Multidimensional Representation of Personal Quality of Vowels and its Acoustical Correlates
(Matsumoto, Hiki, Sone, Nimura; 1973)

Pedro Davalos
CPSC 689-604
Feb 27, 2007
Outline

• Introduction
• Test 1: /a/
• Test 2: Hybrid
• Test 3: Vowels
• Conclusions
Intro: Goal

• Determine how acoustical properties influence recognizing speakers.
Intro: Background

• “Personal Quality”
  – Is NOT high quality as required to perform at an opera
  – It refers to the speaker’s characteristics and the voice attributes that allow speaker recognition
Intro: Approach

Psychological Auditory Space (PAS)

\[
\begin{bmatrix}
v_1 \\
v_2 \\
\vdots \\
v_n
\end{bmatrix}
\]

Classified by Listeners

Voice Samples

Sensory Auditory Space (Physical Space)

Feature Extraction

Acoustical Parameters

Kruskal’s Scaling

\[ S = \text{mdscale}(X,d) \]

Graph Theory

Multiplication Correlation

\[ C = \text{xcorr2}(S,P) \]

Linear Regression

\[ V = \text{regress}(p,C) \]
Intro: Acoustical Parameters

- Mean Fundamental Pitch Frequency
  - $\log F_0$
- Fluctuation of Fundamental Pitch Period
  - $\sigma(\Delta T/T)$
- Slope of Glottal Source Spectrum
  - $\alpha$
- Formant Frequencies
  - $F_1, F_2, F_3$

Glottal Source Characteristics - $U(s)$

Vocal Tract Characteristics - $T(s)$

$P(s) = U(s) T(s) R(s)$
Intro: ____ Recognition

Speech Recognition:
They are all the same!
⇒ “Hello World”

Speaker Recognition:
They are all the different!
⇒ s1, s2, s3, & s4
Test 1 - /a/: Specs

- **Data Samples:**
  - 8 speakers, vowel /a/ at 3 freq: 120, 140, & 160 Hz → 24 samples

- **Listener Testing:**
  - 6 listeners, listen 9 times to each pair twice (order) → 108 values/pair
  - Listeners classify voice pairs as “same talker” or “different talker”
Test 1 - /a/: Results

3D-PAS

F1 & F2 are related to both A1 & A2

Correlation between PAS and Acoustical Parameters

Lower F0, greater contribution to personal quality
Test 1 - /a/: ROC

Receiver Operating Characteristics

Lower F0, greater contribution to personal quality
Test 1 - /a/: Results (var)

<table>
<thead>
<tr>
<th>MEAN FUNDAMENTAL PITCH FREQUENCY</th>
<th>FLUCTUATION OF FUNDAMENTAL PITCH PERIOD</th>
<th>SLOPE OF GLOTTAL SOURCE SPECTRUM</th>
<th>FORMANT FREQUENCIES, $F_1$, $F_2$, AND $F_3$</th>
<th>EXPLAINED VARIANCE IN %</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>86</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>84</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>84</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>81</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>71</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>55</td>
</tr>
</tbody>
</table>
Test 2 – Hybrid: Specs

• Data Samples:
  – 5 speakers, vowel /a/ at 140 Hz.
  – Data set altered by generating fixed glottal source (removing fluctuation of fundamental pitch period variable)

• 6 listeners repeat 10 trials each pair
Test 2 – Hybrid: Results

- F3 became similar to F1 & F2
- Vocal Tract has greater contribution than Glottal (other than F0) since hybrid voices tend to be closer to the original with the same formants.

2D PAS of Hybrid Voices
$V_g$: V-Formant, g-glottal source
Test 3 – Vowels: Specs

• Data Samples
  – 8 Speakers, 5 vowels (40 Voices) all at 164 Hz

• Listeners
  – 13 people listened 3 times to all voice pairs
  – (78 Samples)
Since Talkers are clustered,

The perceptual cues of personal quality common to different vowels is involved in listener judgment
Test 3 – Vowels: Results (ROC)

Receiver Operating Characteristics

[Diagram showing Receiver Operating Characteristic (ROC) curves for different vowel sounds.]
Test 3 – Vowels: Results (xcorr)

α: Slope of glottal source spectrum

σ(ΔT/T): Rapid fluctuation of pitch period

Large Correlations and similar directions
Test 3 – Vowels: Results (var)

### TABLE II
Relation Between Various Sets of the Acoustical Parameters and the Explained Variance of the Configuration of Voice Set II on the PAS

<table>
<thead>
<tr>
<th>Glottal Source Characteristics</th>
<th>Vocal Tract Characteristics</th>
<th>Explained Variance in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fluctuation of P fundamental</td>
<td>Source Spectra, $F_1$, $F_2$, and $F_3$</td>
<td>/i/ /e/ /a/ /o/ /u/</td>
</tr>
<tr>
<td>Pitch Period</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
Conclusions (1)

• F0 is the relative most significant contributor to perception of personal quality

• Vocal Tract and Glottal Characteristics contribute to different perceptual dimensions from each other with F0 constant

• Vocal Tract contributions to perception of personal quality varies with different vowels

• The perceptual dimensions of F0, F1, $\alpha$-slope of glottal, and fluctuation of F0 period are independent of vowel
Conclusions (2)

• Authors claim success because…
  – Talkers cluster on the A1-A2 PAS
  – The P(c) from the listeners was about 60-70%
  – There is uniformity of the results despite different listeners
  – Acoustical parameters were found to influence perception of personal quality

• Future Work:
  – Evaluate other Parameters