

1pSC28

Intelligibility of frequency-shifted speech

(Jack M. Scott, Peter F. Assmann, University of Texas at Dallas, and
Terrance M. Nearey, University of Alberta)

Introduction	Experiment 1: vowels	Results: Mean identification	Confusion matrix	Experiment 2: sentences	Conclusions
Hypotheses	Formant scale factors	Summary of results	Statistical pattern recognition model	Stimulus construction and synthesis	Summary
Synthesis Method	Experiment 1: Stimuli	Individual Vowel errors	Predictions of mean accuracy	Results: Mean identification	References

Poster presentation at the 141st Meeting of the Acoustical Society of America, Chicago, June 4, 2001

Introduction

A significant fact about speech perception is that intelligibility is preserved when the formant pattern is shifted up or down along the frequency scale, and when the fundamental frequency (F_0) is raised or lowered. There is a moderate relationship between F_0 and formant frequencies in natural speech (Nearey, 1989). To study the *combined* effects of upward shifts in formant frequency and F_0 on intelligibility, we used a high-quality vocoder (Kawahara, 1997) to process a sample of vowels in /hVd/ words and sentences from the HINT (Hearing in Noise Test).

Hypotheses

- The sparse spectral sampling hypothesis (Ryalls and Lieberman, 1982; Diehl et al., 1995) predicts a drop in intelligibility as F_0 is raised because the shape of the spectrum is less accurately defined.
- Pattern recognition models of vowel identification (e.g. Miller, 1989; Hillenbrand and Nearey, 1999) suggest that F_0 makes an independent contribution to vowel identity. These models also predict reduced intelligibility, but attribute the decline to disruption of learned relationships between F_0 and formants.

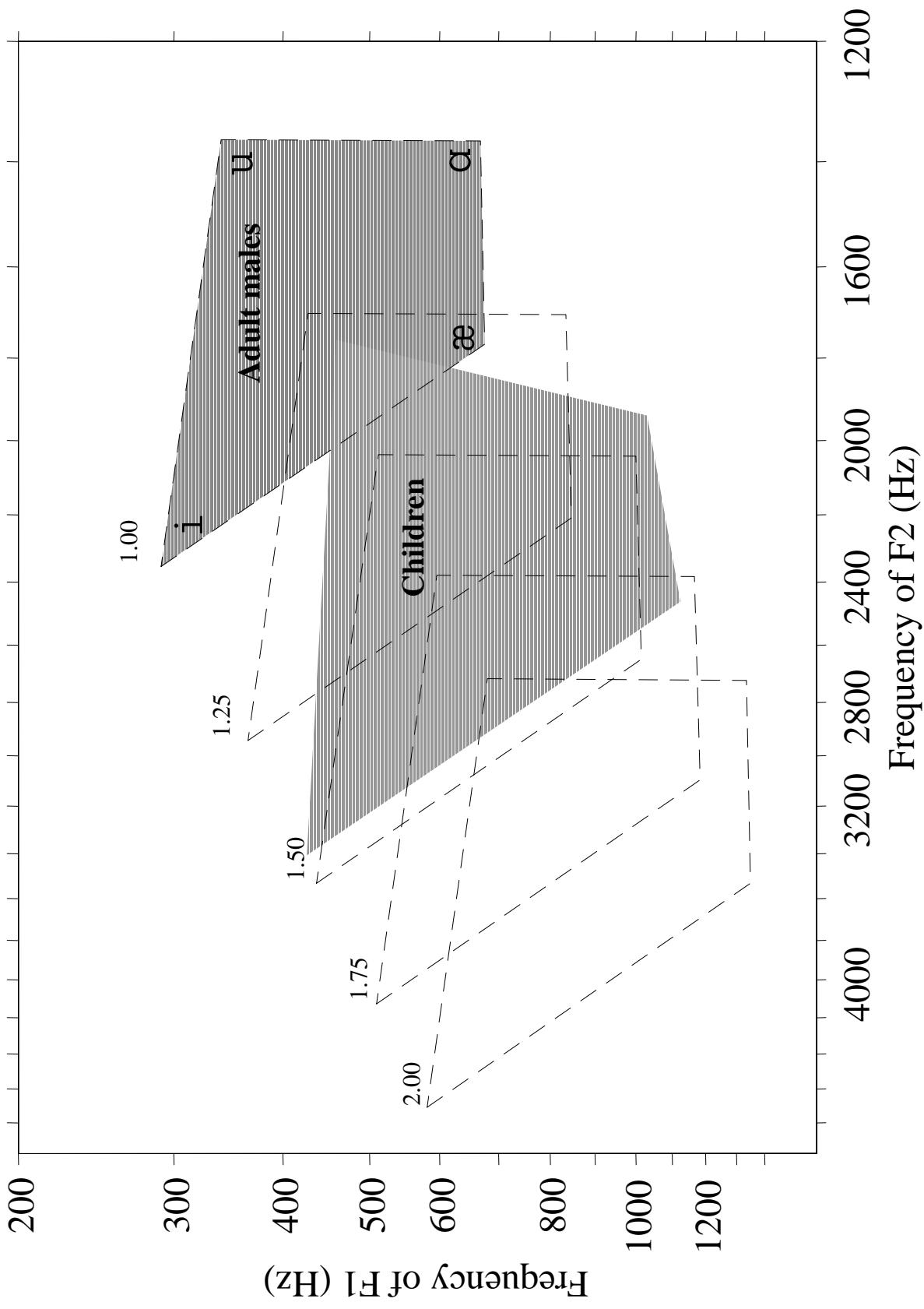
Synthesis method

- Two experiments used the STRAIGHT analysis-synthesis program (Kawahara, 1997) to manipulate formant frequencies and F_0 .
- STRAIGHT performs an accurate decomposition of speech into source and filter components. The program allows the user to shift the formant frequencies by applying a linear scale factor to the spectrum envelope. A linear scale factor can be applied independently to manipulate the F_0 .

Experiment 1: Stimuli

- **Original recordings:** 11 vowels in /hVd/ words spoken by 3 adult males from a larger sample of recordings of adults and children (Assmann & Katz, 2000).
- **Synthesis conditions:** From these original signals, synthesized vowels were constructed: 5 conditions of spectral envelope shift (1.0, 1.25, 1.5, 1.75, 2.0) combined with 3 conditions of F_0 shift (1.0, 2.0, 4.0). The largest shifts produced child-like vowels; intermediate shifts were heard as female voices.

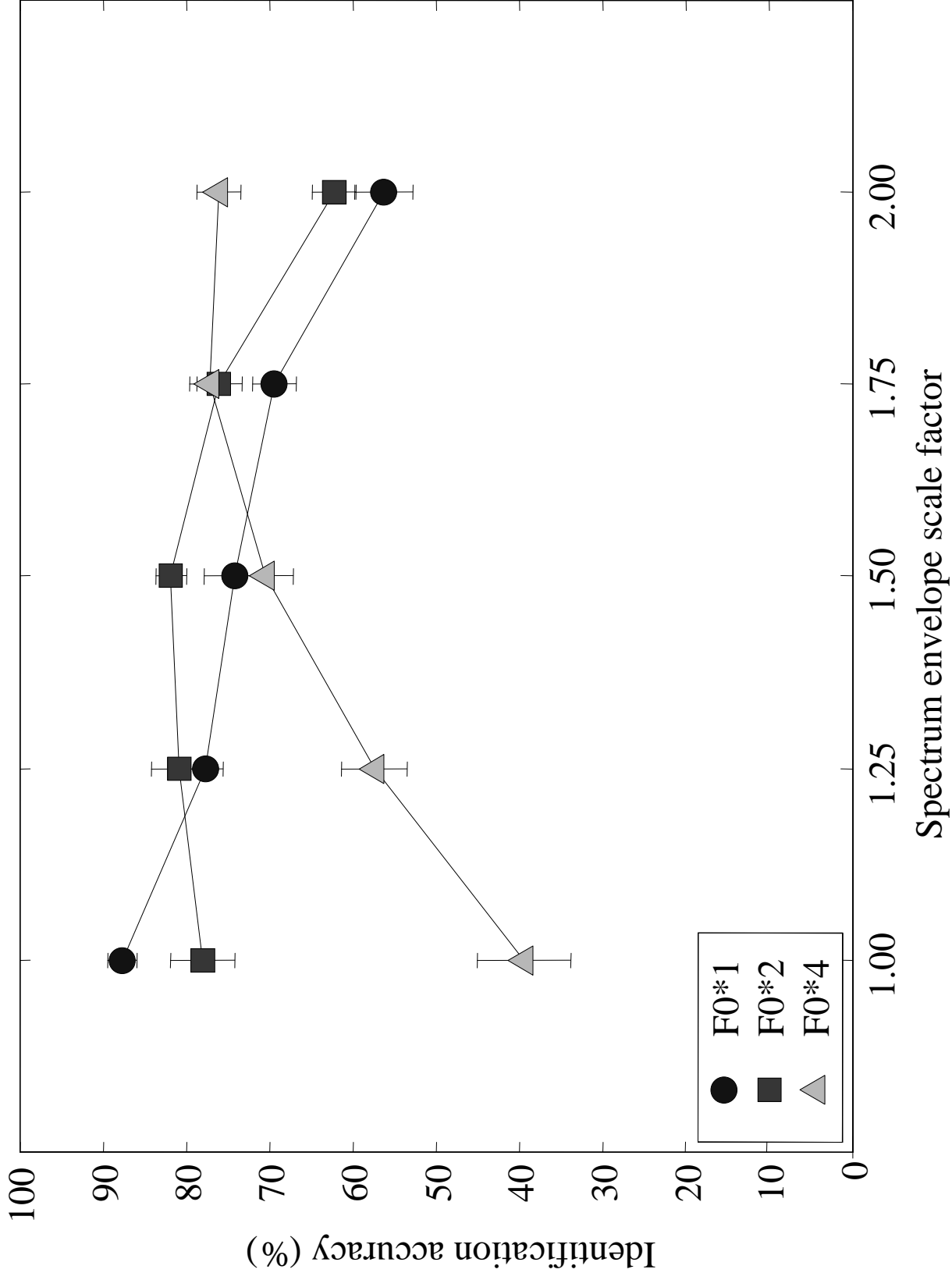
Formant scale factors



Experiment 1: Method

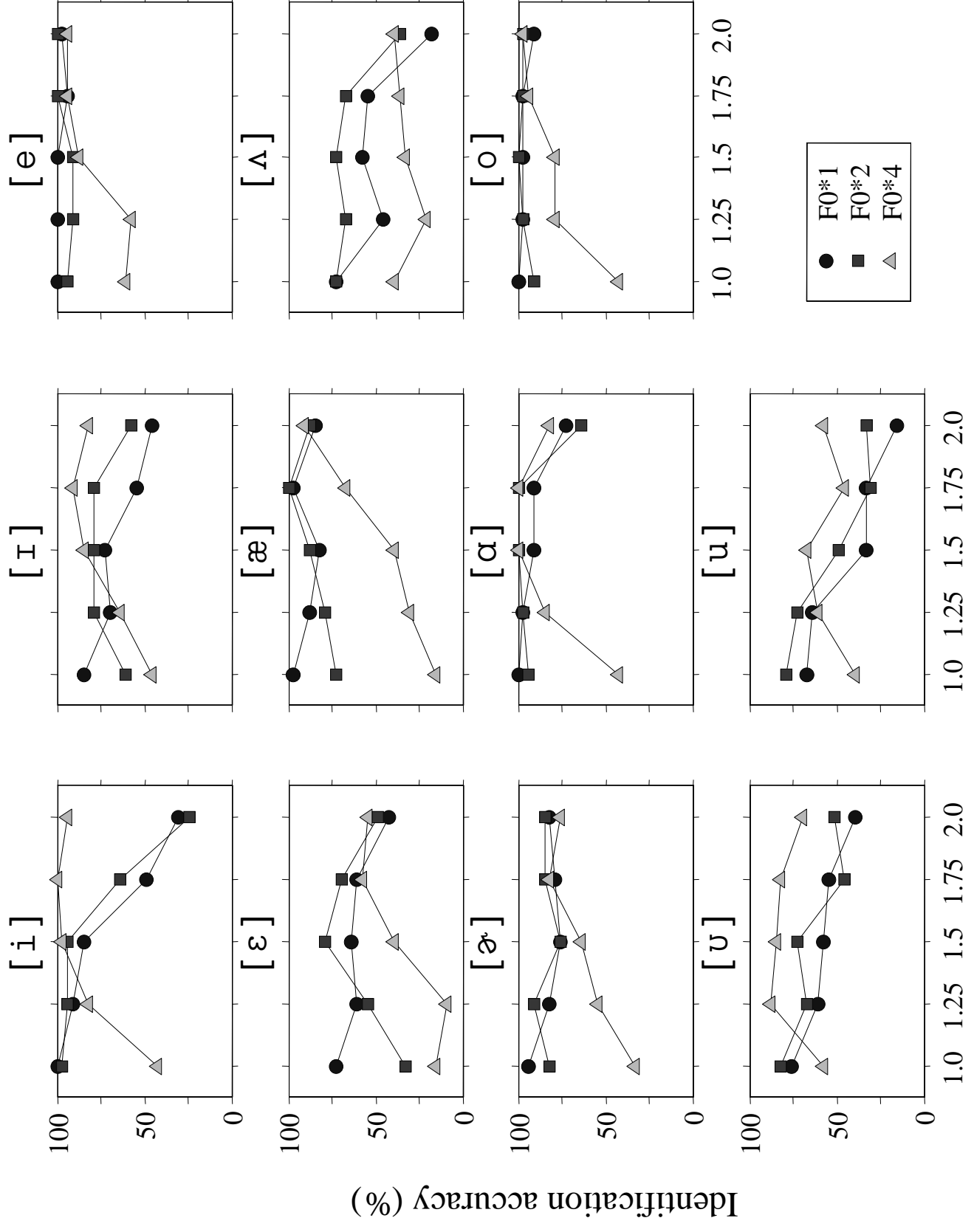
- 11 young adult listeners with normal hearing.
- All were native speakers of American English from the North Texas region.
- 495 stimuli (11 vowels x 3 talkers x 5 spectrum envelope shifts x 3 F_0 shifts) were presented monaurally using headphones; conditions were randomized.
- Listeners first completed practice sessions and were required to score 85% or better on the 11-vowel task before starting the main experiment.

Experiment 1: Results



Summary of results

- The decline in performance with increasing F_0 is predicted by the sparse spectral sampling hypothesis, but the *improvement* in identification of high- F_0 vowels with raised formant frequencies is hard to reconcile with the sparse sampling account.
- Instead, the interaction suggests a synergistic relationship between F_0 and formant pattern shifts: identification accuracy is higher when increases in formant frequency are accompanied by increases in F_0 (and *vice versa*).



Spectral envelope shift factor

FFx1**F0x4**

	i	I	e	ε	æ	Λ	ø	ɑ	O	U	u
i	42	21	18	3	0	0	0	3	0	6	6
I	3	45	3	12	0	0	0	0	0	12	24
e	6	9	61	3	0	3	0	3	0	6	9
ε	0	15	3	15	3	3	0	0	3	27	30
æ	0	3	3	36	15	9	0	27	0	6	0
Λ	0	0	6	0	6	39	6	0	18	15	9
ø	0	0	0	0	0	21	33	0	27	15	3
ɑ	0	0	0	0	0	0	0	42	12	15	30
O	0	0	3	0	0	0	0	9	42	9	36
U	0	0	0	3	3	3	3	0	6	58	24
u	6	9	0	6	3	3	0	3	9	21	39

58

103

97

79

30

82

42

88

118

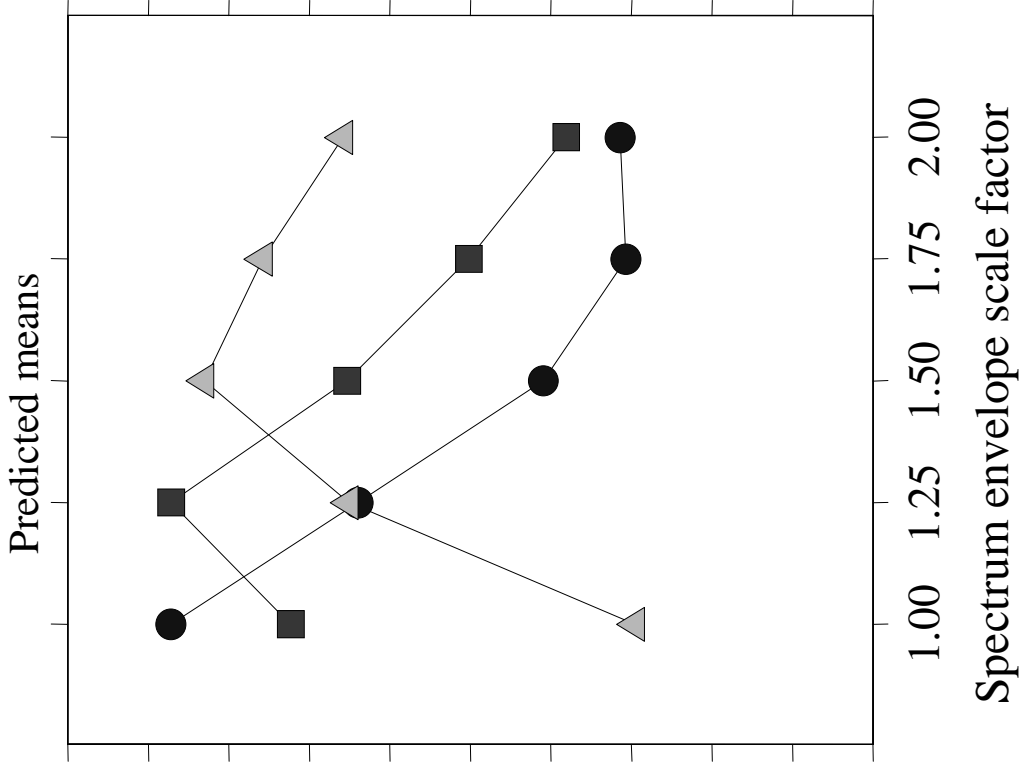
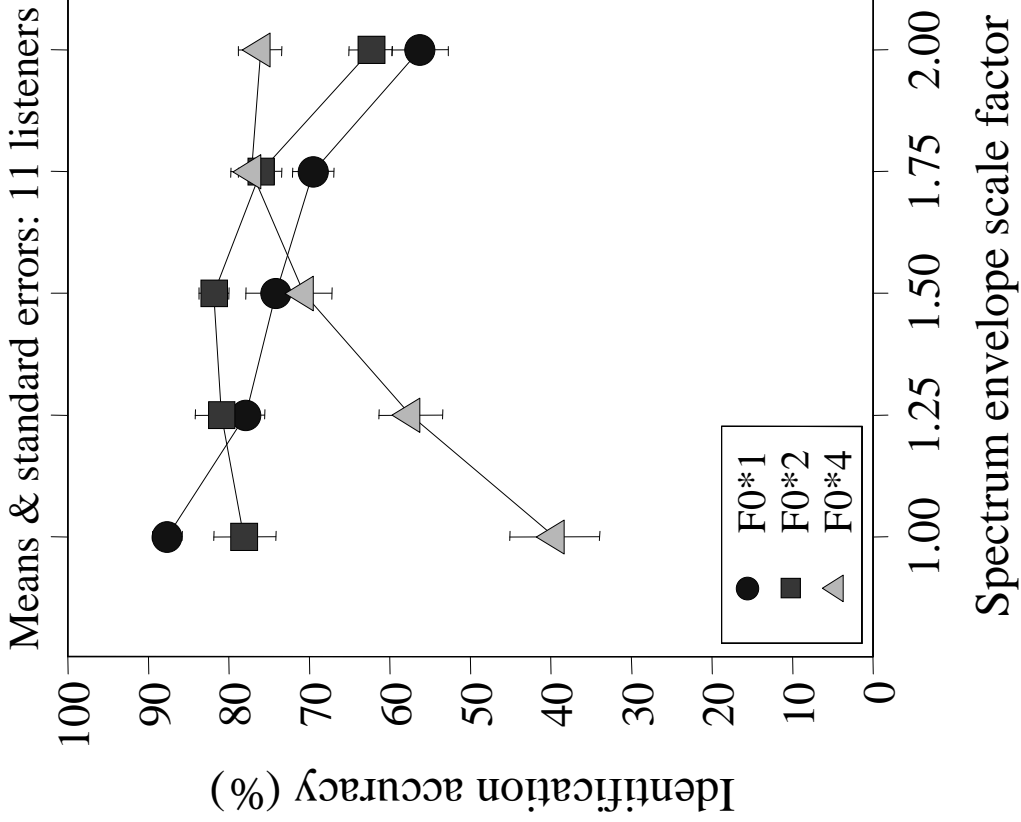
191

212

Pattern recognition model

- Hillenbrand & Nearey (1999) dual-target model
- Parameters: duration, F_0 , and F_1 , F_2 , F_3 sampled at 20% and 80% points
- Training data: 1500+ vowels spoken by men, women and children from the W. Michigan vowel database collected by Hillenbrand et al. (1995)
- *A posteriori* probabilities derived from linear discriminant analysis for each stimulus vowel

Predicted identification



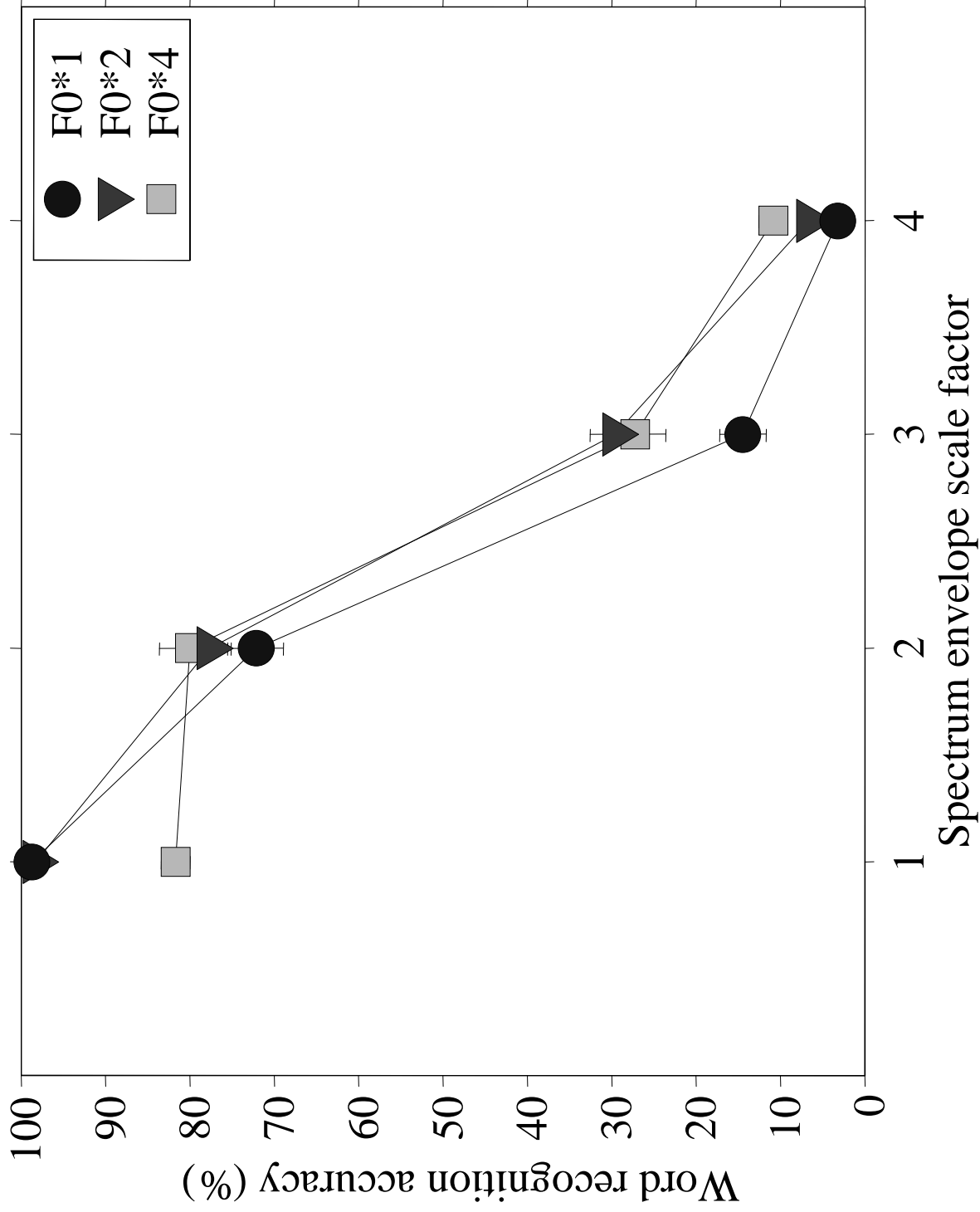
Experiment 2: Stimuli

- **Original recordings:** 280 sentences from the HINT (Nilsson et al., 1994) were selected.
- **Synthesis conditions:** synthesized versions were constructed: 4 conditions of spectral envelope shift (1.0, 2.0, 3.0, 4.0) combined with 3 conditions of F_0 shift (1.0, 2.0, 4.0).
- Larger spectrum envelope shifts were included because we expected that sentences would be more resilient to frequency shifts than vowels.

Experiment 2: Method

- 15 young adult listeners with normal hearing
- All native speakers of American English from the north Texas region
- Listeners first completed a practice session with feedback using 10 unshifted sentences (different from those used in the main experiment).
- In the main experiment each listener heard a random subset of 120 sentences (10 sentences x 4 **spectrum envelope shifts** x 3 **F₀ shifts**).
- Stimuli presented monaurally using headphones; all conditions were randomized.

Experiment 2: Results



Summary of results

- Upward shifts in the formant frequencies have similar effects in sentences and vowels: performance drops by about 25% when formants are raised by a factor of 2. Performance continues to decline to near zero with a scale factor of 4.
- Increasing F_0 by a factor of 4 leads to a 17% drop in recognition accuracy.
- Performance drops less steeply when upward shifts in formant frequency are *combined* with increases in F_0 , supporting the interaction found for vowels.

Conclusions

- For both vowels and sentences, upward shifts in formant frequency lead to lower intelligibility. However, the decline is reduced when combined with an increase in F_0 .
- Similarly, increasing F_0 leads to a drop in performance, but this effect is reduced when the formants are raised.
- F_0 has a smaller effect on sentences than vowels.
- Pattern recognition models can simulate the effects of upward shifts in F_0 and formant frequency in vowels, suggesting that learned relationships between F_0 and spectral envelope cues are responsible for the interaction.

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. Diehl RL, Lindblom B, Hoemeke KA, Fahey RP. (1996). On explaining certain male-female differences in the phonetic realization of vowel categories. *J Phonetics* 24(2): 187-208.
3. Hillenbrand JM, Nearey TM. (1999). Identification of resynthesized /hVd/ utterances: effects of formant contour. *J Acoust Soc Am.* 105(6): 3509-3523.
4. 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.
5. Nilsson M, Soli SD, Sullivan JA. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J Acoust Soc Am.* 95(2): 1085-1099.
6. Nearey TM. (1989) Static, dynamic, and relational properties in vowel perception. *J Acoust Soc Am.* 85(5): 2088-2113.
7. Ryalls JH, Lieberman P. (1982) Fundamental frequency and vowel perception. *J Acoust Soc Am.* 72(5): 1631-1634.