NORMALIZED FORMANT VALUES ARE GENERALIZED ACROSS TALKERS

CHELSEA SANKER CHELSEA.SANKER@YALE.EDU YALE UNIVERSITY LabPhon 18

Introduction

- Listeners can accurately recognize phonological category membership despite variation across individuals (Mullennix et al 1989; Verbrugge et al 1976)
- Vowel interpretation is informed by relative formant values and other cues from the talker's voice (Fujisaki & Kawashima 1968)
- Are vowels transformed in perception to normalize across talkers, or are the raw acoustic details retained (cf. Johnson 1997)?

This Study: A test of how perceptual vowel category boundaries with a novel talker are impacted by earlier exposure to talkers with different natural vowel spaces and manipulated F2

Evidence only for effects of relative formants, not raw values

Characteristics of Stimuli

Table 2: Formants (Hz) by talker, /i/

	F1	F2 (lower'd,rais'd)	F3
fem short	477	2176, 2820	3129
fem long	337	2176, 2820	3012
male short	271	1775, 2304	2837
male long	249	1775, 2304	2818

Table 3: Formants (Hz) by talker, /u/

	F1	F2 (lower'd,rais'd)	F3
fem short	446	1270, 1674	2559
fem long	419	1270, 1674	2646
male short	336	1110, 1478	2430
male long	260	1110, 1478	2303

Table 4: Formants (Hz) by talker, /æ/

	F1	F2 (lower'd,rais'd)	F3
fem short	927	1478,1927	2516
fem long	889	1478,1927	2515
male short	754	1325, 1735	2288
male long	751	1325, 1735	2502

Table 5: Formants (Hz) by talker, /ʌ/

	F1	F2 (lower'd,rais'd)	F3
fem short	759	1330, 1735	2556
fem long	752	1330, 1735	2452
male short	714	1135, 1505	2400
male long	633	1135, 1505	2388

Methodology

Participants: 384 native English speakers (192 m, 192 f), online through Prolific

Task: Exposure phase then testing phase

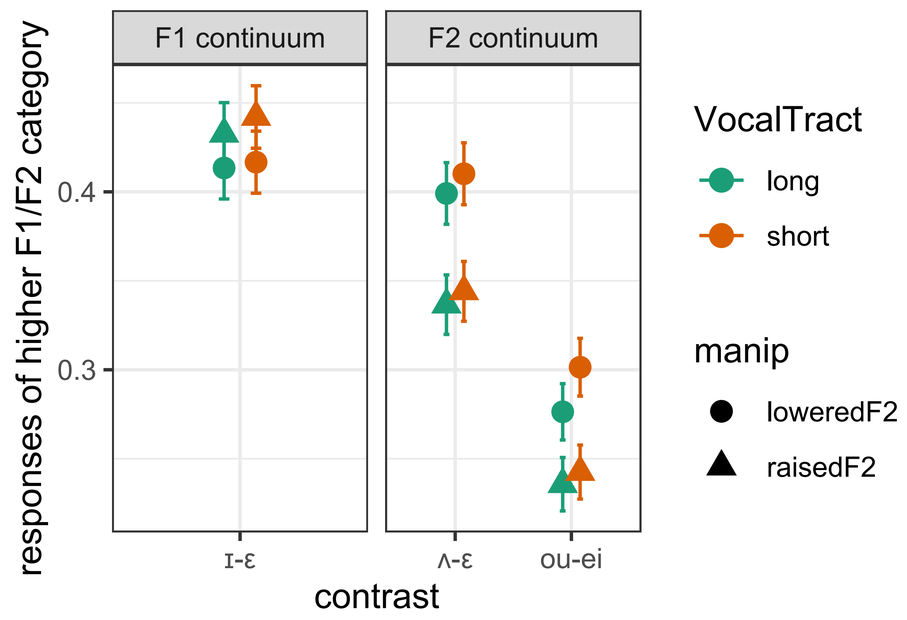
- Exposure:
  - Decisions between consonants in words with the same vowel (e.g. *suit*, *soup*); stimuli included /i u æ ʌ/.
  - Across participants, 8 conditions: 2 talker sexes (male, female) * 2 talker vocal tract lengths (talker's natural vowel space; long, short) * 2 F2 manipulations (raised, lowered)
- Testing:
  - Decisions between vowels, for vowels manipulated along a /ɪ-ɛ/ F1 continuum and /ʌ-ɛ/ and /ou-ei/ F2 continua
  - All participants heard the same testing stimuli, produced by a novel talker

Main Results

	β	SE	z-value	p-value
(Intercept)	-4.0	0.46	-8.7	< 0.001
FormantStep	1.7	0.019	88.8	< 0.001
BaseVowel LowerFormant	-2.2	0.035	-63.6	< 0.001
Contrast /ɪ-ɛ/	0.19	0.65	0.29	0.77
Contrast /ou-ei/	-0.99	0.65	-1.5	0.12
Contrast /ʌ-ɛ/ : Manip Raised	**-0.55**	**0.069**	**-8.0**	$\mathbf{< 0.001}$
Contrast /ɪ-ɛ/ : Manip Raised	**0.18**	**0.067**	**2.6**	**0.0081**
Contrast /ou-ei/ : Manip Raised	**-0.46**	**0.071**	**-6.5**	$\mathbf{< 0.001}$
Contrast /ʌ-ɛ/ : VocalTract Short	0.081	0.069	1.2	0.24
Contrast /ɪ-ɛ/ : VocalTract Short	0.052	0.067	0.77	0.44
Contrast /ou-ei/ : VocalTract Short	**0.15**	**0.07**	**2.1**	**0.032**
Contrast /ʌ-ɛ/ : TalkerSex Male	0.062	0.069	0.9	0.37
Contrast /ɪ-ɛ/ : TalkerSex Male	-0.078	0.067	-1.2	0.24
Contrast /ou-ei/ : TalkerSex Male	0.095	0.07	1.4	0.18

Table 1: Logistic model for responses of the vowel category with the higher formant (F2 for /ʌ-ɛ/, /ou-ei/; F1 for /ɪ-ɛ/). *Intercept: Manip = lowered formant, TrainingTalkerVocalTract = Long, TrainingTalkerSex = Female, BaseVowel = higher formant, Contrast = /ʌ-ɛ/.*

Figure 1: Proportion of responses of the vowel category with higher formant along each continuum.



Exposure to raised F2 in training resulted in fewer responses of /ɛ/ and /ei/, in their respective F2 continua

Exposure to raised F2 in training resulted in more /ɛ/ responses for /ɪ-ɛ/ decisions, in the F1 continuum

A shorter vocal tract (larger vowel space) of the training talker resulted in more /ei/ responses for /ou-ei/ decisions, i.e. a lowered F2 boundary

Conclusions

- Exposure to raised F2 increased listeners' F2 boundaries, but made the naturally produced F1 values in training stimuli lower relative to the vowel space suggested by F2, which lowered F1 boundaries
- F2 values in training were interpreted relative to the vowel space suggested by other formants; the same F2 is relatively lower for a shorter vocal tract than a longer one, so lowered F2 boundaries after exposure to a talker with a shorter vocal tract
- Formants heard during training are extended to a novel talker – how they extend indicates that vowel representations are based on relative formant values, which are directly mapped across talkers
- Implications for convergence and perception studies: raw formants are unlikely to be informative predictors

Selected References

Fujisaki, H., & Kawashima, T. (1968). The roles of pitch and higher formants in the perception of vowels. *IEEE Transactions on Audio and Electroacoustics*, 16(1) 73-77.
Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In Johnson & Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 145-165). Academic Press.
Mullennix, J., Pisoni, D., & Martin, C. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America* 85, 365-378.
Verbrugge, R., Strange, W., Shankweiler, D., & Edman, T. (1976). What information enables a listener to map a talker's vowel space? *Journal of the Acoustical Society of America* 60, 198-212.