Within- and Between-Talker Variability in Voice Quality in Normal Speaking Situations

Jody Kreiman, Patricia Keating, Soo Jin Park, Shaghayegh Rastifar, and Acerbe Alwan

1 Department of Head and Neck Surgery, School of Medicine, University of California, Los Angeles, USA
2 Department of Linguistics, University of California, Los Angeles, USA
3 Department of Electrical Engineering, University of California, Los Angeles, USA

Little is known about how, and how much, individual talkers vary their voice quality across normal everyday speaking situations. In theory, this makes it difficult to specify what we actually mean when we describe a voice as an “auditory pattern.” In practice, lack of knowledge about normal within-talker variability limits ability to predict or explain confusions among voices.

**Current Objectives**

1. Are talkers consistently more similar to themselves than they are to other talkers?
2. What is the relationship between acoustic and perceptual similarity?

**Step 1: Develop a database**

**Database Description**

- 200 UCLA undergraduate talkers (100 female)
- 3 recording sessions on separate days
- A wide range of speaking tasks chosen to sample normal day-to-day and situation-to-situation variation in voice quality

- Steady-state vowels (all sessions)
- Paired sentences (all sessions)
- Spontaneous speech
- Conversational speech
- Lower- and higher-affect speech
- Pet-directed speech
- Lower- and higher-affect speech
- Harvard sentences (all sessions)
- Steady-state vowels (all sessions)

**Introduction**

This work was supported by NSF under grant number IIS 1450992, and by NIH/NIDCD under grant number DC 01797.

**Database Sampling**

- 200 UCLA undergraduate talkers (100 female)
- 3 recording sessions on separate days
- A wide range of speaking tasks chosen to sample normal day-to-day and situation-to-situation variation in voice quality

**Acknowledgments**

Thanks to Anya Haniccia and Brenda Garcia for recording the speakers in the database.

**Acoustic Analyses**

**Selection of measures**

- A large set of measures (F0, H1*-H2*, H2*-H4*, H4*-2K*, 2K*-5K, CPP, F1 - F3)
- Measured every 100 msec across entire duration of each vowel
- Source measures collected with VoiceSauce and validated with analysis-by-synthesis
- Formant frequencies measured using the Snack option within VoiceSauce

**Acoustic distance**

- Measures normalized from 0-1 using known ranges
- Average (Euclidean) acoustic distance with equal weighting of measures

**A wide range of speaking tasks chosen to sample normal day-to-day and situation-to-situation variation in voice quality**

**Acoustic distance**

- Measures normalized from 0-1 using known ranges
- Average (Euclidean) acoustic distance with equal weighting of measures

**Within- and Between-Talker Variability in Voice Quality in Normal Speaking Situations**

**Perceptual Experiment**

- 180 “same talker” pairs and 450 “different talker” pairs
- 2 randomized sets, each of which was divided into 3 subsets (total 6 subsets)
- 60 normal hearing listeners in 6 groups
- Each pair played once in each order (AB/BA)
- Listeners judged whether the voices represented one talker or two different talkers

**Results and discussion**

- Listener accuracy: Hit rates were quite high overall, but false alarm rates were also high (Table 3), suggesting that listeners had difficulty distinguishing different talkers.
- D’ rates in Table 4 confirm that the listening task was difficult.

**Table 3:**

<table>
<thead>
<tr>
<th>Talker # misses</th>
<th>Predictive variables</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>41.8</td>
<td>.83</td>
</tr>
<tr>
<td>2</td>
<td>90</td>
<td>.45</td>
</tr>
<tr>
<td>3</td>
<td>262</td>
<td>.21</td>
</tr>
<tr>
<td>4</td>
<td>128</td>
<td>.22</td>
</tr>
</tbody>
</table>

**Table 4:**

<table>
<thead>
<tr>
<th>Talker A</th>
<th>Talker B</th>
<th>d’ Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>.90</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>.83</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>.91</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>.92</td>
</tr>
</tbody>
</table>

**Acoustic Analyses**

**Result and discussion**

- Within-talker acoustic variability predicted incorrect “same talker” responses.
- Parameters implicated in failures of self-similarity varied from talker to talker.

**Table 5:**

<table>
<thead>
<tr>
<th>Talker</th>
<th>F0</th>
<th>F1, F2, H1*-H2*, CPP, F1 - F3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>1.00</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1.00</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1.00</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>1.00</td>
</tr>
</tbody>
</table>