CONTRAST SALIENCE AND TALKER NORMALIZATION IN NONSIBILANT FRICATIVE PERCEPTION

Molly Babel[1] and Grant McGuire[2]
[1]Dept. of Linguistics, University of British Columbia, 2613 West Mall, Vancouver, BC, CANADA, V6T 1Z4
[2]Dept. of Linguistics, University of California at Santa Cruz, Stevenson Faculty Services, Santa Cruz, CA, USA 95064

1. INTRODUCTION

The contrast between /f/ and /θ/ is notoriously difficult to differentiate acoustically; Ladefoged and Maddieson (1996: 173) attribute this difficulty to individual differences in production. Such a claim would predict that when presented with auditory tokens of /f/ and /θ/, listeners would exhibit improved performance when faced with the voice of a single talker. We investigate this claim by comparing listener performance in an /f/ and /θ/ Yes-no style task where one group of participants is presented with the stimuli blocked by talker while the other is presented with a mixed-talker design.

2. METHODS

Stimuli. Five male and five female native speakers of North American English with some phonetics training provided the stimuli. Audio recordings were made in a sound attenuated room. Subjects wore a head-mounted AKG C250 microphone positioned about two inches to the side of the mouth. Productions were digitally recorded to the hard drive of a PC at a 44K sampling rate. Stimuli were displayed visually to the talker in a randomized order and consisted of the fricatives /f/ and /θ/ in CV, VCV, and VC contexts where the vowel was either /a/, /i/, or /u/ for a total of 18 stimuli. VCV tokens were consistently produced by the talkers with a H* accent on the initial vowel and L% on the second vowel.

Procedures. Both experiments used a Yes-no style task. Subjects were presented a single token per trial and responded <f> or <th> on a button-box. Subjects were encouraged to respond in less than 1000 ms. Subjects classified all tokens for the 10 talkers three times for a total of 540 trials. In the BLOCKED condition participants completed one talker before moving on to the next, creating a total of 10 blocks. In the MIXED condition participants were presented with all ten talkers in a single block; presentations were randomized across a total of 3 blocks.

Subjects. Thirty-four participants from UC Santa Cruz participated in the Blocked condition and thirteen participants from UBC completed the Mixed Condition. All participants were native speakers of North American English and had no speech, language, or hearing disorders.

3. RESULTS

Sensitivity. Accuracy scores were converted to *d'* according to MacMillan and Creelman (2005). Correct /f/ responses were assigned as 'hits' and identification of an /f/ response to a /θ/ trial was labeled as a 'false alarm'. A d' score of 0 indicates no sensitivity to the contrast and that subjects are responding randomly. Considering the unbalanced number of participants in the two conditions, the data was subjected to a linear mixed effects model for analysis. The dependent measure was d', Condition and Talker Gender were entered as fixed effects, and Subject was treated as a random effect. Main effects were found for Condition ($\beta = 0.23$, $t = 2.2$, $p < 0.05$) and Talker Gender ($\beta = 0.08$, $t = 2.2$, $p < 0.05$). These effects are shown in Figure 1. Listeners demonstrated increased sensitivity to male talkers and increased sensitivity overall in the Mixed condition.

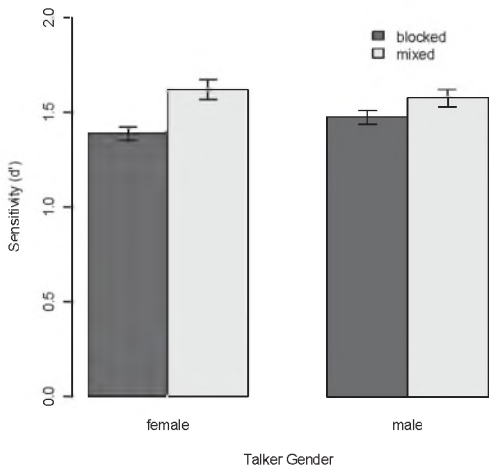

Figure 1. Main effects of Condition and Talker Gender on *d'* for the blocked and mixed conditions.

The criterion measure, or bias, was also subjected to a linear mixed effects model. Based on the arbitrary choice of calling correct /f/ identification a "hit", a positive criterion value indicates a bias to respond /f/ and a negative criterion value indicates a bias to respond /θ/. Criterion served as the dependent measure in the model, while Condition and Talker Gender served as fixed effects and Subject as a random effect. The model revealed a single main effect of Gender ($\beta = -0.34$, $t = -13.4$, $p < 0.001$). This effect is

shown in Figure 2; in both Conditions, listeners had a bias to respond /f/ to female talkers and /θ/ to male talkers.

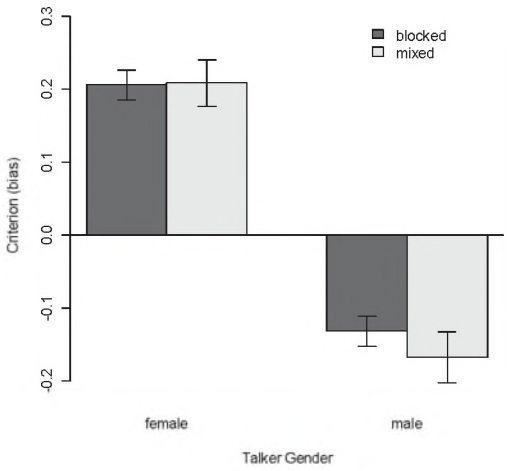

Figure 2. Main effect of Talker Gender on bias, shown for the blocked and mixed conditions.

Reaction Time. Log reaction times from correct responses were also subjected to a linear mixed effects model as the dependent measure. Prior to this, responses logged in under 200 ms were removed and outliers more than three standard deviations from the mean were removed as well. Fricative, Talker Gender, and Condition were entered as fixed effects, while Subject was a random effect. Condition (β = 0.08, t = 2.3, $p < 0.05$), Talker Gender (β=0.02, t = 4.3, $p < 0.001$), Fricative (β = 0.01, t = 2.4, $p < 0.05$) returned as main effects. There was a significant two-way Condition x Fricative interaction (β = 0.03, t = 3.3, $p < 0.001$) and a significant three-way Condition x Fricative x Talker Gender interaction (β = -0.03, t = -2.3, $p < 0.05$). Listeners responded more quickly in the blocked talker condition. Responses were also logged faster in response to female talkers and to /f/ tokens. These main effects are relatively consequential compared to the three-way interaction shown in Figure 3. Listeners responses in the mixed talker condition were fastest in response to female talkers /f/ tokens.

4. DISCUSSION

The results of this project both replicates previous work and contributes novel findings. The increased response latencies for the mixed talker condition replicates classic findings within the talker normalization literature (Martin et al., 1989; Mullennix et al., 1989), whereby increasing the number of talkers in the stimulus set prompts a delay in response latencies due to increased processing demands. However, the increase in listener sensitivity in the mixed talker condition is surprising and novel, as these same researchers who find a delay in response time in multi-talker conditions also find a decrease in accuracy. Our results suggest that for this particularly difficult acoustic contrast (Miller & Nicely 1955, Tabain 1998), even a slight increase in response latency produces more accurate (as measured by the sensitivity measure *d'*) responses. Despite an increase in response time across conditions, note that the bias effect remains constant. Regardless of whether listeners hear male and female talkers intermingled or separated, listeners have a bias to response /θ/ to male talkers and /f/ to female talkers.

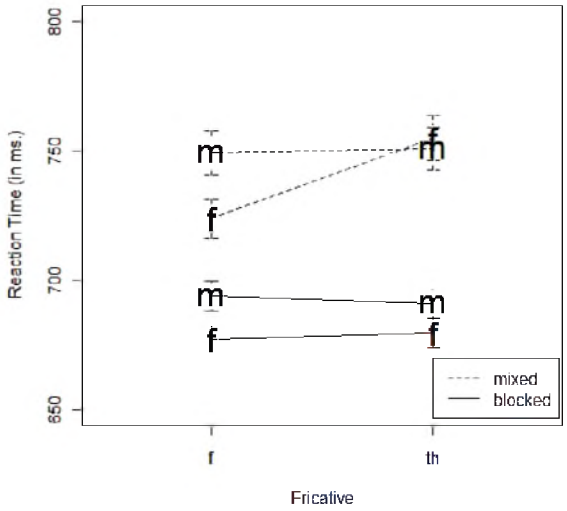

Figure 3. Reaction time by fricative for the mixed and blocked talker conditions. In this figure, "m" = male talkers and "f" = female talkers

REFERENCES

Ladefoged, P. & I. Maddieson. (1996). *The Sounds of the World's Languages*. Cambridge, MA: Blackwell Publishers.

Martin, C. S., Mullennix, J. W.. Pisoni. D. B., & Summers, W. V. (1989). Effects of Talker Varability on Recall of Spoken Word Lists. J. of Exp. Psych., 15, 676-684.

Miller, G. A. & P. E. Nicely. (1955). An analysis of perceptual confusions among some English consonants. *JASA*. 27 (2): 338-352.

Mullennix, J. W., Pisoni, D. B., & *Martin*, C. S. (1989). Some effects of talker variability on spoken word recognition. *JASA*. 85, 365-376.

Tabain, M. (1998). Non-sibilant fricatives in English: Spectral Information above 10 kHz. *Phonetica*, 55, 107-130.

ACKNOWLEDGEMENTS

Thanks to UBC and UC Santa Cruz for project funding. Thanks to Ryan Bennett Meaghan Delaney, Soraya Savji, and Travis Whitebread for assistance with this project.

AUTHOR NOTES

Author order is alphabetical; both authors contributed equally to this project.