Adult listeners' processing of indexical versus linguistic differences in a pre-attentive discrimination paradigm

Dadwani, R.; Peter, V.; Chládková, K.; Geambasu, A.; Escudero, P.

Adult listeners' processing of indexical versus linguistic differences in a pre-attentive discrimination paradigm

Rozmin Dadwani[1], Varghese Peter[1], Kateřina Chládková[2], Andreea Geambaşu[3], Paola Escudero[1]

[1]University of Western Sydney, [2]University of Amsterdam, [3]Leiden University

r.dadwani@uws.edu.au, v.peter@uws.edu.au, k.chladkova@uva.nl, a.geambasu@hum.leidenuniv.nl, paola.escudero@uws.edu.au

ABSTRACT

The human ability to comprehend speech regardless of variation across speakers and accents has long puzzled researchers. Human listeners appear to employ separate mechanisms to cope with speaker versus accent variation. The present study uses event-related potentials (ERP) to test whether such different mechanisms exist at a pre-attentive level of speech processing.

We assessed Australian English monolinguals' and bilinguals' perceptual sensitivity to four types of variation in vowels: namely, variation in speaker identity, gender, accent, and vowel category. Interestingly, listeners showed similar results regardless of their linguistic background. As expected, listeners showed large sensitivity to accent changes. Rather surprisingly, however, they were more sensitive to changes in speaker gender than to changes in vowel category. These results are not in line with those of overt vowel classification but are explained by adults' sensitivity to large differences in voice quality when discriminating speech sounds.

Keywords: pre-attentive discrimination, indexical versus linguistic information, vowels, MMN

1. INTRODUCTION

An intrinsic property of human speech is the large variation with which different speakers produce the same speech sounds. For linguistic processing, adult listeners seem to readily ignore indexical differences (such as between-gender differences in vowel formants) in order to classify speech sounds as the same phoneme. We hypothesize that speaker/gender and accent variation are handled by different normalization mechanisms because variation between speakers or genders is influenced by both physiological and sociocultural factors [2, 5, 23], whereas variation between accents is influenced by only the latter. That is, between-speaker and between-gender differences are partially attributable to variation in vocal tract anatomy [7] and can be normalized automatically, even by non-human animals [14, 22, 24].

Some previous studies have assumed that during speech comprehension, differences between speakers and accents are normalized using the same underlying mechanism [20]. However, accent normalization by adults and older children appears to rely on lexical knowledge and exposure to the specific accent [3, 25]. In contrast, speaker normalization seems to occur automatically and without lexical knowledge since not only young human infants [8, 11] who do not possess a lexicon, but also other mammals and birds can discriminate human speech sounds and syllables across different voices [6, 9, 12, 17]. This suggests that speaker normalization might be a general auditory process whereas accent normalization might be a learned linguistic process.

The results of a recent behavioral study involving vowel categorization [10] showed 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. Kriengwatanna et al. [10] used a Go/No-go task where listeners judge whether a sound is associated with the phoneme that corresponds to a button press. They tested categorization of the Dutch and Flemish vowels /ɪ/ and /ɛ/ (see Figure 1) by Australian-English listeners. The listeners 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. These results confirm that speaker/gender and accent variation are indeed handled by separate mechanisms. The question is whether these differences will be seen in a pre-attentive discrimination paradigm. The present study examines whether these behavioral findings are reflected at a pre-attentive level of processing.

Following [10], we also used a case that constitutes one of the most compelling demonstrations of the extreme variability in the speech signal, namely that in the acoustic properties of vowels [19, 21]. Fig. 1 illustrates this variability with the Dutch vowels /ɪ/, /ɛ/, /u/, and /ɔ/ produced by female and male speakers from North Holland (NL) and from East Flanders (VL) (from the corpus of Adank et al. [1]). The figure shows that there are large differences between the

first and second formants (F1 and F2) of vowels produced by women and those produced by men. It is also shown that the front vowels /ɪ/ and /ɛ/ exhibit large differences between the two accents: the VL vowel /ɪ/ is acoustically closer to the NL vowel /ɛ/ than it is to the NL vowel /ɪ/.

Figure1: Gender and accent variation in F1 and F2 of Dutch vowels. Larger light symbols: vowels produced by women, smaller dark symbols: vowels by men. Circled: vowels from North Holland, plain: vowels from East Flanders.

Accent normalization may appear to be similar to speaker normalization in the context of familiar words, i.e. in the presence of abstract linguistic (lexical) knowledge. Isolated speech sounds, on the other hand, can deliver the required degree of control that will assure that the results are specifically related to the two types of variability (speaker and accent) without any possible covariate. Therefore, we followed a previous behavioral vowel categorization study [10] in using isolated vowels.

We recorded behavior-independent responses of the auditory system to different types of variation in isolated vowels using electroencephalography (EEG), examining the mismatch negativity (MMN). The MMN is widely considered a marker of pre-attentive change detection and is obtained for simple and complex patterns of auditory changes. It is also affected by listeners' experience with abstract linguistic categories for sounds [15, 16]. We predicted that changes in accent will elicit an MMN comparable to the MMN elicited by a change in vowel category but larger than the MMN elicited by changes in speaker or gender. If this prediction is supported, it would corroborate recent findings [10] of the behavioral vowel categorization task mentioned above. Finding similar results across both implicit and explicit paradigms will support the hypothesis that speaker

and accent variation are mediated by distinct mechanisms.

2. METHODS

2.1. Participants

Eight bilinguals who spoke at least one other language other than Australian English (age range: 20-40; mean = 29; 4 females) and eight Australian English monolinguals (age range: 19-26; mean = 22.5; 6 females) from the University of Western Sydney participated in the study in exchange for course credit. All participants were right handed, and reported that they had normal hearing and no language or neurological impairments. Testing took place at the University of Western Sydney. Participants were tested individually in a single session in a sound proof, electrically-shielded room. They were seated in a comfortable chair and were instructed to avoid excessive blinking and movements. During stimulus presentation, they watched a silenced movie with subtitles in English, and were told not to pay attention to the sounds.

2.2. Stimuli

Fig. 2 shows the stimuli used in the present study, which were isolated, naturally produced tokens of Dutch vowels /ɪ/ and /ɛ/, which were selected from the same corpus mentioned in the introduction [1] and were a subset of those used in [10].

Figure 2: The standard (circled) and the four deviant stimuli from the ERP experiment. IPA symbols show the intended vowel, subscripts indicate the type of change.

We used a multi-deviant oddball paradigm where a frequently-repeated standard stimulus was interspersed with infrequent repetitions of four different deviant stimuli. The standard was a vowel produced by a female NL speaker and the four

deviants differed from the standard in speaker, gender, accent, and category membership, respectively. The ratio of presentation was 0.80 for the standard and 0.05 for each of the four deviants. The standards and deviants were presented in a pseudorandom order with at least three standards between the deviants. The inter-stimulus interval was randomly varied between 600-700 ms. The oddball block started with 20 standards, and contained a total of 3470 stimuli, which resulted in a total duration of 42 minutes.

After the oddball block, there were control blocks for each deviant type during which every deviant was repeatedly presented 120 times (which equals approximately 1.5 minutes per deviant type). To prevent fatigue or habituation, there were two breaks: one after 1735 trials (after the first 20 minutes), and one at the end of the oddball block.

2.3. EEG recording and pre-processing

The EEG was recorded from monolingual participants using a 64 channel BioSemi active2 system and from bilingual participants using a 128 channel EGI system.

2.3.1. EEG recording: Monolinguals

EEG was recorded from 64 active Ag-AgCl electrodes placed according to the International 10/20 placement in a cap (BioSemi) fitted to participant's head size. Six external electrodes were used: below and above the right eye, on the left and right temple (ocular activity), and on the right and left mastoid. The EEG signal was recorded at a sampling rate of 512 Hz.

2.3.2. EEG recording: Bilinguals

EEG was recorded from 128 active HydroCel electrodes placed in a Geodesic sensor net (HCGSN) connected to the Netstation 4.2 software. Data from each of the 128 electrodes were digitized at 1000 Hz with a vertex reference and band pass filter of 0.1–400 Hz, while electrode impedance was maintained below 50 kΩ.

2.3.3. EEG processing

The EEG was offline processed using EEGLAB toolbox [4] running in MATLAB 2012a for the monolinguals and BESA 6.0 (Brain Electrical Source Analysis) for the bilinguals. Portions of EEG with large artifacts were removed by visual inspection. The continuous EEG was divided into epochs from -100ms to +600ms relative to stimulus onset. For subsequent baseline correction the mean voltage in the 100-ms pre-stimulus interval was subtracted from each sample in the epoch. The epochs were low pass filtered at 30 Hz and referenced to the average of the mastoid electrodes. Eye blink artifacts were removed using independent component analysis (ICA). Noisy channels were interpolated using spherical spline interpolation. Epochs with amplitude exceeding +/- 75 µV were removed before averaging. The epochs were averaged separately for standards (excluding the first 20 standards and standards that immediately followed a deviant), for each deviant type, and for each control stimulus type.

Per participant, we obtained a difference wave for each deviant type by subtracting their average response to a stimulus in the control block from the average response to the same stimulus when it was presented as deviant in the oddball block. This was done in the same way for both experiments for all participants. Subsequently, for each deviant type we computed the grand average difference waveform by pooling across participants. In the grand average difference wave per deviant type we determined the latency of the negative peak in the window from 150 to 250 ms relative to stimulus onset. At this grand-average latency value, we centered a smaller 40-ms window for each participant individually. The average voltage within this 40-ms window served as our measure of the MMN amplitude. MMN amplitude was calculated at 9 electrode sites, F3, FC3, C3, Fz, FCz, Cz, F4, FC4 and C4.
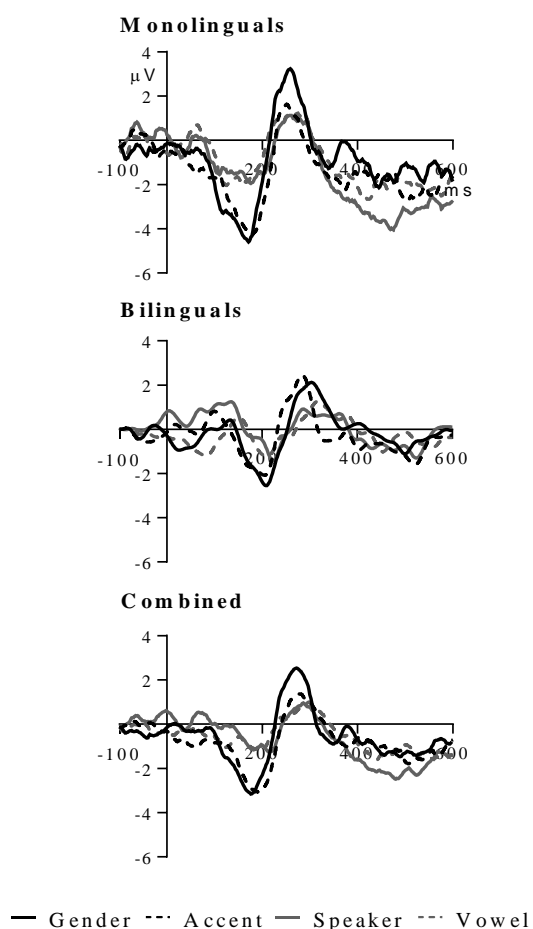
2.3.4. Statistical Analysis

The MMN amplitudes were submitted to a 4-way repeated-measures analysis of variance (ANOVA) with the between-subject factor Group (monolinguals, bilinguals) and the within-subjects factors: Deviant type (4 levels: Vowel category, Speaker, Gender, Dialect), Anteriority (3 levels: Frontal, FrontoCentral, Central) and Laterality (3 levels: left, midline, right). An alpha level of .05 was set as the criterion for statistical significance.

3. RESULTS

Figure 3 shows the grand average difference waves (i.e. deviant – control) for each deviant type for monolinguals (top row), bilinguals (middle row) and combined (bottom row) at the electrode FCz. Table 1 reports the mean MMN amplitudes for the 4 deviant types at the channel FCz.

Figure 3: The grand average difference waveforms at FCz for each deviant type.

Monolinguals

Bilinguals

Combined

— Gender ‑‑‑ Accent — Speaker ‑‑‑ Vowel

Table 1. MMN amplitudes for the four deviant types at the channel FCz.

Deviant Type	Group	MMN amplitude (µV)
Accent	Monolinguals	-4.36 (0.93)
	Bilinguals	-3.02 (0.43)
Gender	Monolinguals	-4.41 (1.15)
	Bilinguals	-3.63 (0.63)
Speaker	Monolinguals	-2.68 (0.78)
	Bilinguals	-2.36 (0.53)
Vowel	Monolinguals	-2.56 (0.72)
	Bilinguals	-1.93 (0.41)

The 4-way ANOVA on the MMN amplitudes revealed a main effect of deviant type $F(3,42)=6.23$, $p=.001$. Bonferroni corrected pairwise comparisons showed that both accent (M=-3.45 µV, SE=0.48) and gender (M=-3.61 µV, SE=0.62) deviances elicited larger MMN compared to speaker (M=-2.09 µV, SE=0.39) and vowel (M=-2.05 µV, SE=0.35) deviances. There was also a significant main effect of laterality $F(2, 28)=9.28$, $p=0.001$. MMN at central electrodes (M=-3.12 µV, SE=0.40) was significantly larger than it was at left (M=-2.57 µV, SE=0.35) and right (M=-2.57 µV, SE=0.42) electrodes. The main effect of group was not significant $F(1,14)=0.12$, $p=.73$, neither was the main effect of anteriority $F(2, 28)=0.48$, $p=0.62$. None of the interactions were significant (all $F< 2.2$, all $p>.05$).

4. DISCUSSION AND CONCLUSION

The present ERP experiment shows that adult listeners, irrespective of their language background have less sensitivity to variation across speakers than to variation across accents during pre-attentive listening. This is evident from the stronger mismatch response to an accent change. This is in line with the overt categorization results that show that accent differences, unlike speaker and gender differences, are not normalized in vowel classification [10].

The strong MMN elicited by the gender change demonstrates that listeners readily detect this type of indexical differences in a pre-attentive task, a result that is contrary to that of a previous overt vowel categorization task [10] using the same stimuli. It seems that the difference in paradigms (behavioral versus pre-attentive) partially influenced adult listeners' processing of the speaker versus accent variation. The strong sensitivity to changes in gender seems to be due to a large difference between the voice qualities of the male and female speakers used in the present study. It seems that voice quality is a particularly salient cue in an unattended discrimination task. Future research should confirm whether the difference between the present results and those in [10] are indeed task dependent or related to the specific design we used in the present ERP study.

5. REFERENCES

[1] Adank, P., Van Hout, R. and Van de Velde, H. 2007. An acoustic description of the vowels of northern and southern standard Dutch II: Regional varieties. Journal of the Acoustical Society of America 121, 1130-41.

[2] Bachorowski, J. A. and Owren, M. J. 1999. Acoustic correlates of talker gender and individual

talker identity are present in a short vowel segment produced in running speech, J. Acoust. Soc. Am. 196, 1054-1063.

[3] Clarke, C. and Garrett, M. 2005. Rapid adaptation to foreign accented English. Journal of the Acoustical Society of America 116, 3647-3658.

[4] Delorme, A. and Makeig, S. 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods 134, 9-21.

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

[6] Dooling, R.J. and Brown, S.D. 1990. Speech perception by budgerigars (Melopsittacus undulatus): spoken vowels. Perception & psychophysics 47(6), 568-574.

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

[8] Jusczyk, P.M., Pisoni, D.B. and Mullenix, J. 1992. Some consequences of stimulus variability on speech processing by two-month-old infants. Cognition 43, 253–291.

[9] Kluender, K.R., Diehl, R.L. and Killeen, P.R. 1987. Japanese quail can learn phonetic categorie. Science 237, 1195-97.

[10] Kriengwatana, B., Escudero, P. Terry, J. 2014. Listeners cope with speaker and accent variation differently: Evidence from the Go/No-go task. Proc. SST, Christchurch, 2014.

[11] Kuhl, P. K. 1983. Perception of auditory equivalence classes for speech in early infancy. Infant Behavior and Development 6, 263-285.

[12] Kuhl, P. and Miller, J. 1975. Speech perception by the chinchilla: Voiced voiceless distinction in alveolar plosive consonants. Science 190, 69–72.

[13] Ladefoged, P. and Broadbent, D.E. 1957. Information conveyed by vowel. Journal of the Acoustical Society of America 29, 98–104.

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

[15] Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. 2007. The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clinical Neurophysiology, 118, 2544-2590.

[16] Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. 2001. Primitive intelligence in auditory cortex. Trends in Neuroscience 25, 283-288.

[17] Ohms, V.R., Gill, A., Van Heijningen, C.A.A., Beckers, G.J.L. and Ten Cate, C. 2010. Zebra finches exhibit speaker-independent phonetic perception of human speech. Proceedings of the Royal Society London B. 277, 1003-1009.

[18] Peterson, G.E. and Barney, H.L. 1992. Control methods used in a study of the vowels. Journal of the Acoustical Society of America 23(2), 175-184.

[19] Pierrehumbert, J. 2002. Word-specific phonetics. In: Gussenhoven C, Warner N (eds.) Laboratory phonology. Berlin: Mouton de Gruyter, 101-139.

[20] Potter, R.K., and Steinberg, J.C. 1950. Toward the Specification of Speech. Journal of the Acoustical Society of America 22, 807-819.

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

[22] Van Bezooijen, R. 1995. Sociocultural aspects of pitch differences between Japanese and Dutch women. Lang. Speech 38, 253-265.

[23] Von Kriegstein, K., Warren, J. D., Ives, D. T., Patterson, R. D., and Griffiths, T. D. 2006. Processing the acoustic effect of size in speech sound. NeuroImage 32, 368-375, 2006.

[24] White, K.S. and Aslin, R.N. 2011. Adaptation to novel accents by toddlers. Developmental Science 14(2), 372-384.