A REVIEW OF RESULTS ON FREQUENCY SHIFTS AND VOWEL IDENTIFICATION

Peter F. Assmann[1] and Terrance M. Nearey[2]
1) School of Behavioral and Brain Sciences, University of Texas at Dallas, Richardson, TX 75083 USA assmann@utdallas.edu
2) Department of Linguistics, University of Alberta, Edmonton, AB T6G 2E7 Canada tnearey@ualberta.ca

1. INTRODUCTION

Listeners experience a wide range of variation in fundamental frequency (F_0) and formant frequencies in everyday speech communication. Recent studies have shown that speech remains intelligible when the spectrum envelope is shifted up or down along the frequency scale across a fairly wide range. However, intelligibility declines when the shift factor is greater than about 1.5, or less than 0.7 (Fu and Shannon, 1999). In this paper we present evidence from a perceptual experiment and describe a model of vowel identification that accounts for the effects of frequency shifts in terms of listeners' sensitivity to statistical regularities in natural speech. We focus on the effects of upward and downward shifts in spectrum envelope on the identification of vowels in hVd syllables.

Figure 1 illustrates the statistical covariation of F_0 and formant frequencies in spoken vowels associated with size differences in the larynx and vocal tract across talkers. The figure shows acoustic measurements of a set of more than 3000 vowels spoken by men, women and children from the north Texas region. There is a moderate correlation between F_0 and formant frequencies (r=0.45). On average, the change from male to female speech involves an upward shift of 15-20% in formant frequencies, accompanied by an 80-100% increase in F_0.

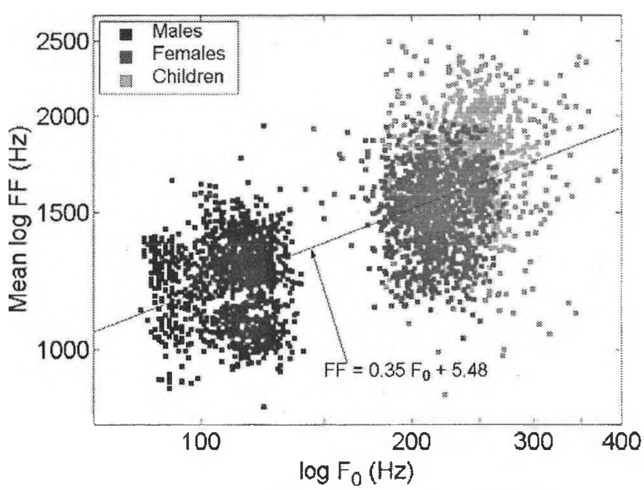

Fig. 1. Geometric mean of formant frequencies (F1-F3) vs. F0 for 11 vowels in hVd words spoken by 10 males, 10 females and 30 children aged 3-7 years (Assmann and Katz, 2000).

Assmann, Nearey and Scott (2002) used a high-quality source-filter vocoder called STRAIGHT (Kawahara, 1997) to investigate the interaction of F_0 and formant frequencies in frequency-shifted vowels. The baseline stimuli were /hVd/ words spoken by three adult male talkers. Identification accuracy dropped by more than 30% when the formant frequencies were shifted upward by a factor of 2.0, and by 50% when F0 was scaled by a factor of 4.0. However, when upward shifts in formant frequencies were combined with upward shifts in F0, performance improved, suggesting that listeners are sensitive to the pattern of covariation of F0 and formant frequencies in natural speech.

A pattern recognition model of vowel identification was implemented, using the modeling strategy described by Nearey and Hillenbrand (1999). The model incorporated measurements of the formant frequencies (F1,F2,F3) sampled at the 20% and 80% points in the vowel, combined with the mean F0 and duration. Linear discriminant analysis was carried out to obtain *a posteriori* probabilities of group membership for each frequency-shifted vowel. The training data for the model was a set of 3000+ vowels (examples of each of the 11 vowels from 50 talkers, including 10 males, 10 females, and 30 children from the north Texas region). The model accurately reproduced the overall pattern of results from the experiment, including the decline in identification when F0 and formant frequencies were shifted separately and the improvement when upward shifts in F0 were combined with upward shifts in the formant frequencies.

The experiment reported here extended this approach, applying both upward and downward frequency shifts to vowels from a larger sample of talkers, including adult males, adult females and children.

2. METHOD

The stimuli were /hVd/ words (*heed*, *hid*, *hayed*, *head*, *had*, *hud*, *hawed*, *hoed*, *hood*, *who'd*, *herd*) spoken by two adult males, two adult females, and two 7-year old children. These vowels were selected from a larger sample of vowels recorded from 10 men, 10 women, and 30 children, ages 3, 5, and 7 years from the North Texas region (Assmann and Katz, 2000). The STRAIGHT vocoder was used to resynthesize each vowel in 15 conditions (3 levels of F_0 shift factor x 5 levels of spectrum envelope shift factor):

spectrum envelope scale factor = 0.6, 0.8, 1.0, 1.5, 2.0

F_0 scale factor = 0.5, 1.0, 4.0

Vowel=/ɪ/, /I/, /ε/, /E/, /Θ/, /℘/, /∈∏/, /A/, /o/, /Y/, /υ/

Vowels were presented diotically over headphones to 14 listeners with normal hearing. Listeners identified the vowels using an 11-category response box on the computer screen. In the main experiment they heard 990 vowels, with all conditions randomly interspersed (11 vowels, 6 talkers, 5 spectrum envelope shifts, 3 F_0 shifts).

3. RESULTS

Figure 2 displays the interaction between spectrum envelope and F_0 shifts as a function of talker group. There are several key aspects of the results. First, upward and downward shifts in spectrum envelope resulted in a decline in identification accuracy, consistent with earlier studies. Second, downward shifts produced smaller effects on children's vowels than those of adults, while upward shifts showed smaller effects on vowels spoken by men compared to vowels of women and children. Third, there was a significant *improvement* in performance when an upward shift in spectrum envelope was combined with an upward shift in F0 for the adult males, consistent our previous study (Assmann et al, 2002). There was also a small but significant improvement in several conditions when a downward shift in spectrum envelope was accompanied by a downward shift in F0 for vowels of women and children.

4. DISCUSSION

The experiment replicated earlier studies showing a decline in vowel identification accuracy with upward and downward shifts in spectrum envelope. Adult male vowels were more resistant to upward shifts, while children's voices were more resistant to downward shifts. Children's voices have higher formant frequencies than those of adults, and hence this finding may reflect relatively fixed frequency limits on the shifts in spectrum envelope that preserve intelligibilty.

Shifts in F_0 also produced lower identification accuracy, and children's vowels were particularly susceptible to upward shifts in F_0. Further support for the perceptual learning account comes from the interaction of F_0 and spectrum envelope shifts. For the male vowels, identification improved when an upward shift in F_0 accompanied an upward shift in spectrum envelope. For the women and children, improved accuracy was observed in some conditions when downward shifts in F_0 were combined with downward shifts in spectrum envelope. Taken together, the results are consistent with the idea that learned relationships between F0 and formant pattern are responsible for the effects of frequency shifts on vowel identification.

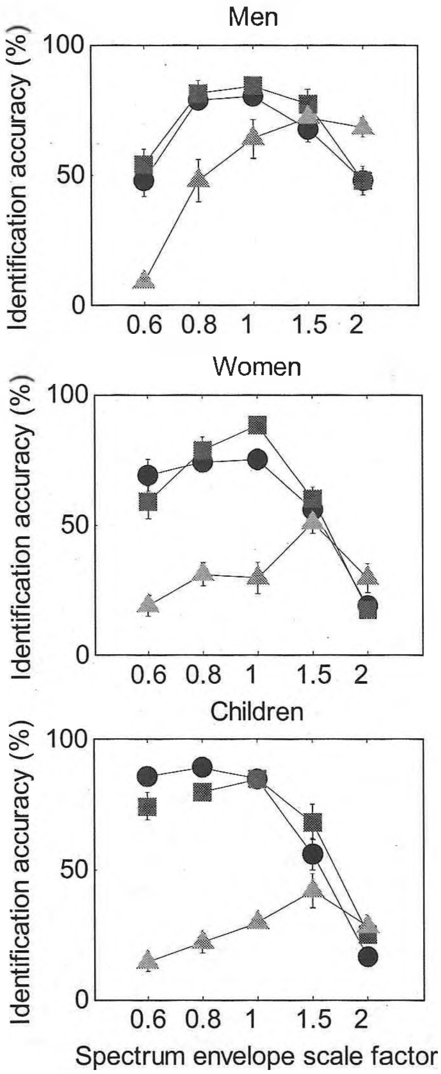

Figure 2: Effects of spectrum envelope and F0 on vowel identification accuracy

REFERENCES

Assmann, P.F. and Katz, W.F. (2000). Time-varying spectral change in the vowels of children and adults *J. Acoust. Soc. Am.* 108: 1856-1866.

Assmann, P.F., Nearey, T.M. and Scott, J.M. (2002). Modeling the perception of frequency-shifted vowels. *Proceedings of the 7th International Conference on Spoken Language Processing*, pp. 425-428.

Fu, Q-J. and Shannon, R.V. (1999). Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing. *J. Acoust. Soc. Am.* 105: 1889-1900.

Hillenbrand, J.M and Nearey, T.M. (1999). Identification of resynthesized /hVd/ utterances: effects of formant contour. *J. Acoust. Soc. Am.* 105: 3509-3523.

Kawahara, H. (1997). Speech representation and transformation using adaptive interpolation of weighted spectrum: Vocoder revisited. *Proceedings of the ICASSP*, pp. 1303-1306.