

Asymmetries in patterns of confusion of American English vowels

Robert Felty^a and José Benki^b

1 Introduction

- From a speech-in-noise task we observe certain vowels to be highly confusable.
- We also observe an asymmetry in the perception of /a æ/, which manifests as a bias towards /æ/

Hypothesis:

Perceptual bias is due to a difference in acoustic properties between listener's expectations and the presented stimuli.

2 Method

- 125 Listeners identified American English phonemes in noise at signal-to-noise ratios (S/N) ranging from 0 to -12 dB.
- Stimuli consisted of three different types, CV and VC syllables with all phonotactically possible English consonants and vowels [a i u] by California speakers (Shannon et al., 1999), and hVd stimuli by Michigan speakers (Hillenbrand et al., 1995). Each listener heard two tokens of each phoneme, spoken by different talkers.
- This presentation will focus on the vowel results.

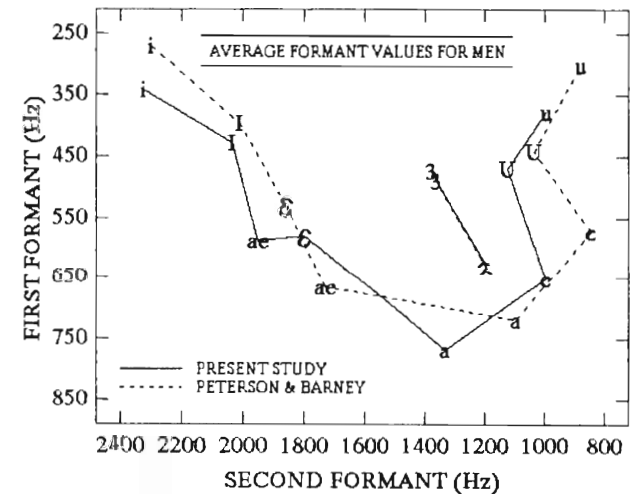
3 Analysis

- Data were first analyzed as separate confusion matrices for each S/N. The original confusion matrices were transformed into 2x2 submatrices.

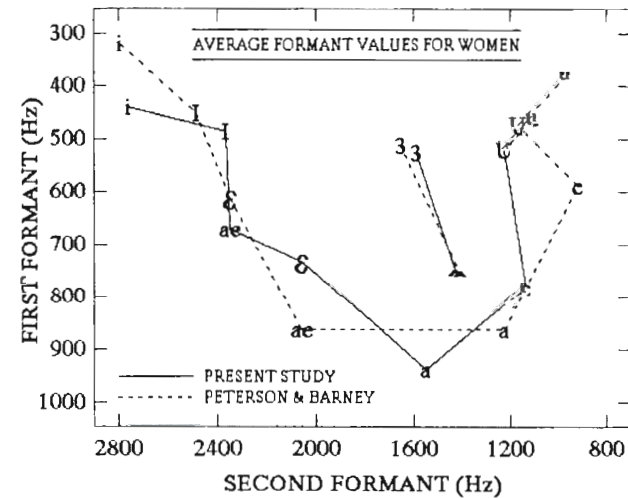
Research supported by the NIH/NIDCD and the University of Michigan Undergraduate Research Opportunity Program.

^aUniversity of Michigan Department of Linguistics
robfelty@umich.edu

^bUniversity of Michigan Kresge Hearing Research Institute
benki@umich.edu



(a) men



(b) women

Figure 1: Hillenbrand et al. (1995)

- A signal detection theory (SDT) analysis was then applied to each

submatrix. These are shown in figure 3

- SDT measures the sensitivity of distinguishing two stimuli, using the metric, d' . We apply SDT to analyze phonemes.
- SDT also provides a measure of bias, c' , which indicates whether one is more or less likely to respond with a particular phoneme. Positive values of c' indicate a bias towards a response; negative values indicate a bias against a response.

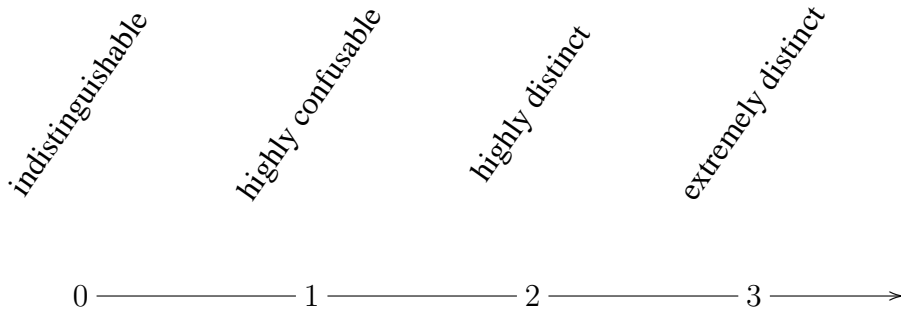


Figure 2: d' values

4 Results

The results of the SDT analysis are summarized in table 1 and shown in figure 3 (page 4).

5 Explanations for confusability and bias

5.1 Type of Noise

- It is possible that the noise could introduce a bias between contrasts, by either masking a cue, or by being interpreted as a cue for a given phoneme.
- Chang et al. (2001) have found evidence for this in the case of velar stops being perceived as palatal or alveolar, but not the reverse.
- For the confusable pair $/a \text{ } \text{æ}/$, we find a bias against $/a/$ even in the clear condition, which indicates that noise cannot be the cause of the bias. See figure 3(c).

Table 1: Sensitivity and bias results

contrast	sensitivity	bias
$/a \text{ } \text{ɔ}/$	$d' < 1$ highly confusable even without noise	$-.8 < c' < .8$ no significant bias
$/a \text{ } \text{æ}/$	$1 < d' < 3$ somewhat confusable esp. with noise	$c' < .02$ except $\leq -9\text{dB}$ bias towards $/\text{æ}/$
$/ɔ \text{ } \text{æ}/$	$.9 < d' < 4.5$ highly distinct except at -12 dB	$c' < 0$ except -12dB no significant bias
$/\text{æ} \text{ } \text{ɛ}/$	$.5 < d' < 3$ somewhat confusable esp. with noise	$-.2 < c' < 0.2$ no significant bias
$/a \text{ } \text{ʌ}/$	$2.2 < d' < 3.7$ highly distinct even with noise	$-.8 < c' < 0.8$ no significant bias
$/ɔ \text{ } \text{ʌ}/$	$1.8 < d' < 3.7$ highly distinct even with noise	$-.8 < c' < 0.8$ no significant bias
$/\text{ɛ} \text{ } \text{ʌ}/$	$.8 < d' < 4.5$ highly distinct except below -9dB	$-.8 < c' < 0.8$ no significant bias

5.2 Acoustics

- Given that $/a/$ and $/ɔ/$ are highly confusable, and that $/a/$ and $/\text{æ}/$ are also fairly confusable, one would expect that $/ɔ/$ and $/\text{æ}/$ would also be confusable, but this is not the case.
- This can be explained by looking at F2. From table 2, it can be seen that the primary acoustic distinction between $/a/$ and $/ɔ/$ is the difference in F2. F2 also seems to be the main distinction between $/a/$ and $/\text{æ}/$.

- Thus it seems that the difference in F2 between /ɑ/ and /æ/ is not great enough to enable accurate identification, but the difference in F2 between /ɔ/ and /æ/ is sufficient.

Table 2: Formant frequency and duration differences for confusable phoneme contrasts. Values are mean differences for men and women as measured by Hillenbrand et al. (1995). Checkmarks indicate the primary acoustic distinction for each contrast.

contrast	Δ F1	Δ F2	Δ duration
/ɑ ɔ/	135Hz	370Hz	15ms
/ɑ æ/	210Hz	700Hz (✓)	10ms
/ɔ æ/	90Hz	1075Hz ✓	16ms
/æ ε/	34Hz	220Hz	85ms ✓
/ɑ ʌ/	170Hz	126Hz	91ms ✓
/ɔ ʌ/	29Hz	246Hz	115ms ✓
/ε ʌ/	30Hz	595Hz ✓	17ms

5.3 Listener expectations

- The bias for /æ/ in the /a æ/ pair cannot be solely explained by acoustics.
- Rather, the bias seems to be due to listener expectations.
- Figure 1 shows the vowel space of their recordings compared with those of Peterson and Barney (1952) (P&B). If the listeners are expecting to hear vowels similar to P&B, but instead are hearing the vowels recorded by Hillenbrand et al. (1995), the bias can be explained.
- The /ɑ/ tokens that our listeners heard have formant frequencies that overlap with those recorded by P&B, which explains why /ɑ/ is frequently heard as /æ/.
- However, the /æ/ our listeners heard is acoustically only similar to P&B’s /æ/, not P&B’s /ɑ/, which explains why /æ/ is rarely heard as /ɑ/

- Both vowels have shifted forward in the F1 F2 space as a result of NCS. This movement makes [ɑ] tokens ambiguous for ”standard (i.e., P&B)” perceptual categories /ɑ/ and /æ/, but the [æ] tokens are not as ambiguous.
- In other words, our listeners are not treating the speakers as speakers of a Northern Cities Shift (see e.g. Gordon, 1997) variety.

6 Conclusions

- The asymmetries in perception in our data can be best explained by the interaction of the acoustics of the presented stimuli and the listener’s expectations.

References

- Chang, Steve, Madelaine Plauché, and John Ohala. 2001. Markedness and consonant confusion asymmetries, in Elizabeth Hume and Keith Johnson, (Eds.) *The role of speech perception in phonology*, San Diego: Academic Press.
- Gordon, Matthew. 1997. *Urban sound change beyond city limits: the spread of the Northern Cities Shift in Michigan*, Ph.D. thesis, University of Michigan dissertation.
- Hillenbrand, J.M., L.A. Getty, M.J. Clark, and K. Wheeler. 1995. Journal of the acoustical society of america, *JASA*, 97(5), 3099–3111.
- Peterson, G. E. and H. L. Barney. 1952. Control methods used in a study of the vowels, *Journal of the Acoustical Society of America*, 24, 175–184.
- Shannon, R.V., A. Jensvold, M. Padilla, M.E. Robert, and X. Wang. 1999. Consonant recordings for speech testing, *Journal of the Acoustical Society of America*, 106(6), L71–L74.

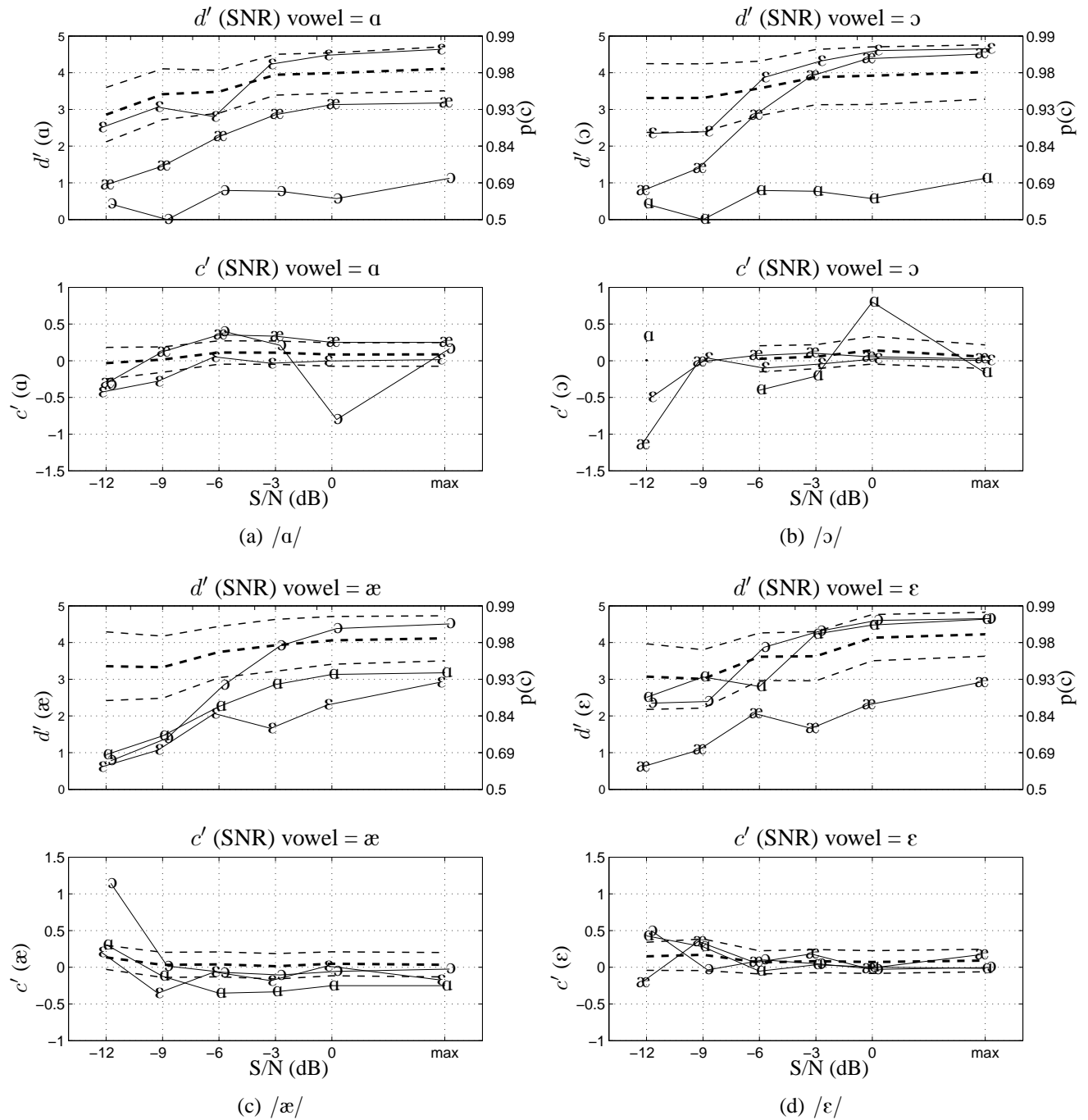


Figure 3: SDT analysis - The thick dotted line in each figure represents the average over all 14 possible phoneme pairs, with 95% confidence intervals represented by the thin dotted line. Results for relevant phonemes are plotted with their IPA symbol as a marker. Equivalent proportion correct is indicated on the right-hand axis of the d' plots.