

Effects of frequency shifts on vowel category judgments

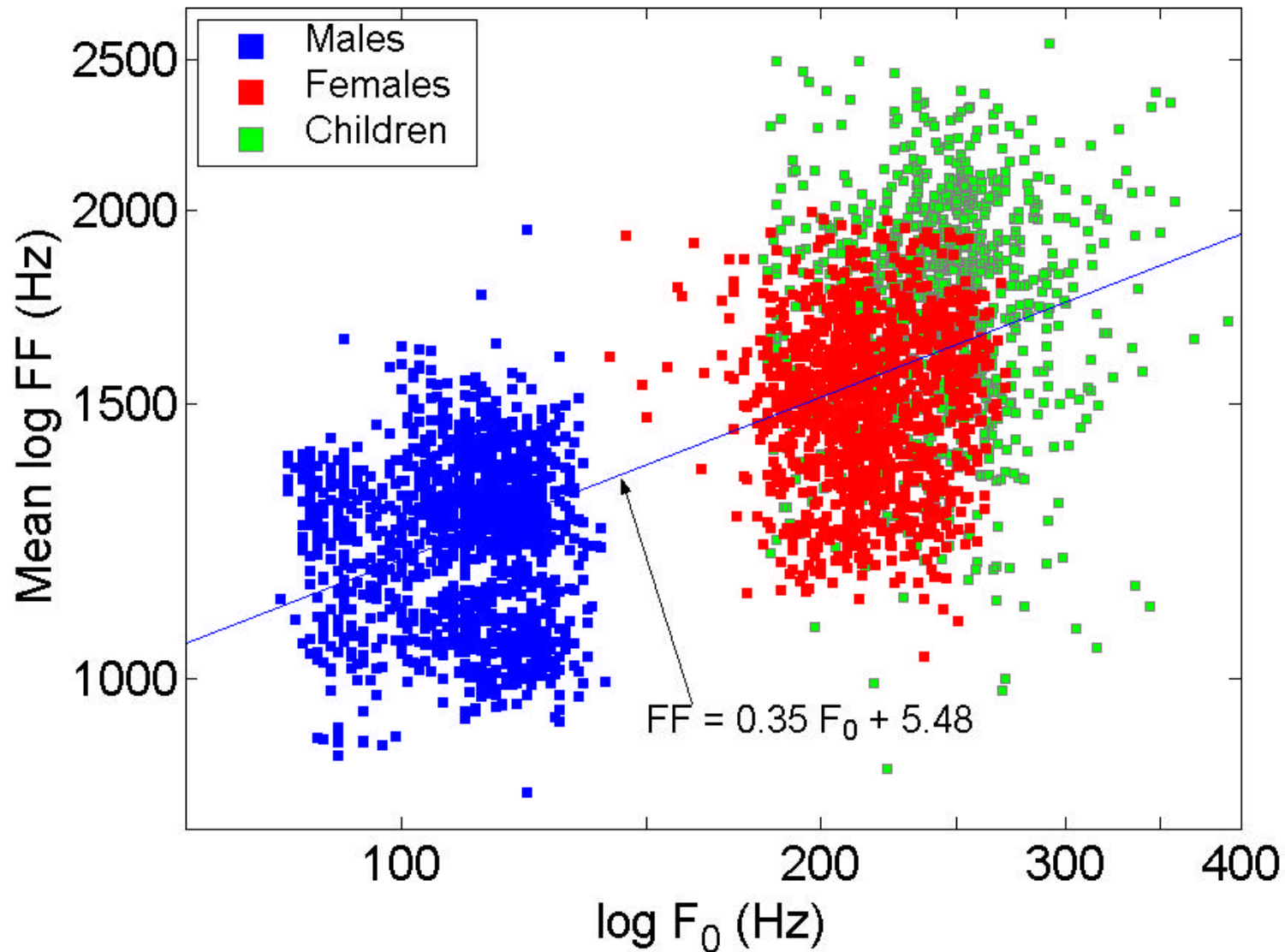
Catherine Glidden, Peter F. Assmann (School of Human Development, Univ. of Texas at Dallas, Box 830688, Richardson TX 75083)

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 a vowel continuum
 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)

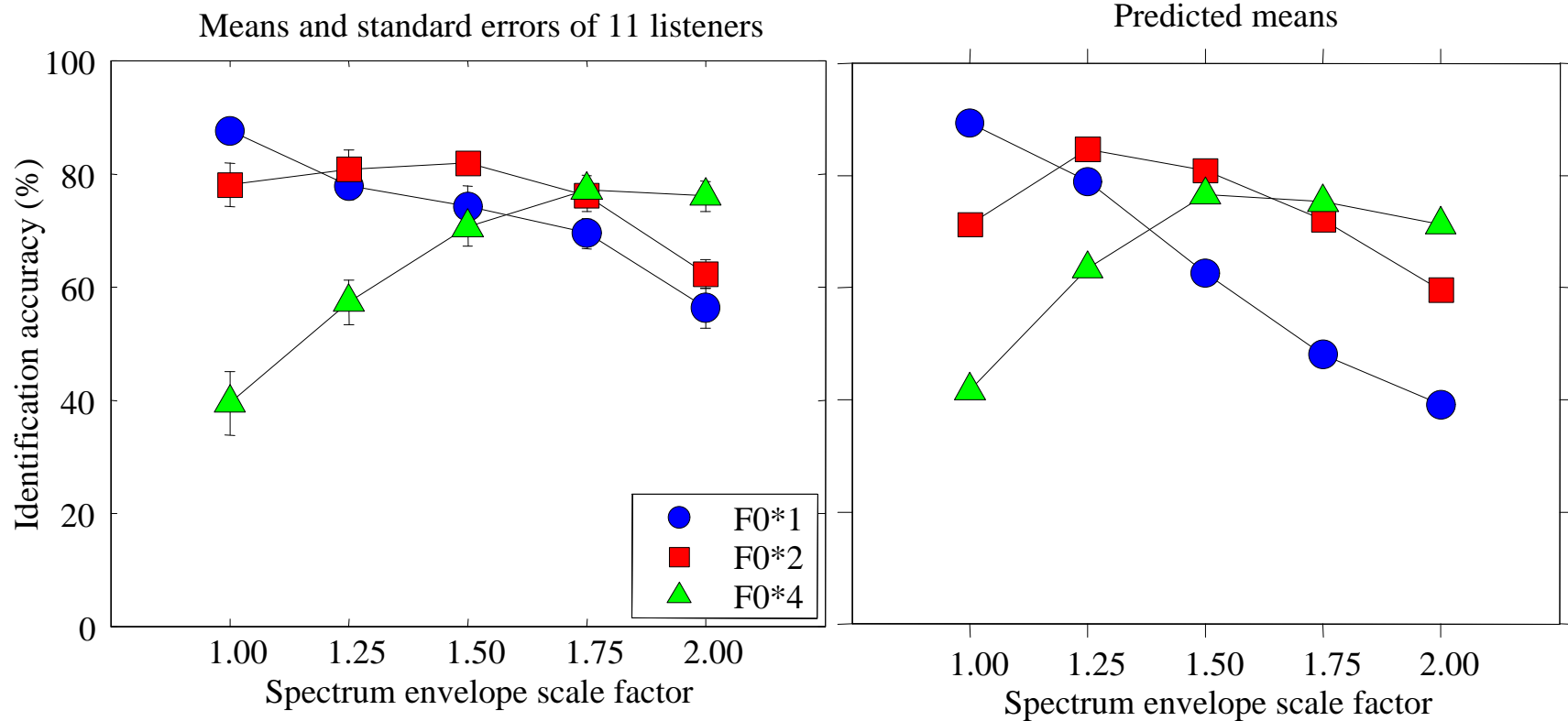
Pattern recognition model

- Hillenbrand & Nearey (1999) dual-target 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 F_0 or formant frequencies (FF) resulted in lower vowel identification accuracy.
- However, *combining* upward shifts in F_0 with upward shifts in FF led to *improved* identification accuracy.
- This finding that vowel identification accuracy is higher with coordinated shifts in F_0 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 F_0 and FF in natural speech.

Vowel Identification Accuracy



(Assmann, Nearey, and Scott, ICSLP 2002).

Vowel Identification Experiment

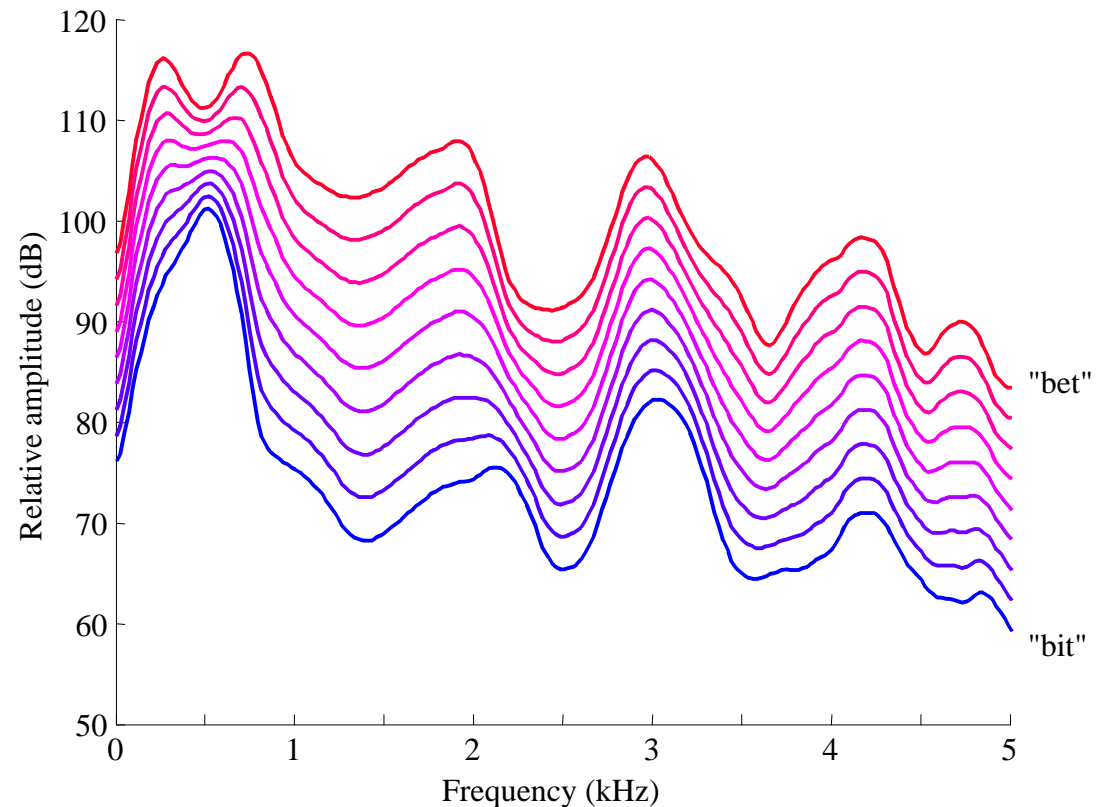
- The present study examined the effects of frequency shifts on vowel category boundaries along a continuum from /bɪt/ to /bɛt/
- Natural recordings of /bɪt/ and /bɛt/ from an adult female talker were analyzed and resynthesized using the STRAIGHT vocoder, and upward and downward frequency shifts were introduced

STRAIGHT vocoder

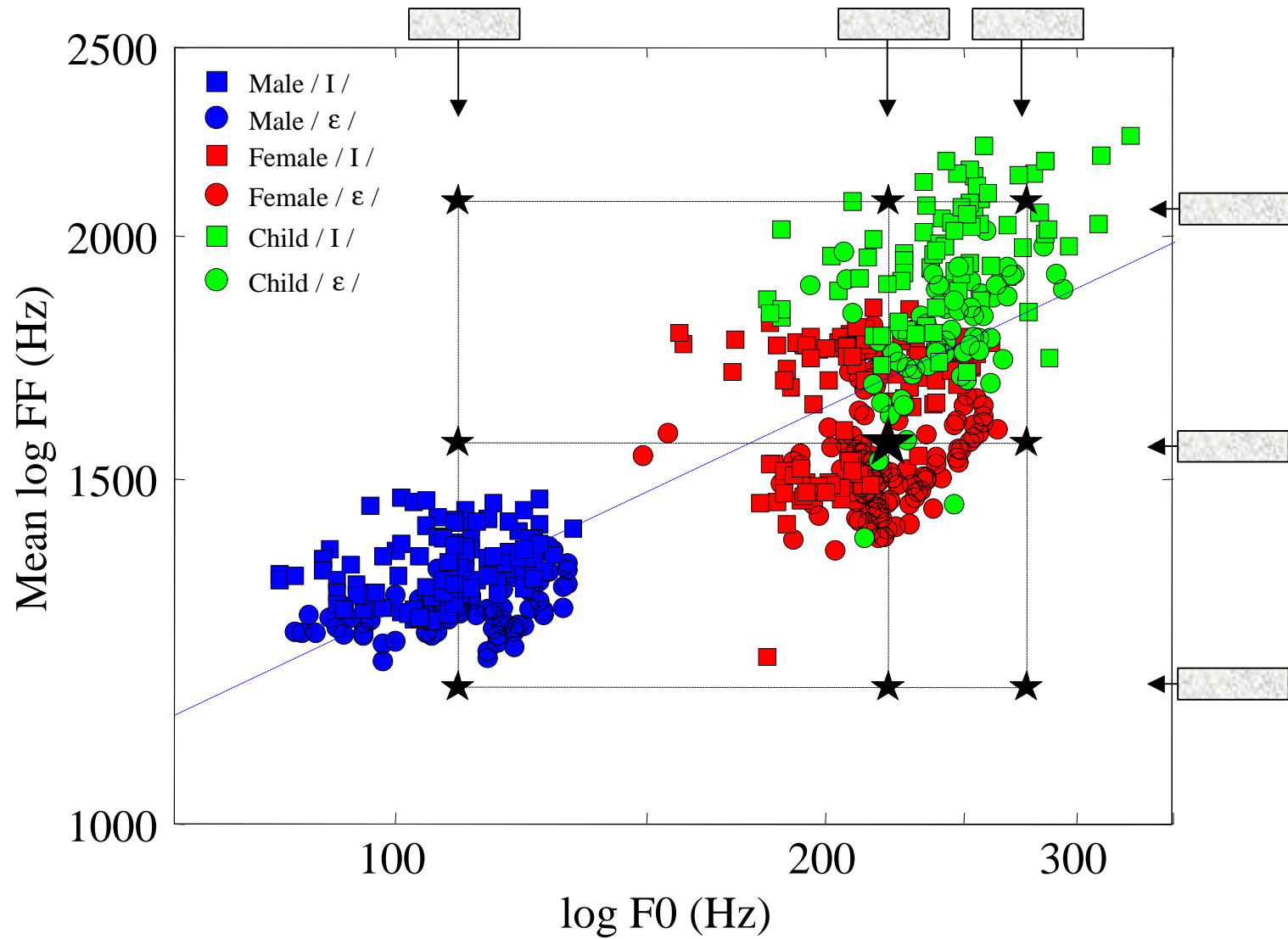
- High-quality vocoder, STRAIGHT (Kawahara, 1997) used to generate 9-step continuum from /bɪt/ to /bɛt/.
 - High-resolution analysis of time-varying spectrum envelope
 - Wavelet-based instantaneous frequency F_0 extraction
 - Spectrum envelope (**FF**) scaling
 - Fundamental frequency (**F₀**) scaling

Synthesis of Stimuli

- The 9-step continuum from “bit” to “bet” was generated by interpolation of the time-varying spectrum envelope on the dB scale.
- Here we show a sampled spectrum envelope for each step on the baseline continuum, taken from the vowel midpoint (average of 20 frames; 1-ms frame rate). The “bet” endpoint is shown in red; “bit” is shown in blue.



Scale Factors



Scale Factors

F ₀ scale factors		
0.5	1.0	1.25

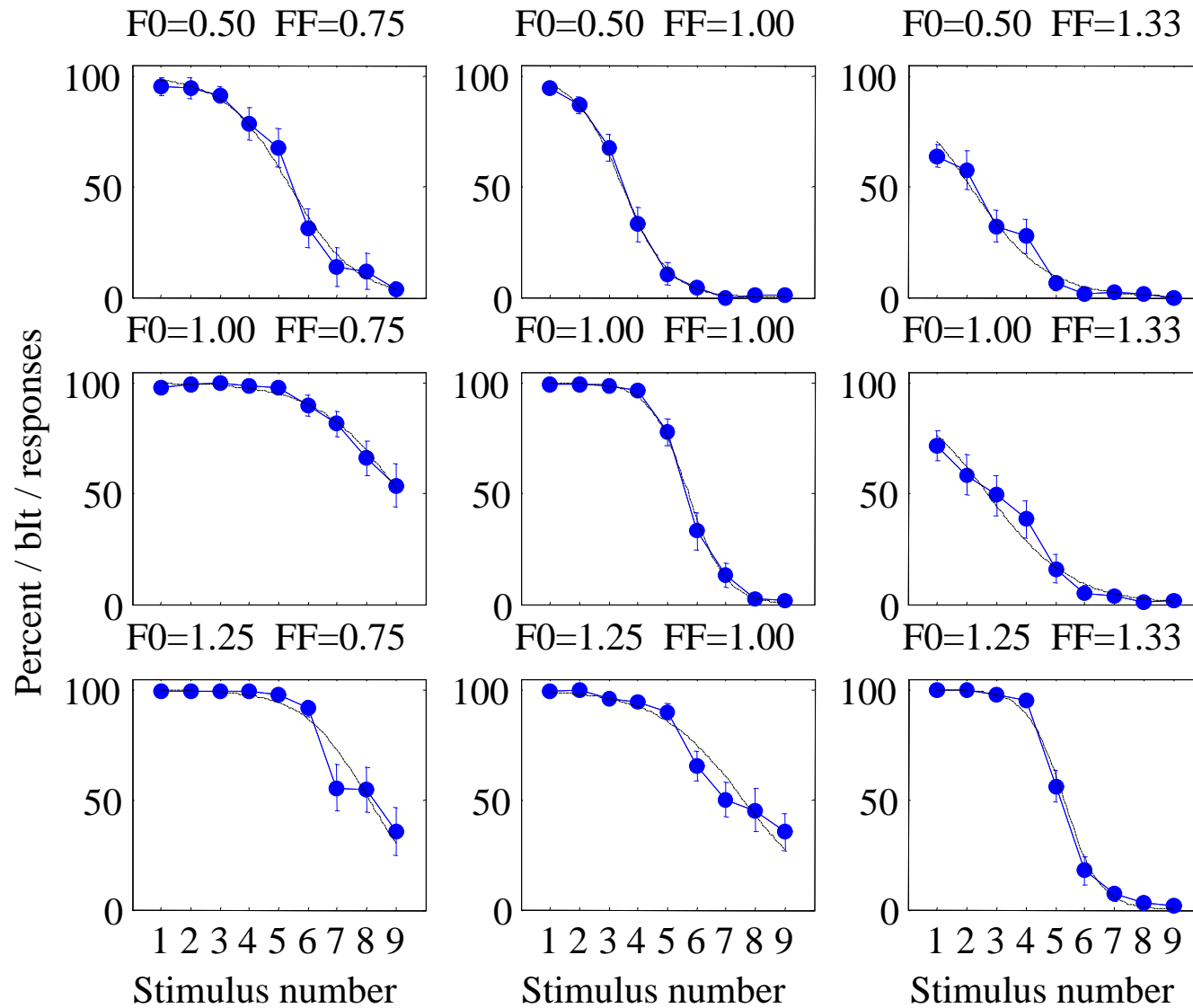
FF scale factors		
0.75	1.0	1.33

- Baseline voice (scale factors: F₀=1.0, FF=1.0) is adult female
- Downward shifts tend to produce male-like voices; upward shifts heard as child-like voices

Method

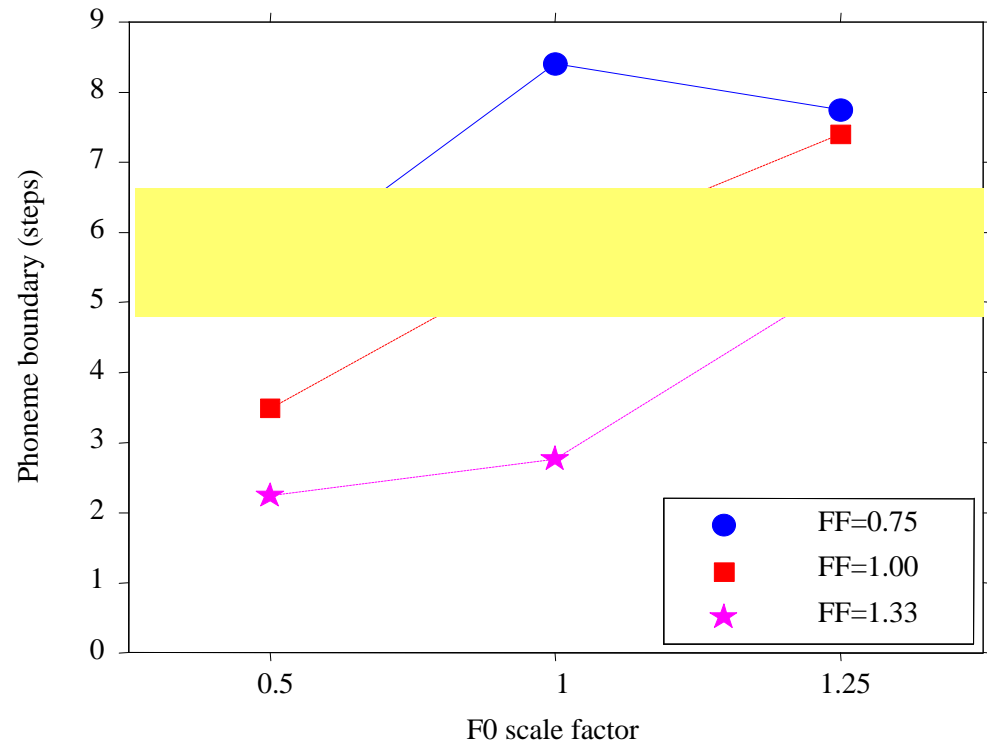
- 810 syllables (9-step continuum x 3 F_0 scale factors x 3 FF scale factors x 10 trials) presented diotically in double-walled sound booth
- 13 listeners identified the words “bit” or “bet” using a forced-choice two-button response box drawn on computer screen
- Stimuli blocked by voice type (F_0 x FF condition) and trials, but randomized within blocks

Observed Identification Functions



Observed Mean Boundaries

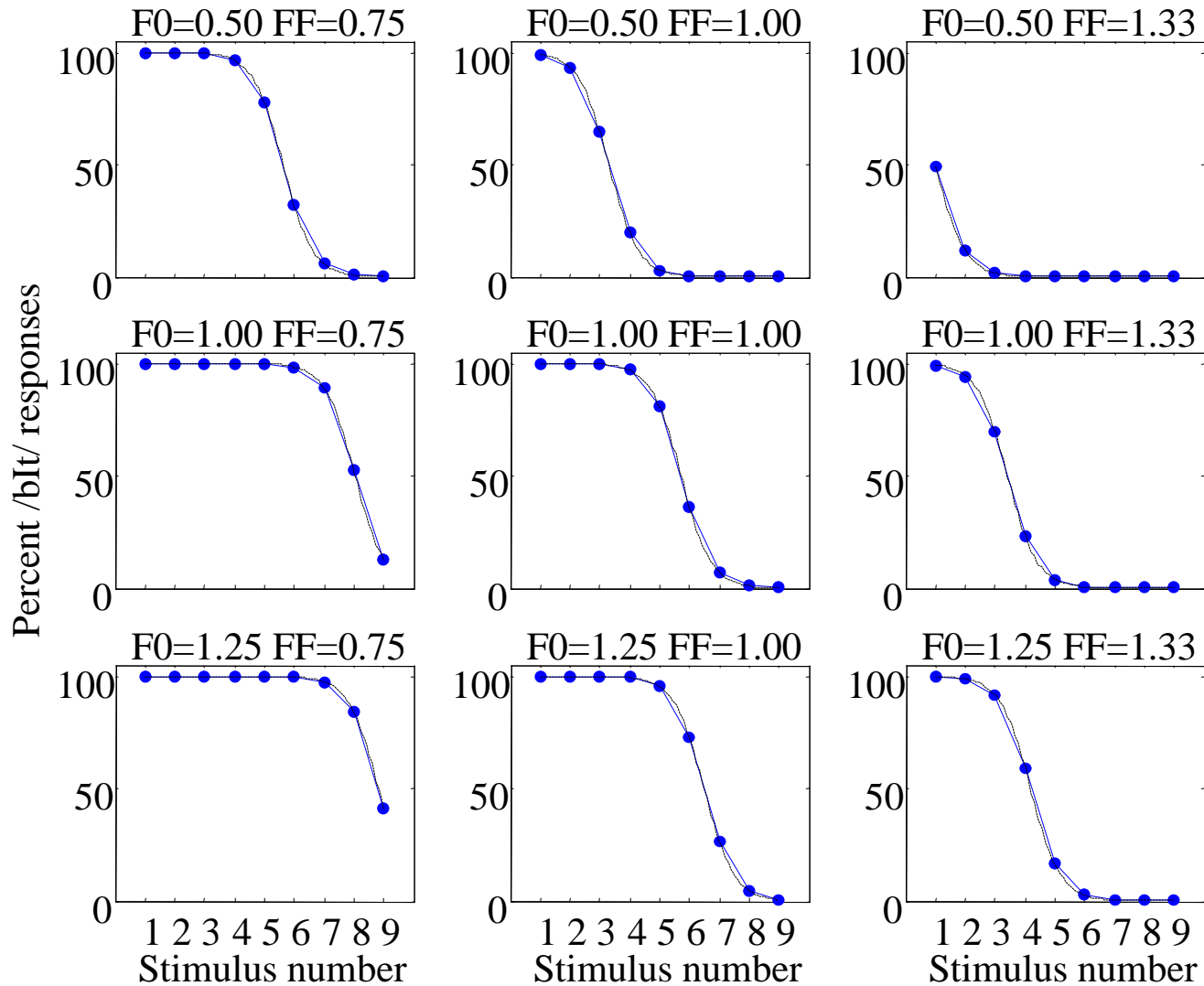
- Increasing the F_0 scale factor caused boundaries to shift towards the /bɪt/ endpoint
- Increasing the FF scale factor caused boundaries to shift towards the /bɛt/ endpoint
- Coordinated shifts (highlighted in yellow) show relatively little change



Model Implementation

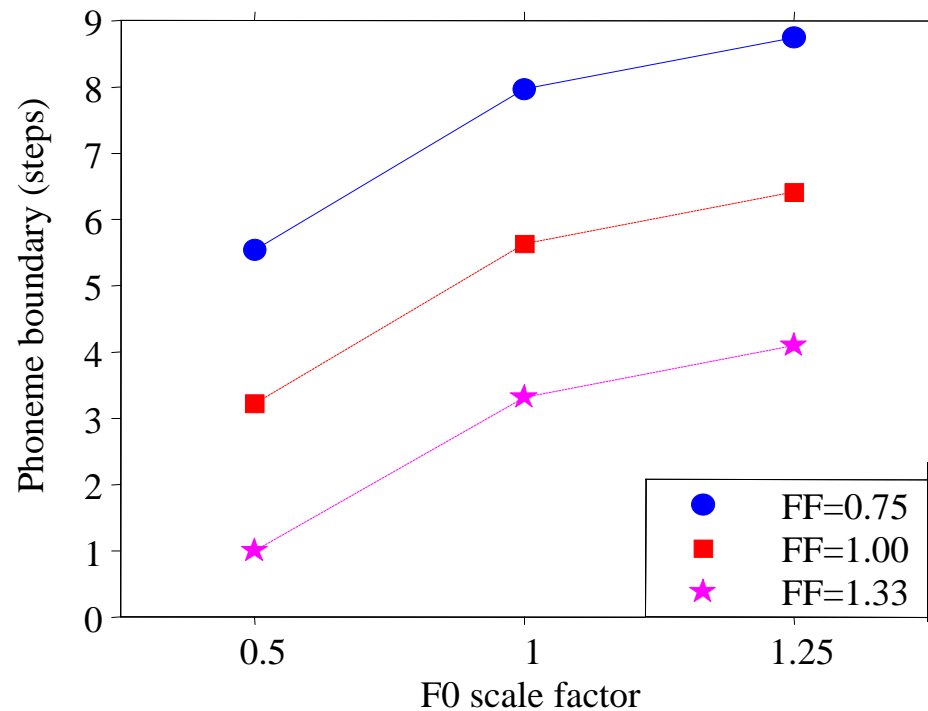
- Formant frequencies and F_0 were measured from each of the 9 stimuli along the baseline continuum and were scaled up or down according to the condition
- The model generated *a posteriori* probabilities of /ɪ/ or /ɛ/ responses for each of the stimuli

Predicted Identification Functions

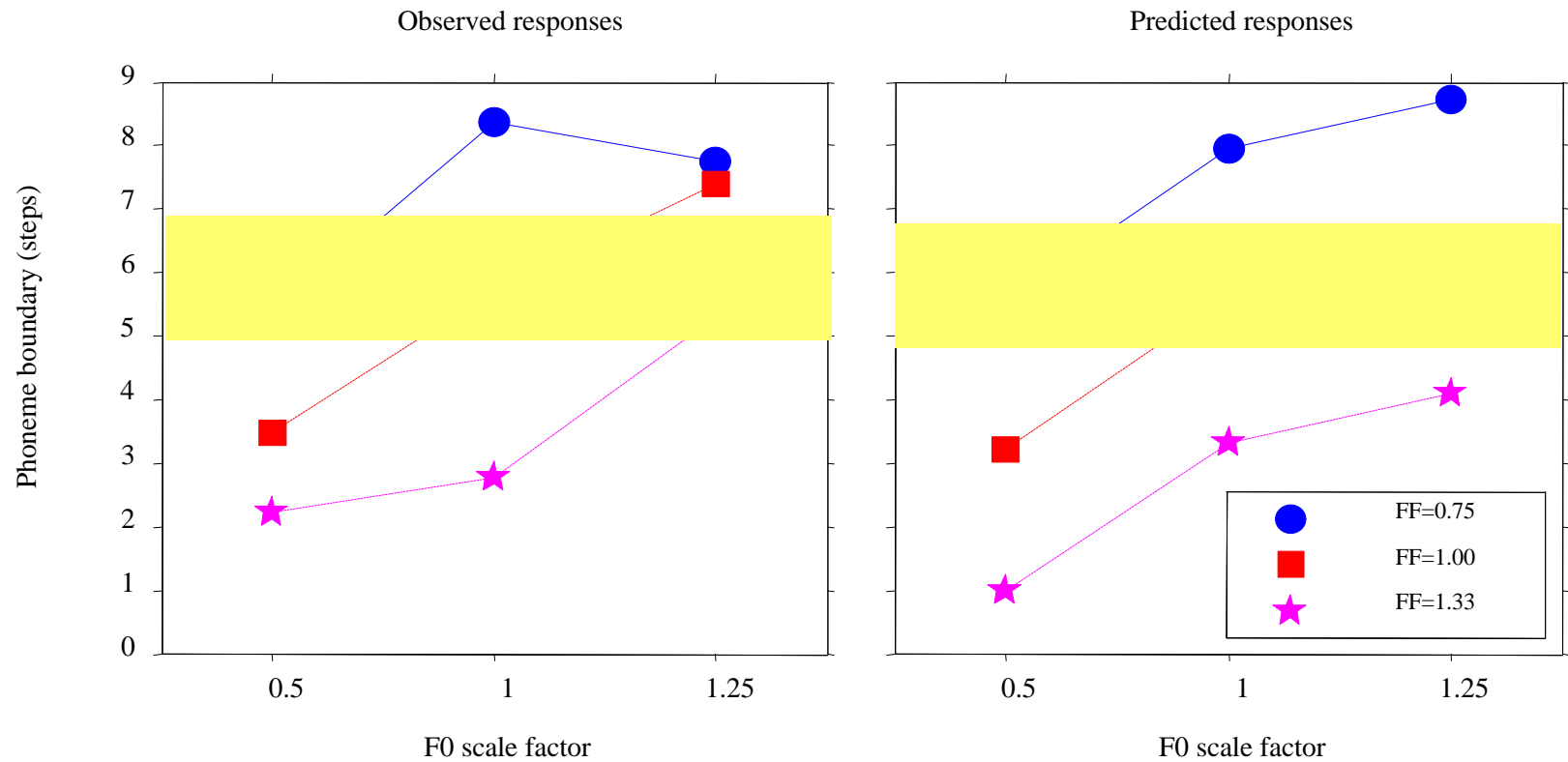


Predicted Mean Boundaries

- Increasing F_0 scale factor predicts boundaries shift towards /bɪt/ endpoint
- Increasing FF scale factor predicts boundaries shift towards /bɛt/ endpoint
- Vowel boundaries shift in predictable but opposite directions as a function of F_0 and FF



Observed and Predicted Boundaries



Conclusions

- Upward and downward shifts in F_0 and FF cause vowel category boundaries to shift in opposing directions, as predicted by the model.
- Coordinated shifts (raised F_0 and raised FF, or *vice versa*, highlighted in yellow) have little effect.

Conclusions

- Discrepancy between the model and the data at the extreme edges (F_0 : 1.25, FF: 0.75 and F_0 : 0.5, FF: 1.33)
- Possible reasons:
 - Ambiguous quality created by an apparent change in vocal tract size is inconsistent with the shift in pitch
 - Range effects (blocking may lead listeners to avoid all “bit” or all “bet” responses in a block)
 - Boundary estimates are less reliable when the identification functions lack clear crossovers

Follow-up experiment

- Smaller scale factor to avoid extreme edges
- Embed syllables in carrier sentence (and apply shifts to both carrier and test syllable) to eliminate trial-to-trial context effects
- Randomize order of conditions to eliminate possible range effects

References

1. Assmann PF, Katz WF. (2000) Time-varying spectral change in the vowels of children and adults. *J Acoust Soc Am.* 108(4): 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. 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.
4. Hillenbrand JM, Nearey TM. (1999) Identification of resynthesized /hVd/ utterances: effects of formant contour. *J Acoust Soc Am.* 105(6): 3509-3523.
5. Kawahara, H. (1997) Speech representation and transformation using adaptive interpolation of weighted spectrum: vocoder revisited. *Proc. IEEE Int. Conf. on Acoustics, Speech & Signal Processing (ICASSP '97)*, vol.2, pp.1303-1306.