38] Barsties B, De Bodt M. Assessment of voice quality: current state-of-the-art. 2015. Auris Nasus Larynx. 2015;42:183-188.

[39] Khomsi A. ELO: Evaluation du Langage Orale [language test]. ECPA Pearson. 2001;

[40] Macchi L, Descours C, Girard E, et al. ELDP: Epreuve Lilloise de Discrimination Phonologique [ELDP1 protocol & manual]. 2018. [cited 2019 August 28]. Available from: http://orthopho-nie.univ-lille2.fr/stocks/stock-contents/epreuve-lilloise-de-dis-crimination-phonologique.html.

[41] Boersma P, Weenink D. Praat (version 6.0.29) [Computer software]. 2017. [cited 2019 August 28]. Available from: http:// www.fon.hum.uva.nl/praat/.

Published in : Logopedics Phoniatrics Vocology (2019)
DOI: 10.1080/14015439.2019.1659410
Status: Postprint (Author's version)

[42] Schiller I, Remade A, Morsomme D. NODYS: NOrmophonic and DYsphonic Speech samples [database]. 2019 Jan 18 [cited 2019 Jan 18]. In: Mendeley data [internet]. Available from: http://dx.doi.org/10.17632/g2fmkw8t85.l

[43] Light RJ. Measures of response agreement for qualitative data: some generalizations and alternatives. Psychol Bull. 1971;76:365.

[44] Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. TQMP. 2012;8:23.

[45] Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20:37-46.

[46] Kreiman J, Gerratt BR. Sources of listener disagreement in voice quality assessment. J Acoust Soc Am. 2000;108: 1867-1876.

[47] Xie Z, Gadepalli C, Farideh J, et al. Machine learning applied to GRBAS voice quality assessment. Adv Sci Tech Eng Syst J. 2018; 3:329-338.

[48] Shah RK, Engel SH, Choi SS. Relationship between voice quality and vocal nodule size. Otolaryngol Head Neck Surg. 2008; 139:723-726.