The evaluation of singing voice accuracy: Are we good judges even if we are not musicians?

Larrouy-Maestri P., Roig-Sanchis V., Morsomme D.

Department of Psychology – University of Liège - Belgium
What is “in tune”? 

Accurate version

Contour error at the note 2

Interval errors of 200 cents between notes 1-2 and 2-3

Tonality modulation at the note 2
How to evaluate the vocal accuracy?

Overview

- Self-evaluation
  - Asking the singer himself
    - Difficult (Cuddy et al., 2005; Sloboda et al., 2005; Wise & Sloboda, 2008)
    - Under-estimation (Pfordresher & Brown, 2007)

- Music experts
  - Grids, scales, global score
    - e.g. Alcock et al., 2000a, 2000b; Hébert et al., 2003; Lévêque et al., 2012; Racette et al., 2006; Schön et al., 2004; Wise & Sloboda, 2008

- Computer-assisted method
  - Acoustic analyses, extraction F0, computation of errors
    - e.g. Dalla Bella & Berkowska, 2009; Dalla Bella et al., 2007; Lévêque et al., 2009; Pfordresher & Brown, 2007; Pfordresher et al., 2010
Are music experts good judges?


Procedure described in Larrouy-Maestri & Morsomme (in press)
- Estimation of the f0 for each note
- Computation of errors
  - Number of contour errors
  - Mean interval deviation
  - Number of modulations

166 sung performances

18 Musicians

http://sldr.org/sldr000774/en

1----2----3----4----5----6----7----8----9
very inaccurate
very accurate
Conclusions

High correlation between the raters \((r = .77, p < .01)\)
- Same definition of vocal accuracy

High correlation between the two methods \((r = .87, p < .001)\)
- Objectivity of the judges
- 3 judges are sufficient to keep a correlation of .85

81% of the variance of the judges is explained
- Pitch interval deviation \((\beta = 0.51; p < .001)\)
- Number of tonality modulations \((\beta = 0.45; p < .001)\)

Answer...
But …

| Effects of music expertise | We are all experts of our culture  
(Bigand & Delbé, 2010) |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discrimination</strong></td>
<td>• Early development of melodic perception</td>
</tr>
<tr>
<td>Micheyl et al., 2006;</td>
<td>Chang &amp; Trehub, 1977; Ferland &amp; Mendelson, 1989;</td>
</tr>
<tr>
<td>Tervaniemi et al., 2005</td>
<td>Plantinga &amp; Trainor, 2005; Stalinski et al., 2008</td>
</tr>
<tr>
<td><strong>Pitch perception</strong></td>
<td>• Enculturation</td>
</tr>
<tr>
<td>Hutchins &amp; Peretz, 2012</td>
<td>Miyamoto, 2007; Stalinski &amp; Schellenberg, 2012;</td>
</tr>
<tr>
<td></td>
<td>Trainor, 2005; Trainor et al., 2012</td>
</tr>
<tr>
<td><strong>Interval size estimation</strong></td>
<td>• Implicit learning</td>
</tr>
<tr>
<td>Russo &amp; Thompson, 2005</td>
<td>Jonaitis &amp; Saffran, 2009; Loui et al., 2010;</td>
</tr>
<tr>
<td></td>
<td>Saffran et al., 1999; Schön et al., 2008</td>
</tr>
<tr>
<td><strong>Complex tones</strong></td>
<td>• Melodic expectations</td>
</tr>
<tr>
<td>Hutchins et al., 2012;</td>
<td>Marmel et al., 2008</td>
</tr>
<tr>
<td>Vurma et al., 2010;</td>
<td>• Ability to sing in tune</td>
</tr>
<tr>
<td>Zarate et al., 2012</td>
<td>Dalla Bella et al., 2007</td>
</tr>
<tr>
<td><strong>Error detection</strong></td>
<td></td>
</tr>
<tr>
<td>Fujiroka et al., 2004;</td>
<td></td>
</tr>
<tr>
<td>Hutchins et al., 2012 ;</td>
<td></td>
</tr>
<tr>
<td>Warrier &amp; Zatorre, 2002</td>
<td></td>
</tr>
<tr>
<td><strong>Melodic recognition</strong></td>
<td></td>
</tr>
<tr>
<td>Orsmond &amp; Miller, 1999</td>
<td></td>
</tr>
</tbody>
</table>
Are we good judges?

- Methods
- Results
- Conclusions
Methods

Acoustic analyses

166 sung performances

http://sldr.org/sldr000774/en

18 NON Musicians

Procedure described in Larrouy-Maestri & Morsomme (in press)
- Estimation of the f0 for each note
- Computation of errors
  - Number of contour errors
  - Mean interval deviation
  - Number of modulations

1---2---3---4---5---6---7---8---9
very inaccurate very accurate
Methods

- Participants
  - Paired in age, gender, sociocultural background
  - Non musicians
    - Less than 2 years of musical training
    - Occasional listeners
  - Audio and MBEA OK

- Procedure
  - 166 x “Happy birthday”
    - 5 lists in random order
  - Global pitch accuracy
  - Two times (test and retest with 8-15 days in between)
Are we good judges?

- Methods
- Results
- Conclusions
Results and discussion

Judges reliability

- Intra-judge reliability
  - Mean Spearman correlation T1/T2: .66 (SD = .06)
    \(\Rightarrow\) OK

- Inter-judge reliability
  - Intra-class correlation coefficient: .89; \(p < .01\)
    \(\Rightarrow\) OK

- Correlation with computer-assisted method
  - \(r(166) = .81; p < .01\)
    \(\Rightarrow\) Higher scores for accurate performances: OK
Results and discussion

Correlation with decreased number of judges

Non experts

Experts
Results and discussion
Comparison Non experts / Experts

- Comparison mean ratings
  - $r(166) = .84; p < .01$
  - $U$ Mann Whitney $p < .01$
  - Non experts are more severe than experts

- Criteria

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Non experts</th>
<th>Experts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>$F(3,165) = 104.44; p &lt; .01$</td>
<td>$F(3,165) = 231.51; p &lt; .01$</td>
</tr>
<tr>
<td>% variance</td>
<td>66%</td>
<td>81%</td>
</tr>
<tr>
<td>Interval deviation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tonality modulations</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Are we good judges?

- Methods
- Results
- Conclusions
Conclusions
Are we good judges?

YES

Correlation with Music Experts
Correlation with computer assisted method
Inter-judge reliability
Intra-judge reliability

NO

Difference with Music Experts
Bigger groups of raters
Just one criterion
% of variance

Larrouy-Maestri, Roig-Sanchis, & Morsomme  08/08/2013
Conclusions
Are we good judges?

- **Reliability and objectivity**
  - Similar definition of vocal accuracy
  - Same strategy each time
  - Relationships with music experts’ ratings and objective method

- **However**
  - More severe
  - Just sensitive to the precision of the intervals
  - 34% of the variance unexplained

- **Perspectives**
  - Other criteria which influence the rating of non musicians
  - Perception and tolerance of music experts and non experts
  - Why?
The evaluation of singing voice accuracy

THANK YOU FOR YOUR ATTENTION
References


References

References


