Frequency shifts and vowel identification

Peter F. Assmann (School of Behavioral and Brain Sciences, Univ. of Texas at Dallas, Box 830688, Richardson TX 75083)
Terrance M. Nearey (Dept. of Linguistics, University of Alberta, Edmonton, Alberta, Canada T6E 2G2).
Introduction

- Listeners can understand frequency-shifted speech across a wide frequency range (Fu & Shannon, 1999).
- We hypothesize that this ability can be explained in terms of listeners’ sensitivity to statistical variation across talkers in natural speech.
- The aims of the present study were:
  1. To study the effects of frequency shifts on the identification of vowels spoken by 2 men, 2 women and 2 children (age 7).
  2. To test the predictions of a model of vowel perception that incorporates measures of fundamental frequency ($F_0$) and formant frequencies (FF) associated with size differences in larynx and vocal tract across talkers.
Co-variation of formant frequencies and F0 in natural speech

Mean log FF: Geometric mean of formant frequencies: F1,F2,F3
>3000 vowels in hVd words (Assmann & Katz, 2000)
Pattern recognition model

- Parameters: duration, mean $F_0$, and $F_1$, $F_2$, $F_3$ sampled at 20% and 80% points
- Training data: 3000+ vowels spoken by 10 men, 10 women and 30 children from the N. Texas region (Assmann & Katz, 2000)
- *A posteriori* probabilities derived from linear discriminant analysis for each stimulus vowel
Frequency shifts and vowel identification

• In a previous study (Assmann, Nearey & Scott, 2002) we confirmed that upward shifts in F0 or formant frequencies (FF) resulted in lower vowel identification accuracy.

• However, combining upward shifts in F0 with upward shifts in FF led to improved identification accuracy.

• The finding that vowel identification accuracy is higher with coordinated shifts in F0 and FF is well predicted by the model of vowel identification outlined below, and supports the idea that listeners are sensitive to the pattern of co-variation of F0 and FF in natural speech.
Vowel Identification Accuracy

Means and standard errors of 11 listeners

Predicted means

(Assmann, Nearey, and Scott, ICSLP 2002).
Vowel Identification Experiment

- The present study examined effects of upward and downward frequency shifts on vowel identification.
- 11 vowels (/i/, /ɪ/, /ɛ/, /ɛ/, /æ/, /ʌ/, /ə/, /ɔ/, /ʊ/, /u/) in hVd context spoken by 3 men, 3 women, and 3 children from the N. Texas region.
- Upward and downward frequency shifts were introduced by means of the STRAIGHT vocoder (Kawahara, 1997).
STRAIGHT vocoder

• High-resolution analysis of time-varying spectrum envelope
• Wavelet-based instantaneous frequency $F_0$ extraction
  – Spectrum envelope ($\mathbf{FF}$) scaling
  – Fundamental frequency ($\mathbf{F}_0$) scaling
Scale Factors

<table>
<thead>
<tr>
<th>FF scale factors</th>
<th>0.6</th>
<th>0.8</th>
<th>1.0</th>
<th>1.5</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>F₀ scale factors</td>
<td>0.5</td>
<td>1.0</td>
<td>4.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- For females and children, downward shifts tend to produce male-like voices; for adults, upward shifts heard as child-like voices.
Method

• Listeners were 14 Psychology undergraduates participating for partial course credit. Since the majority had no phonetics training, they first completed 3 practice sets:

  Set 1: passive listening with feedback (24 resynthesized but not frequency-shifted vowels; no response required).

  Set 2: practice identification (a different set of 24 vowels presented for identification; repeated until a score of 21/24 or better was obtained).

  Set 3: passive listening with feedback (24 frequency-shifted vowels; shift factors randomly chosen from the 15 conditions of the experiment; no response required)
Method

- Main experiment: 990 syllables (11 vowels x 2 talkers per group x 3 talker groups x 3 F0 scale factors x 5 FF scale factors). All conditions randomly interspersed.
- Vowels were presented diotically over headphones in a double-walled sound booth. Listeners identified the vowels using an 11-button response box drawn on computer screen labeled with keywords for the vowel category.
Effects of FF shifts

Identification accuracy (%)

Spectrum envelope scale factor

Men
Women
Children
Interaction of FF and F0 shifts

- For men’s vowels, accuracy is higher when upward shift in FF is accompanied by upward shift in F0.
- For women and children, there is a recovery from downward shifts in FF when F0 is also shifted down.
Conclusions

- Identification accuracy drops significantly when vowels are shifted upward in formant frequency by a factor of 1.5 or more, or downward by a factor of 0.6 or less. Adult males are less susceptible to upward shifts than females and children, while children are less affected by downward shifts.
- In several conditions, the drop in intelligibility was reduced by combining formant shifts with corresponding changes in F0.
- Pattern recognition models predicted the effects of frequency shifts on vowel identification, including the synergistic link between F0 and formant frequency. A plausible account is that learned relationships between F0 and spectral envelope cues are responsible for this interaction.
References


Average correct ID per synthetic voice
Disk and spoke plot
Disks = Observed ID

- The colored disks represent listeners correct identification rate
  - **Blue**: male speakers’ synthesized voices (scaled and unscaled); **Red**: female speakers; **Green**: child speakers;
  - The position of the center of the disk indicates the average F0 and formant frequencies of the voice
  - The area of each disk is proportional to the average % corrected identification by listeners of the voice
  - The circles in the legend box indicate the correct identification rate
Disk and spoke plot
Spokes = Predicted ID

• Length of the spokes indicate predicted ID rate by LDFA
  – Trained on natural measurements of Assmann and Katz, predictions on scaled values of this experiment
  – Patterns:
    • Accurate predictions: ‘Basketball’ spoke length matches disk radius
    • Under predictions: ‘Asterisks’ in disks, listeners do better
    • Over predictions: ‘Spiked disks’, model does better than listeners
% ID well predicted by smooth function of mean F0 and mean FF
Accounts for 88% of variance