Listeners cope with speaker and accent variation differently: Evidence from the Go/No-go task

Buddhamas Kriengwatana[1], Paola Escudero[2], Josephine Terry[2]

[1]Institute for Biology Leiden, Leiden University, the Netherlands
[2]MARCS Institute, University of Western Sydney, Australia

bkrieng2@alumni.uwo.ca, paola.escudero@uws.edu.au, j.terry@uws.edu.au

Abstract

The present study tests the hypothesis that speaker and accent normalization are mediated by distinct mechanisms, using a Go/No-go paradigm. Listeners naive to Dutch vowels and to Dutch and Flemish accents were trained to discriminate isolated /ɪ/ and /ɛ/ vowel tokens produced by a female Dutch speaker, and then tested on their categorization of /ɪ/ and /ɛ/ vowels from a different female Dutch speaker, a male Dutch speaker, a female Flemish speaker, and a male Flemish speaker. Our results demonstrate that listeners can correctly categorize vowels in the context of a speaker and gender change, but are unable to do so in the context of an accent or an accent and gender change. This supports our hypothesis that human listeners have separate mechanisms to cope with speaker versus accent variation in vowel productions: a mechanism that is intrinsic for speaker and gender versus a learned mechanism for accent. These results also demonstrate that the XAB task and Go/No-go task produce comparable results, which enables comparisons of Go/No-go responses in humans to non-human listeners.

Index Terms: speech perception, speaker normalization, accent normalization, vowels, Go/No-go paradigm

1. Introduction

The partitioning of relevant phonetic information is fundamental to speech perception and linguistic comprehension. However, an intrinsic property of human speech is the large variation with which different speakers produce the same speech sounds. The acoustic properties of vowels provide compelling evidence of this problem: the same vowel produced by different speakers varies tremendously, especially if speakers are of different ages, genders, and socio-linguistic backgrounds [1, 2]. Figure 1 illustrates this variation with F1 and F2 values for the Dutch vowels /ɪ/ and /ɛ/.

We hypothesize that different normalization mechanisms handle speaker/gender and accent variation because variation between speakers and genders is influenced by both physiological and sociocultural factors [3, 4, 5] whereas variation between accents is influenced by only the latter. Specifically, differences between speakers and genders are partially attributable to differences in vocal tract anatomy [6] and seem to be normalized automatically at a pre-attentive and subcortical level [7, 8, 9]. Demonstrations of speaker and gender normalization of phonemes by pre-linguistic infants and non-human animals with little or no previous language exposure support this claim [10, 11, 12].

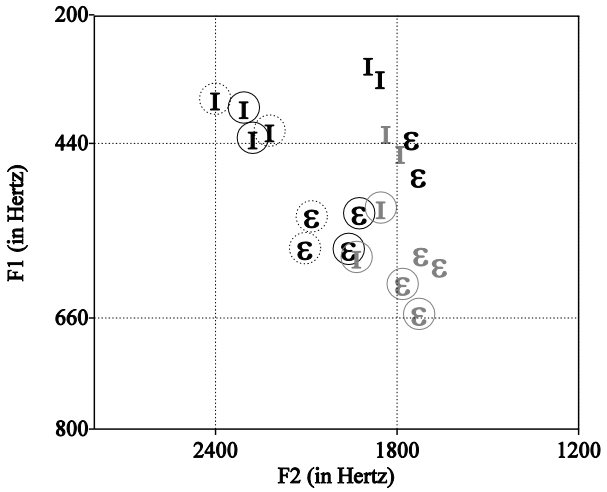

Figure 1: *F1 and F2 values of the vowels used in the present study: Females circled, trained Dutch female (black), unfamiliar Dutch female (dotted), Dutch male (black), Flemish female and male (grey).*

Sensitivity to relative formant ratios instead of absolute formant frequencies may explain the automaticity of speaker and gender normalization. Specifically, transformation of American English vowels into formant ratios between the first and third formant frequency (F1/F3) and the second and third formant frequency (F2/F3) eliminates much of the speaker and gender variation [7]. Figure 2 shows that a conversion of the Dutch and Flemish values in Figure 1 as F1/F3 and F2/F3 ratios greatly reduces speaker and gender, but not accent variation. This suggests that the human perceptual system may handle speaker and accent variation very differently, as dialect variation results from cultural transmission and has no consistent physiological correlates. In other words, speaker and gender normalization mechanisms may not be adequate to deal with unfamiliar accents because of the learning required to normalize accents.

In support of this, empirical studies have shown that successful accent normalization of words requires experience [13, 14], recruits lexical processing in children and adults [15, 16], and that distinct neural regions are activated in response to different speakers versus different accents [17]. Taken together, these findings suggest that human listeners have two separate mechanisms to cope with speaker versus accent

variation in vowel productions: a mechanism that is intrinsic and phylogenetically-shared for speaker and gender versus a learned mechanism for accent. However, empirical evidence showing that listeners can cope with speaker and gender variability but require further information to do away with accent variation is currently lacking.

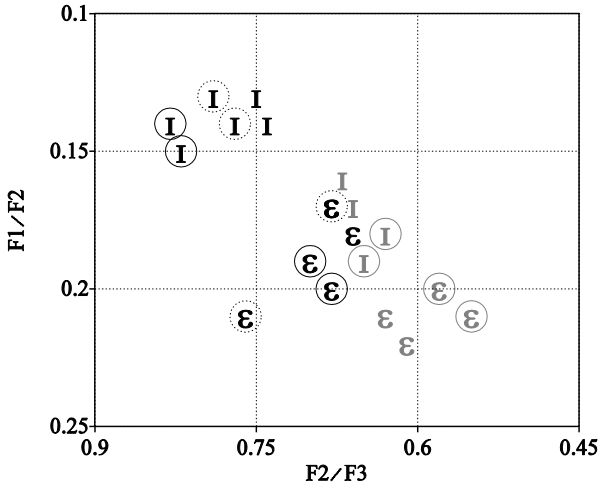

Figure 2: *Formant ratios of the stimuli in Figure 1: Females circled, trained Dutch female (black), unfamiliar Dutch female (dotted), Dutch male (black), Flemish female and male (grey).*

To test this hypothesis, it is imperative that studies on human infants, adults, and non-human animals use comparable experimental procedures. However, this is not a simple undertaking, as the different populations have their own specific constraints.

The XAB and Go/No-go tasks are two behavioral paradigms that have been used to assess perceptual categorization of speech sounds [18, 19]. In the XAB task, listeners hear a sequence of three stimuli and must indicate which of latter two stimuli belong to the same category as the first stimulus. In the Go/No-go task, listeners are first trained to respond (Go) or withhold responses (No-go) to particular sets of stimuli, and then tested on their responses to novel, unfamiliar stimuli. While the XAB task has its advantages (i.e. is less time consuming, places few demands on short-term memory, and many comparisons can be tested in one session), the Go/No-go task is much more suitable for tests of speech perception in non-human animals because animals can be trained to discriminate Go from No-go stimuli using feedback, without the need for verbal instructions. Consequently, studies using identical stimuli in both the XAB and Go/No-go tasks are required to see whether performance on these tasks is comparable across humans, with the goal of comparing Go/No-go responses in humans versus non-human listeners.

The present study used the Go/No-go paradigm to test the hypothesis that normalization of vowels produced by a different speaker/gender in the same accent is handled by a different mechanism than normalization of vowels produced by a different speaker/gender in a different accent (which do not contain lexical information or any other cue as to the accent divergence). This is because accent normalization requires context establishing that the accent has changed or prior exposure with feedback in order to map accent-related deviations in pronunciation to the same phoneme representations, whereas speaker and gender normalization will occur automatically and without any context or prior knowledge, as this normalization is performed by the human auditory system.

We tested categorization of the Dutch and Flemish vowels /ɪ/ and /ɛ/ by Australian-English listeners who are naïve to the Dutch and Flemish accents. From Figure 1, it is evident that the Dutch /ɛ/ and Flemish /ɪ/ are acoustically very similar. Consequently, we predicted that Australian listeners who have no knowledge of Dutch and Flemish accent differences would incorrectly categorize the Flemish /ɪ/ as a Dutch /ɛ/. That is, we predict that in the absence of lexical information or knowledge that the stimuli are produced in a different accent, Australian listeners will not be able to normalize the Flemish vowels because even after speaker/gender normalization the Flemish /ɪ/ still occupies the same vowel space of a different Dutch vowel, namely /ɛ/. If this prediction is supported, it would corroborate recent findings using a more common task in speech perception research, namely the XAB task [20. The XAB task requires an overt response regarding phonemic identity with two specific options, namely /ɪ/ and /ɛ/ in this case, while the Go/No-go task can be regarded as phoneme monitoring task, where listeners judge whether a sound is associated with the phoneme that corresponds to a button press. Finding similar results across tasks will demonstrate that despite their methodological differences, the XAB and Go/No-go tasks test vowel categorization in comparable ways.

2. Methods

2.1. Participants

Sixteen listeners (11 females, 5 males) who were undergraduate students at the University of Western Sydney participated in the study in exchange for course credit (Mean age = 24.06 years, *SD* = 8.88). Eight listeners were monolingual Australian English, while the other eight spoke another language other than English (Arabic, French, Mandarin Chinese, Japanese, Serbian, Korean, Macedonian, and Thai).

None of the listeners had prior exposure to Dutch or to any Dutch accent. The same number of multilingual listeners was assigned to each of the two stimuli conditions reported below.

2.2. Stimuli and Procedure

Stimuli were natural isolated Dutch vowels /ɪ/ and /ɛ/ extracted from the syllable [s-Vowel-s] which was read in the carrier sentence /ɪn sVs ən ɪn sVsə zɪt də V/ [21]. Two tokens of each vowel from each speaker were used. They were ramped at the beginning and ended with a 5-ms fade period, and duration and intensity were equalized using the PRAAT program [22]. They were presented via headphones attached to a laptop computer running E-prime version 2.

Stimuli were presented within the Go/No-go task, which as mentioned above requires participants to learn to respond to one stimulus set (Go) and inhibit responses to another stimulus set (No-go). For our task to be comparable with Go/No-go tasks administered to non-human animals (i.e. zebra finches), the listeners of the present study were given minimal verbal instructions. That is, they were simply informed that they were to listen to sounds and to determine what was required to elicit a correct response and to avoid an incorrect response in order to earn as many points as possible. To earn points, participants

had to learn to press "spacebar" to one set of sounds (Go tokens), but to withhold a response and not press any key to the other set of sounds (No-go tokens).

Participants initiated each trial by pressing the spacebar. After the presentation of a token, the text "Press Spacebar if appropriate" appeared on the screen. If participants pressed the spacebar within 2000 ms after hearing the Go tokens, they were positively reinforced with a "smiley face", a pleasant "ding" sound, and 1 point was awarded. If they pressed the spacebar within 2000 ms after hearing the No-go tokens, they were penalized with a "sad face", an unpleasant "punch" sound, and no point was awarded. Feedback appeared on the screen for 2000 ms. Points were awarded to motivate participants to try to make correct responses, much in the same way zebra finches in Go/No-go tasks are motived to make correct responses in order to obtain food [19, 23].

Participants completed two training phases, and a final test phase. For task familiarization, the first phase (two blocks of 10 trials) rewarded participants for correct Go, and No-go responses to easily discriminable syllables "deet" and "pon". In the second phase, Go and No-go tokens were the vowels /ɪ/ and /ɛ/ spoken by a Dutch female as shown in Figures 1 and 2 (i.e. the Trained Go and Trained No-go tokens). There were three blocks of 20 randomly presented trials and the allocation of the /ɪ/ and /ɛ/ vowels as the Go token was counterbalanced across participants.

In the test phase, trained (80%) and unfamiliar (20%) vowel tokens were presented randomly over six blocks of 20 trials (in total 120 trials, 96 trained, 24 unfamiliar). Feedback was not given for responses to unfamiliar vowel tokens, nor to 25% of responses to trained tokens. Unfamiliar tokens of /ɪ/ and /ɛ/ were either spoken by a different Dutch female speaker (*speaker change*), a Dutch male speaker (*gender change*), a Flemish female speaker (*accent change*), or a Flemish male speaker (*accent+gender change*). The acoustic values of the vowel tokens are shown in Figures 1 and 2. Each participant completed either the speaker and gender change conditions, or the accent and accent+gender change conditions. The presentation order of conditions was counterbalanced. Each experimental session lasted one hour, with each condition taking 25 minutes to complete.

Data analysis was conducted on trials in the test phase where no feedback was given. Responses to these 48 trials (12 Trained Go, 12 Trained No-go, 12 Unfamiliar Go, 12 Unfamiliar No-go) were analyzed using generalized linear mixed models (GLMMs) with a binary logistic regression. Statistical analyses were performed in SPSS 21.0.

Separate models were used to analyze the proportion of Go responses for each type of stimulus change (speaker, gender, accent, and accent+gender). In all models, token type (Trained Go, Trained No-Go, Unfamiliar Go, Unfamiliar No-go), was entered as a fixed effect. Satterthwaite correction for degrees of freedom was applied, as required when performing GLMMs that use pseudolikelihood estimation [24]. To assess differences between categorization of trained and unfamiliar token types, simple planned comparisons comparing the Trained Go to the Trained No-go, Unfamiliar Go and Unfamiliar No-go tokens were performed. If participants classified the unfamiliar tokens correctly, then there should be a significant difference between Trained Go and Trained No-go, and Trained Go and Unfamiliar No-go, whereas there should be no difference between Trained Go and Unfamiliar Go.

3. Results

The results of our analysis for the speaker change and gender change stimuli condition are shown in Figure 3. These results demonstrate that listeners were able to correctly categorize the /ɪ/ and /ɛ/ tokens of the unfamiliar Dutch female and unfamiliar Dutch male (main effect of token type, $F(3, 15) = 37.07$, $p < 0.001$ and $F(3, 16) = 13.80$, $p < 0.001$ for speaker and gender, respectively).

For both speaker and gender change, simple planned comparisons revealed that responses to the Trained No-go and Unfamiliar No-go differed from the Trained Go tokens ($p < 0.001$ for all; Figure 3). Responses to the Unfamiliar Go were similar to responses to the Trained Go tokens (Figure 3).

Participants' responses to accent change and accent+gender change show a very different pattern, as can be seen from Figure 4. Listeners were unable to correctly categorize the /ɪ/ and /ɛ/ tokens of an unfamiliar Flemish female or Flemish male, even if they could correctly classify the /ɪ/ and /ɛ/ tokens of a Dutch female (main effect of token type, $F(3, 13) = 12.24$, $p < 0.001$ and $F(3, 13) = 8.59$, $p = 0.002$ for accent and accent+gender, respectively).

When confronted with an accent change, participants' responses to the Unfamiliar Go and Unfamiliar No-go tokens did not differ significantly from their responses to the Trained Go tokens. However, participants continued to distinguish between the Trained Go and Trained No-go tokens ($p < 0.001$; Figure 4). This indicates that despite continuing to correctly categorize vowel tokens produced by a Dutch female, listeners could not correctly classify vowel tokens if they were produced by an unfamiliar Flemish female.

Similarly, when confronted with an accent+gender change, participants' responses to the Trained No-go, Unfamiliar Go and Unfamiliar No-go all differed significantly from the Trained Go ($p < 0.01$ for all; Figure 4). This indicates that listeners correctly classified vowel tokens produced by a Dutch female, but could not correctly categorize vowel tokens produced by an unfamiliar Flemish male, as they treated both the Unfamiliar Go and Unfamiliar No-go tokens differently from the Trained Go tokens.

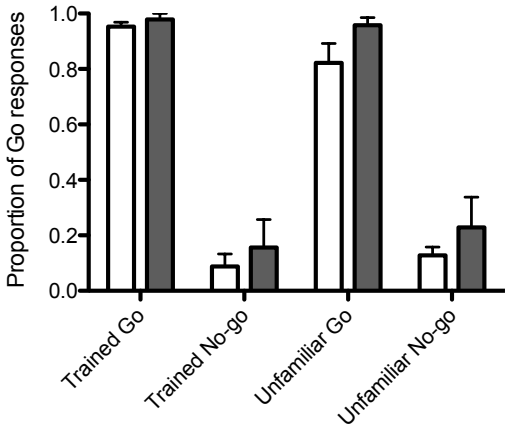

Figure 3: *Proportion of Go responses to the different token types when a Dutch female (white) and a Dutch male (grey) were the unfamiliar tokens. Error bars are SEM.*

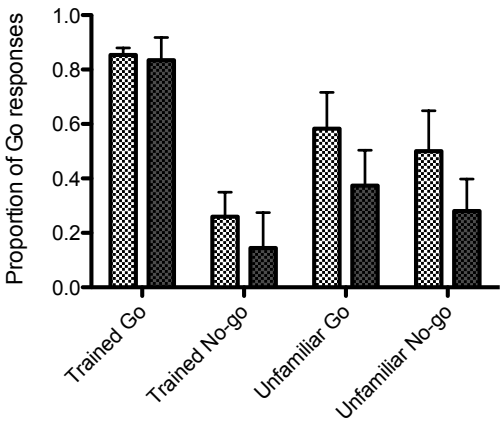

Figure 4: *Proportion of Go responses to the different token types when a Flemish female (hatched) and a Flemish male (grey hatched) were the unfamiliar tokens. Error bars are SEM.*

4. Discussion and Conclusion

The results of the present study clearly show that listeners who are naïve to Dutch vowels and to the Dutch and Flemish accents can normalize speaker and gender differences, but not accent differences. That is, they were able to correctly categorize the /ɪ/ and /ɛ/ vowel tokens spoken by two unfamiliar Dutch speakers, but not those spoken by two unfamiliar Flemish speakers. Therefore, these findings support our hypothesis that separate mechanisms handle speaker/gender and accent variation – especially because we obtained these results despite the diversity of language backgrounds in our sample of participants. Replication of these findings with different voices would indicate that the patterns observed in the present study are generalizable.

Our results are not in complete agreement with those of the XAB task used by Chládková et al. [20] because we find that listeners are close to chance performance when categorizing Flemish tokens of either /ɪ/ or /ɛ/ (Figure 4), whereas they found that listeners from the same subject pool as those reported here classified Flemish /ɪ/ as Dutch /ɛ/. This difference may be explained by the presentation of both Flemish /ɪ/ and /ɛ/ vowels in our study and the presentation of only Flemish /ɪ/ in Chládková et al.'s study [20], which may lead to a different pattern of classification across studies.

Consequently, data from another study using the XAB task including both Flemish vowels is required to ascertain whether listeners' classification in the XAB task results in a similar shift in classification as that shown in the Go/No-go task we used here. Nevertheless, the results of the present study and those of Chládková et al. [20] still point to the same conclusion: speaker and accent variation are handled differently.

5. Acknowledgements

We thank Samantha Taylor for help with participant testing.

6. References

[1] Peterson, G. E. and Barney, H. L., "Control methods used in a study of the vowels", J. Acoust. Soc. Am., 24(2):175-184, 1952.

[2] Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K., "Acoustic characteristics of American English vowels", J. Acoust. Soc. Am., 97(5):3099-3111, 1995.

[3] Bachorowski, J. A. and Owren, M. J., "Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech", J. Acoust. Soc. Am., 196:1054-1063, 1999.

[4] Dolson, M., "The pitch of speech as a function of linguistic community", Music Percept. 11(3):321-331, 1994.

[5] van Bezooijen, R., "Sociocultural aspects of pitch differences between Japanese and Dutch women", Lang. Speech, 38:253-265, 1995.

[6] Fitch W. T. & Giedd J., "Morphology and development of the human vocal tract: A study using magnetic resonance imaging", J. Acoust. Soc. Am., 106(3):1511-1522 ,1999.

[7] Monahan, P. J. and Idsardi, W J., "Early auditory sensitivity to formant ratios: Towards a perceptual account of vowel normalization", Language and Cognitive Processes, 25:808-839, 2010.

[8] Tuomainen, J., Savela, J., Obleser, J., and Aaltonen, O., "Attention modulates the use of spectral attributes in vowel discrimination: behavioral and event-related potential evidence", Brain Research, 1490:170–183, 2013.

[9] von Kriegstein, K., Warren, J. D., Ives, D. T., Patterson, R. D., and Griffiths, T. D., "Processing the acoustic effect of size in speech sounds", NeuroImage, 32:368-375, 2006.

[10] Kuhl, P. K., "Perception of auditory equivalence classes for speech in early infancy", Infant Behaviour & Development, 6:263-285, 1983.

[11] Burdick, C. K. and Miller, J. D., "Speech perception by the chinchilla: discrimination of sustained /a/ and /i/", J. Acoust. Soc. Am., 58:415–427, 1975.

[12] Dooling, R. J. and Brown, S. D., "Speech perception by budgerigars (Melopsittacus undalatus): spoken vowles", Percept Psychophys., 47(6):568-74, 1990.

[13] Bradlow, A. R. and Bent, T., "Perceptual adaptation to non-native speech", Cognition, 106:707-729, 2008.

[14] Van Heugten, M. and Johnson. E. K., "Learning to contend with accents in infancy: Benefits of brief speaker exposure", Journal of Experimental Psychology: General, 143:430-450, 2004.

[15] Clarke, C. M. and Garrett, M. F., "Rapid adaptation to foreign-accented English", J. Acoust. Soc. Am., 116:3647-3658, 2004.

[16] White, K. S. and Aslin, R. N., "Adaptation to novel accents by toddlers", Developmental Science, 14:372-384, 2011.

[17] Adank, P., Davis, M., and Hagoort, P., "Neural dissociation in processing noise and accent in spoken language comprehension", Neuropsychologia, 50:77–84, 2012.

[18] Escudero, P., and Wanrooij, K., "The effect of L1 orthography on non-native vowel perception," Lang. Speech, 53:343–365, 2010.

[19] Ohms, V. R., Escudero, P., Lammers, K., and ten Cate, C, "Zebra finches and Dutch adults exhibit the same cue weighting bias in vowel perception", Animal Cognition, 15:155-161, 2012.

[20] Chládková, K., Geambaşu, A., Dadwani, R., Peter, V., and Escudero, P., "Speaker and dialect variation are handled differently: Behavioural and pre-attentive evidence", in preparation.

[21] Adank, P., Smits, R., van Hout, R., "A comparison of vowel normalization procedures for language variation research", J. Acoust. Soc. Am., 116(5):3099-3107, 2004.

[22] Boersma, P., "Praat, a system for doing phonetics by computer", Glot International 5(9/10):341-345, 2001.

[23] Ohms, V. R., Gill, A, van Heijningen, C. A. A., Beckers, G. J. L., and ten Cate, C., "Zebra finches exhibit speaker-independent phonetic perception of human speech", Proc R Soc B, 277:1003–1009, 2010.

[24] Stroup, W., Generalized Linear Mixed Models: Modern Concepts, Methods and Applications, CRC Press, 2012.