

Frequency Shifts and Vowel Identification

Peter F. Assmann[†], Terrance M. Nearey[‡]

[†] University of Texas at Dallas, Richardson, TX 75083, USA

[‡] University of Alberta, Edmonton, AB, T6G 2E7, Canada

E-mail: assmann@utdallas.edu, t.nearey@ualberta.ca

ABSTRACT

To study the effects of frequency shifts on vowel identification, a high-quality vocoder (STRAIGHT) was used to process a set of vowels in /hVd/ syllables spoken by two adult males, two adult females, and two children. Vowel identification accuracy was reduced when the spectrum envelope was shifted upward by a factor of 2.0, or downward by a factor of 0.6. Upward shifts produced a smaller decline for the vowels of adult males compared to the vowels of adult females and children, while downward shifts led to a smaller drop for children's vowels compared to those of adults. In several conditions, the drop in accuracy with spectrum envelope shifts was counteracted by introducing corresponding upward or downward shifts in fundamental frequency (F_0). The results suggest that listeners' prior exposure to statistical regularities in natural speech is an important factor in the identification of frequency-shifted vowels.

1. INTRODUCTION

In everyday speech communication, listeners experience a wide range of variation in fundamental frequency (F_0) and formant frequencies. Several 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. Chiba and Kajiyama (1941) experimented with frequency-shifted speech by varying the playback speed of a phonograph. They found that slowing down a male voice or speeding up the voices of women or children resulted in lower intelligibility. Chiba and Kajiyama's technique applies the same frequency shift to the formants and F_0 , and does not simulate the natural covariation of F_0 and formant frequencies across talkers. For example, the change from male to female speech involves an upward shift of 15-20% in spectrum envelope features (formants), accompanied by an 80-100% increase in F_0 (although individual voices may deviate considerably from this pattern).

More recently, Fu and Shannon (1999) used a channel vocoder with noise excitation, and found that intelligibility declined when the spectral shift factor was less than 0.6 or greater than 1.5. Fu and Shannon confirmed that vowels spoken by adults were more resistant to upward shifts in spectrum envelope than the vowels of women and children, while children's vowels were more resistant to downward shifts. Fu and Shannon used noise excitation to synthesize their stimuli, and did not investigate the effects of F_0 .

The intelligibility of frequency-shifted speech may be a perceptual adaptation to statistical covariation between F_0 and spectrum envelope features (e.g. formant frequencies) in natural speech associated with size differences in the larynx and vocal tract across talkers (Nearey, 1989). There is a moderate correlation between F_0 and formant frequencies in natural speech, and a number of studies have shown that vowel quality can be altered by increasing or decreasing F_0 (e.g., Miller, 1953; Johnson, 1990; Nearey, 1989). In a recent study, we investigated the separate and combined effects of upward shifts in F_0 and spectrum envelope on the identification of vowels spoken by adult males in /hVd/ syllables (Assmann, Nearey and Scott, 2002). The results showed a drop in identification accuracy when the spectrum envelope was shifted upward by a factor of 2.0, or when F_0 was raised by a factor of 4.0. However, performance *improved* in several conditions where F_0 and spectrum envelope were both shifted up in frequency. The interaction of F_0 and spectrum envelope shifts were predicted by a pattern recognition model incorporating measurements of formant frequencies, F_0 and duration. One explanation for the synergistic interaction between F_0 and spectrum envelope is that listeners' judgements of vowel identity are influenced by learned associations between F_0 and spectrum envelope in natural speech.

The present study extends the results of Assmann et al. (2002) by investigating both upward and downward shifts in spectrum envelope and F_0 , and by using a wider range of talkers (adult males, adult females, and 7-year old children).

2. EXPERIMENT

The experiment was designed to measure the relative contributions of *spectrum envelope* and *fundamental frequency* to frequency-shifted speech. A high-quality source-filter vocoder, STRAIGHT (Kawahara, 1997) was used to process a set of vowels in /hVd/ words.

2.1. Method

The stimuli were /hVd/ words (*heed, hid, hayed, head, had, hud, hawed, hoed, hood, who'd, herd*) spoken by 2 adult males, 2 adult females, and 2 children aged 7 years. 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 &

Katz, 2000; Katz & Assmann, 2001). Each vowel was resynthesized in 15 conditions (3 levels of F₀ shift factor x 5 levels of spectrum envelope shift factor):

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

F₀ scale factor = 0.5, 1.0, 4.0

Vowel=/i/, /ɪ/, /e/, /ɛ/, /æ/, /ʌ/, /ɜ/, /ɑ/, /o/, /u/, /ʊ/

Vowels were presented to 6 listeners with normal hearing. The stimuli were presented diotically over headphones. Listeners identified the vowels using an 11-category response box on the computer screen. Prior to the experiment, listeners completed practice sessions on a separate vowel set until they reached a mean identification score of 85% correct. In the main experiment they heard 990 vowels, with all conditions randomly interspersed (11 vowels, 6 talkers, 5 spectrum envelope shifts, 3 F₀ shifts).

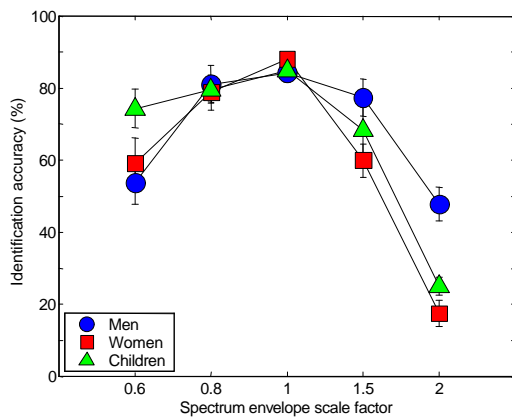


Figure 1: Means and standard errors for 6 listeners as a function of spectrum envelope shifts for vowels of men (circles), women (squares) and children (triangles). The results shown are for conditions where F₀ was not shifted, i.e. the natural variation in F₀ was preserved.

2.2. Results

Figure 1 shows that identification accuracy was reduced when the spectrum envelope was shifted upwards by a factor of 2.0, or downwards by a factor of 0.6. Figure 1 also shows an interaction of spectrum envelope shift and talker group. Consistent with Fu and Shannon (1999), the results showed that the vowels of children were less adversely affected by downward shifts than those of adults, while the vowels of adult males were more resistant to upward shifts. Since the formant frequencies of children are higher than those of adults, this result suggests that there are absolute limits on the extent of spectrum envelope shifts that preserve vowel quality.

Figure 2 shows the interaction between spectrum envelope and F₀ shifts as a function of talker group. The interaction is complex, but several key points can be noted. First, as reported by Assmann et al. (2002), an upward shift in spectrum envelope applied to the vowels of adult males led to reduced identification accuracy, but performance

improved when the upward shift in spectrum envelope was accompanied by an increase in F₀. This improvement was less pronounced in corresponding conditions for the vowels of women and children, where identification dropped to chance level with a spectrum envelope shift factor of 2.0.

A second key finding was an improvement in identification when the vowels of women and children with downward shifts in spectrum envelope (shift factor of 0.6) were synthesized on a lower fundamental (F₀ shift factor of 0.5). The improvement suggests that frequency-shifted vowels may be easier to identify when F₀ and formant pattern preserve their natural relationships. However, the corresponding condition does not lead to an improvement for the adult males.

When F₀ is increased by a factor of 4.0, the amplitude spectra of children's vowels contain few harmonics, and the frequencies of the formants may be difficult to estimate. This condition has been described as "sparse sampling" (Diehl et al., 1996; Assmann & Katz, 2000). It is interesting to see that increasing the spectrum envelope shift factor from 0.6 to 1.5 leads to a progressive improvement from chance level (9%) to about 40-50% correct for the vowels spoken by women and children.

3. PATTERN RECOGNITION MODEL

In a previous study (Assmann, et al., 2002) we showed that identification accuracy by listeners could be predicted fairly accurately using the pattern recognition model described by Hillenbrand and Nearey (1999). The input to the model is a "dual-target" representation of the vowel (Nearey and Assmann, 1986) with 8 parameters: mean F₀, formant frequencies F1, F2, and F3 sampled at the 20% and 80% points in the vowel, and vowel duration. The frequency measures are expressed in log units. 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). Linear discriminant function analysis was used to generate *a posteriori* probabilities of group membership for each test vowel.

A key assumption of the model is that listeners have internalized knowledge of the relationship between F₀ and formant frequencies in natural speech, i.e. that higher formant frequencies are accompanied by higher F₀ and vice versa. We are currently investigating the predictions of the model for the data reported here. Preliminary findings suggest that some aspects of the interaction between F₀ and spectrum envelope shifts are predicted by the model. However, there are also discrepancies between the model and the data, indicating that additional factors need to be considered, such as masking and frequency selectivity. Frequency shifts alter the spacing of spectral features (harmonics and formants) and therefore frequency resolution may play a role.

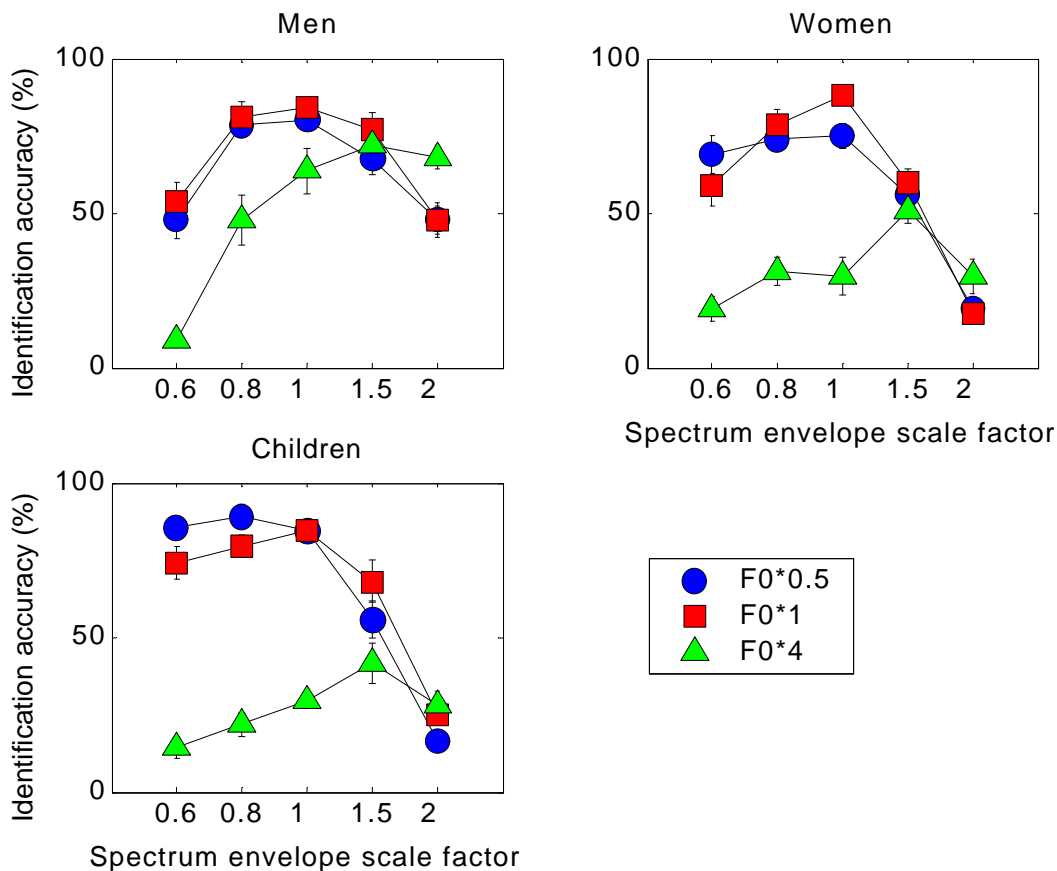


Figure 2: Interaction of spectrum envelope and F_0 shifts on vowel identification accuracy.

4. CONCLUSIONS

The results of the experiment replicated earlier studies showing a drop in vowel identification accuracy when the spectrum envelope is shifted down by a factor of 0.6 or up by a factor of 2.0. 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 intelligibility. One possibility is that these perceptual tolerance limits are rooted in perceptual experience.

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.

REFERENCES

- [1] Assmann, P.F. & Katz, W.F. (2000). Time-varying spectral change in the vowels of children and adults. *J. Acoust. Soc. Am.* 108: 1856-1866.
- [2] 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.
- [3] Chiba, T. & Kajiyama, M. (1941). The vowel: its nature and structure. Tokyo-Kaiseikan.
- [4] Diehl, R.L., Lindblom, B., Hoemeke, K.A., and Fahey, R.P. (1996) On explaining certain male-female differences in the phonetic realization of vowel categories. *Journal of Phonetics* 24: 187-208.
- [5] Fu, Q-J. & Shannon, R.V. (1999). Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing. *J. Acoust. Soc. Am.* 105: 1889-1900.

- [6] Hillenbrand J.M., Nearey T.M. (1999). Identification of resynthesized /hVd/ utterances: effects of formant contour. *J. Acoust. Soc. Am.* 105: 3509-3523.
- [7] Kawahara, H. (1997). Speech representation and transformation using adaptive interpolation of weighted spectrum: Vocoder revisited. *Proceedings of the ICASSP*, pp. 1303-1306.
- [8] Katz, W.F. & Assmann, P.F. (2001). Identification of children's and adults' vowels: Intrinsic fundamental frequency, fundamental frequency dynamics, and presence of voicing. *J. Phonetics* 29: 23-51.
- [9] Miller, R.L. (1953). Auditory tests with synthetic vowels. *J. Acoust. Soc. Am.* 18: 114-121.
- [10] Nearey, T.M. (1989). Static, dynamic, and relational properties in vowel perception. *J. Acoust. Soc. Am.* 85: 2088-2113.
- [11] Nearey, T.M. & Assmann, P.F. (1986). "Modeling the role of inherent spectral change in vowel identification," *J. Acoust. Soc. Am.* 80: 1297-1308.