Interspeech 2018
2-6 September 2018, Hyderabad

Perceptual sensitivity to spectral change in Australian English close front vowels: an electroencephalographic investigation

Daniel Williams[123], *Paola Escudero*[23], *Adamantios Gafos*[1]

[1]Linguistics Department, University of Potsdam, Germany
[2]MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Australia
[3]ARC Centre of Excellence for the Dynamics of Language, Australia
daniel.williams@uni-potsdam.de, paola.escudero@westernsydney.edu.au, gafos@uni-potsdam.de

Abstract

Speech scientists have long noted that the qualities of naturally-produced vowels do not remain constant over their durations – regardless of being nominally "monophthongs" or "diphthongs". Recent acoustic corpora show that there are consistent patterns of first (F1) and second (F2) formant frequency change across different vowel categories. The three Australian English (AusE) close front vowels /iː, ɪ, ɪə/ provide a striking example: while their midpoint or mean F1 and F2 frequencies are virtually identical, their spectral change patterns distinctly differ.

The present study utilizes a pre-attentive discrimination paradigm with electroencephalography to assess AusE listeners' perceptual sensitivity to close front vowels with different F1 × F2 trajectory lengths (TLs) and directions (TDs). When TLs are modest, there is an asymmetry in perceptual sensitivity: *closing* vowels, e.g., /iː/ whose trajectory terminates high in the F1 × F2 vowel space, are perceptually prominent, whereas *centering* vowels, e.g., /ɪ, ɪə/ whose trajectories end more centrally, are not. However, when TLs are exaggerated, the asymmetry in the perceptual sensitivity to the two TDs is substantially reduced.

The results indicate that, despite the distinct patterns of spectral change of AusE /iː, ɪ, ɪə/ in production, its perceptual relevance is not uniform, but rather vowel-category dependent.

Index Terms: vowels, pre-attentive discrimination, speech perception, speech acoustics, English dialects

1. Introduction

Research on vowel segments has been guided by a belief that relevant spectral information lies within a relativity "steady-state" portion of the first (F1), second (F2) and third (F3) formants [1]. Peterson and Barney [2] noted some 60 years ago that the location of such a portion is elusive because vowels are rarely spectrally static. More recent acoustic studies have shown English vowels are better separated when dynamic spectral information is present (e.g., [3], [4] and [5]).

With respect to vowel perception, Strange and colleagues (for a review, see [6]) have repeatedly shown that vowel identification in North American English is very accurate when the "steady-state" portions of nominal monophthongs have been removed. In fact, identification accuracy is much lower for vowels which have had their beginning and end portions removed, leaving only their steady-state portions. This evidence suggests that spectrally-dynamic information contained within vowels is highly relevant. Furthermore, English vowels are generally more intelligible when dynamic spectral information is present than when it is removed [3].

Vowel inherent spectral change (VISC) refers to the "relatively slowly varying changes in formant frequencies associated with vowels themselves, even in the absence of consonantal context" [7]. That is, individual vowel categories show distinct patterns of spectral change that distinguish them from other vowel categories, despite other phonetic factors. For instance, the mean or midpoint F2 frequencies of Standard Southern British English (SSBE) /iː/ and /uː/ are quite similar, but /iː/'s F2 frequency increases over time, whereas /uː/'s F2 decreases, regardless of consonantal context [5]. Chládková *et al.* [8] demonstrated this directional difference in F2 clearly influences SSBE listeners' /iː/-/uː/ categorization, confirming spectral change can play a role in perceptual vowel identity.

Although most vowel perception studies report a global improvement in identification or intelligibility when dynamic spectral information is present versus its absence, there is little direct evidence indicating that this is important for vowel *discrimination*. By contrast, there have been many investigations on discrimination involving information on vowels' static spectral properties, fundamental frequency (f0), and duration (e.g., [9], [10] and [11]).

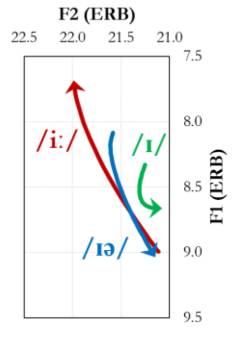

Figure 1: *F1 × F2 trajectories for AusE /iː/, /ɪ/ and /ɪə/ produced male speakers reported in Elvin et al. [4].*

Elvin *et al.* [4] collected a corpus of vowels in the Western Sydney variety of Australian English (AusE). The three close front vowels /iː/, /ɪ/ and /ɪə/ (e.g., in "bead", "bid" and "beard", respectively) are virtually identical in terms of F1 and F2 frequencies averaged across whole formants. The F1 × F2 trajectories for male speakers are displayed in Figure 1. /iː/ has a 1.60 ERB long trajectory proceeding in the direction of a closing diphthong, while /ɪə/'s trajectory length is 1.09 ERB with a direction of a centering diphthong. Finally, /ɪ/'s

10.21437/Interspeech.2018-2505

trajectory direction is centering, but the length is relatively short at 0.52 ERB. Males also produced /iː/, /ɪ/ and /ɪə/ with different durations: 168 ms, 101 ms and 206 ms, respectively.

To assess which combination of acoustic properties best separates AusE /iː/, /ɪ/ and /ɪə/, Elvin *et al.* [4] conducted a classification analysis. A combined measure of F1 × F2 trajectory length with direction as well as vowel duration were the two most effective properties, correctly classifying 77.5% and 70.2% of tokens, respectively. When both properties were entered, 93.1% of the tokens were correctly categorized.

The present study utilizes electroencephalography (EEG) to test the perceptual prominence of spectral change in pre-attentive discrimination. We used an oddball paradigm in which an auditory stimulus (a Standard) is repeated many times, the repetitions of which are infrequently interrupted by physically different stimuli (Deviants). Separately, we also presented listeners with repetitions of each Deviant on its own as a control condition. This particular experimental design has previously been used to examine vowel perception due to vowel category, speaker gender and speaker accent [12]. Evidence confirming pre-attentive change detection comes from a negative difference between the scalp-measured electrophysiological amplitudes from the oddball condition and those from the control condition, which is commonly referred to as a mismatch negativity (MMN) response [13].

Given the previously reported patterns of spectral change for AusE [4], we hypothesize that AusE listeners will show pre-attentive discrimination between a front vowel with a static F1 × F2 trajectory (no spectral change) and vowels containing dynamic F1 × F2 trajectories. As AusE /iː/ is a closing vowel and AusE /ɪə/ is a centering vowel, we predict that these two F1 × F2 trajectory directions will yield similar pre-attentive responses. We also expect that, by exaggerating F1 × F2 trajectory lengths, the perceptual difference between a front vowel with a static F1 × F2 trajectory and one with a dynamic F1 × F2 will be enhanced.

2. Method

2.1. Subjects

11 monolingual speakers of Australian English from Western Sydney who were aged 18-30 years old. None reported hearing impairment or speech and language difficulties.

2.2. Auditory stimuli

The auditory stimuli were created with the Klatt synthesizer [14] in *Praat* [15] based upon average values of the male AusE speakers reported above [4]. The five vowel stimuli were physically identical except for the directions and lengths of their F1 × F2 trajectories, as shown in Table 1. The vowel containing zero spectral change (i.e., flat F1-F4) served as the Standard stimulus and the four vowels containing F1 × F2 spectral change were the four Deviant types. The closing Deviants A and B approximate AusE /iː/, with the latter exhibiting an exaggerated trajectory length, the centering Deviant C approximates /ɪ/ or /ɪə/, while Deviant D approximates an exaggerated instance of /ɪə/.

2.3. Experimental procedure and EEG recording

The experiment consisted of an Oddball condition followed by a Control condition, which each contained four blocks (one for each Deviant type). In the Oddball blocks, the Standard stimulus was played repeatedly and interspersed with repetitions of one of the Deviant types. Across all Oddball blocks, the Standard made up 80% of vowel repetitions and Deviants 20% (= 5% for each type). The Control condition's four blocks each consisted of repetitions of one the four Deviant types on its own (without the Standard). The order of blocks within the two conditions was randomized. The Deviant type of every block was repeated 120 times and the inter-stimulus interval varied randomly between 600-700 ms.

Table 1: *Standard and Deviant type details.*

Stimulus	Duration (ms)	Mean frequency (ERB)	Trajectory direction	Trajectory length (ERB)
Standard		f0 3.80	Static	0.0
Dev. A		F1 8.42	Closing	1.5
Dev. B	129	F2 21.40		3.9
Dev. C		F3 23.72	Centering	1.5
Dev. D		F4 25.55		3.9

The auditory stimuli were played binaurally via in-ear earphones. As pre-attentive sensitivity is automatic and does not require an explicit response, during the experiment subjects watched a subtitled film with the audio muted. The EEG was recorded at a sampling rate of 512 Hz using a 64-channel BioSemi active2 system with electrodes placed according to the international 10/20 placement on a cap fitted to each subject's head size. Six external electrodes were attached to above and below the right eye and on the left and right temples and mastoids.

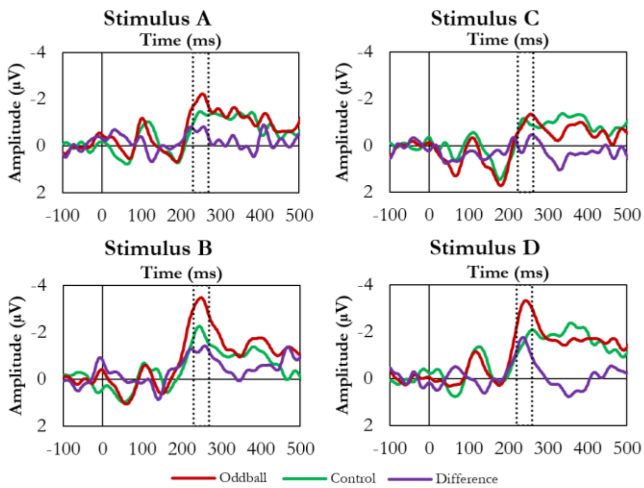

Figure 2: *Grand average waveforms for the four Deviant types in the Oddball and Control conditions at channel FCz. Grand average difference waveforms (MMN) are also shown. The 40 ms window around the peak difference latency is shown as a dotted box.*

2.4. EEG processing

EEG recordings were post-processed with the EEGLAB 13.0 toolbox [16] in MATLAB [17]. Each subject's recording was referenced to their two mastoid channels and was bandpass filtered with a 1 Hz low cut-off of and a 30 Hz high cut-off. Each recording was divided into epochs (time windows) from −100 to 600 ms relative to the onset of the auditory stimulus. Each epoch was baseline-corrected by subtracting the average amplitude of the −100 ms to 0 ms interval from every sample in the rest of the epoch. Eyeblink artefacts were removed by independent component analysis and epochs with amplitudes greater than ±60 μV were rejected (8.97% of all epochs).

We were interested the MMN response for the four Deviant types, i.e., amplitude differences between the Oddball and Control conditions. The following procedure identified the latencies at which this MMN amplitude peaks (cf., [12] and [18]). For each subject, their frontocentral midline electrode (FCz) waveforms were averaged across each block (Figure 2). For each Deviant type, we subtracted each subject's average waveform in the Control condition from that in the Oddball condition, resulting in grand average difference waveforms (Figure 2). Since MMN amplitudes typically peak between 150-250 ms, for each Deviant type the peak's beginning was defined as the first negative excursion of the grand average difference waveform 150 ms after stimulus onset and the peak's end was the first positive excursion at least 40 ms after this. The middle latency in this window served as the *peak difference latency*. For all Deviant epochs, i.e., trials on which Deviant A, B, C or D was played, we subsequently centered a 40 ms window around the peak difference latency of the corresponding Deviant type and averaged over the amplitudes within this small window. This was done for the F3, FC3, C3, Fz, FCz, Cz, F4, FC4 and C4 electrode sites. The bottom two rows of Table 2 show how each of these electrode sites relate to scalp anteriority and laterality.

3. Results

As seen in Figure 2, Deviant types B and D with exaggerated F1 × F2 trajectory lengths evoked larger negative amplitudes than Deviants A and C approximately 200-350 ms after stimulus onset – indicative of the N2 [19]. Contrary to our expectations, it appears there are differences between the closing Deviant types (A and B) and their centering counterparts (C and D). If listeners detected a change from the Standard stimulus (a vowel with a static F1 × F2 trajectory) to the Deviant types (vowels containing dynamic F1 × F2 trajectories), reliable differences between the Oddball and Control conditions, i.e., MMN responses, will emerge according to F1 × F2 trajectory directions and lengths.

Table 2: *Numeric coding for the fixed factors in the mixed-effects regression model.*

Fixed factor	Level	Coding
Condition	Control	−0.5
(Cond)	Oddball	0.5
Trajectory Direction	Centering	−0.5
(TD)	Closing	0.5
Trajectory Length	1.5 ERB	−0.5
(TL)	3.9 ERB	0.5
Anteriority	Central (C)	−1
(Ant)	Frontocentral (FC)	0
	Front (F)	1
Laterality	Right (4)	−1
(Lat)	Midline (z)	0
	Left (3)	1

Statistical analyses of MMN effects are often performed on the differences between each subject's *mean* amplitudes from the Oddball and Control conditions per Deviant type (e.g., [12], [18]). Limitations of prior averaging include the inability to consider the variance within each subject's pool of epochs [20] and across randomly-assigned blocks [21], and the assumption that experimental manipulations have uniform effects across subjects [22]. For these reasons, a linear mixed-effects regression model was fitted to amplitudes from around the peak difference latency from *all* 9,623 Deviant epochs (with the *lme4* package [23] *R* [24]). Five fixed factors, as shown in Table 2, and their interactions were entered. Fixed factor levels were numerically coded and centered on zero, meaning that the model's intercept represents a grand mean amplitude (μV) and the fixed effects are interpretable as main effects. Random by-subject intercepts and slopes were included for all fixed factors and by-block intercepts and slopes were included for Cond, Ant and Lat, as these were repeated across blocks within each condition [22].

Table 3: *Model results for the fixed factors of Condition (Cond), Trajectory Direction (TD), Trajectory Length (TL), Anteriority (Ant) and Laterality (Lat). * indicates significant at α = 0.05.*

Fixed effects	*z*	SE	*t*	*p*
Intercept	−1.63	0.60	−2.71	0.018*
Cond	−0.68	0.56	−1.21	0.273
TD	−0.48	0.29	−1.63	0.133
TL	−0.97	0.35	−2.80	0.019*
Ant	−0.18	0.07	−2.67	0.023*
Lat	−0.02	0.08	−0.28	0.789
Cond×TD	−0.73	0.18	−4.10	< 0.001*
Cond×TL	−1.04	0.16	−6.50	< 0.001*
Cond×Ant	0.01	0.09	0.14	0.885
Cond×Lat	−0.02	0.09	−0.24	0.809
TD×TL	−0.20	0.16	−1.25	0.210
TD×Ant	−0.05	0.09	−0.55	0.585
TD×Lat	−0.06	0.09	−0.66	0.509
TL×Ant	−0.14	0.09	−1.44	0.150
TL×Lat	0.04	0.09	0.48	0.630
Ant×Lat	0.07	0.06	1.23	0.220
Cond×TD×TL	0.68	0.32	2.13	0.033*
Cond×TD×Ant	0.12	0.19	0.61	0.540
Cond×TD×Lat	0.17	0.19	0.90	0.366
Cond×TL×Ant	0.01	0.19	−0.03	0.977
Cond×TL×Lat	−0.07	0.19	−0.35	0.725
Cond×Ant×Lat	0.04	0.12	0.35	0.729
TD×TL×Ant	0.02	0.19	0.12	0.901
TD×TL×Lat	−0.11	0.19	−0.58	0.565
TD×Ant×Lat	0.00	0.12	−0.03	0.977
TL×Ant×Lat	0.05	0.12	0.43	0.667
Cond×TD×TL×Ant	0.16	0.38	0.42	0.677
Cond×TD×TL×Lat	−0.17	0.38	−0.46	0.644
Cond×TD×Ant×Lat	−0.01	0.23	−0.06	0.951
Cond×TL×Ant×Lat	−0.05	0.23	−0.23	0.820
TD×TL×Ant×Lat	−0.03	0.23	−0.14	0.890
Cond×TD×TL×Ant×Lat	0.06	0.46	0.14	0.890

Table 3 displays the results. Satterthwaite approximations were used in the *t*-tests. As the intercept is (expectedly) a negative value, positive coefficients indicate the respective fixed factor shifts amplitudes closer to zero, whereas negative coefficients suggest greater negativity. The main effect of Cond, which indicates the mean MMN response (amplitude difference between the Oddball and Control conditions) over all four Deviant types, is unreliable. For TD, closing Deviant types evoked a −0.48 μV greater amplitude than centering

Deviant types on average, but this difference is also unreliable. For TL, amplitudes from 3.9 ERB Deviant types were on average −0.97 µV greater than from 1.5 ERB Deviant types. The significant effect of Ant suggests frontal electrodes registered slightly higher negative amplitudes than more central electrodes, which is unsurprising because pre-attentional perception is correlated with greater negative anterior neurophysiological activity [19]. As no reliable interactions involve Ant or Lat, the effects of Deviant type or MMN response were not specific to electrode sites.

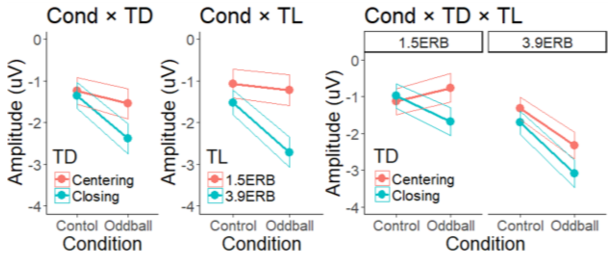

Figure 3: *Model estimates of interactions involving Cond, TD and TL. Error bars are standard errors. A negative slope between the Control and Oddball conditions suggests a MMN response.*

The model's three reliable interactions all involve Cond, TD and TL, as plotted in Figure 3, suggesting MMN amplitudes (differences between the Oddball and Control conditions) depend on the properties of the Deviant type. For the Cond×TD interaction, contrary to our expectation of no MMN difference between TDs, there was a much smaller MMN amplitude for centering Deviant types (ΔM −0.31 µV; CI −0.47, −0.15) than for closing Deviant types (ΔM −1.04 µV; CI −1.25, −0.83), suggesting closing vowels are perceptually more prominent. As for the Cond×TL interaction, the MMN was much stronger when TL is 3.9 ERB (ΔM −1.19 µV; CI −1.44, −0.95) than when it is 1.5 ERB (ΔM −0.15; CI −0.29, −0.02), indicating exaggerated F1 × F2 trajectory lengths are perceptually more prominent. The model's lack of reliable main effects for Cond and TD were therefore due to crossover effects, i.e., when the direction of an effect is observed in one condition but not in the other. The Cond×TD×TL interaction reveals that Deviant C (centering, 1.5 ERB) did *not* evoke a MMN response (ΔM 0.38 µV; CI 0.27, 0.48), suggesting centering front vowels with moderate F1 × F2 trajectory lengths are not discriminable pre-attentively from front vowels without spectral change. This was not the case for Deviant types A (ΔM −0.69 µV; CI −0.85, −0.54), B (ΔM −1.39 µV; CI −1.65, −1.13) and D (ΔM −1.00 µV; CI −1.22, −0.78) where MMN responses are clearly evident.

4. Discussion

We investigated AusE listeners' perceptual sensitivity to spectral change in close front vowels by using a pre-attentive discrimination paradigm. It was found that when the F1 × F2 trajectory length of a close front vowel is 1.5 ERB, there was an asymmetry in perceptual sensitivity which was dependent on the trajectory's direction: when the F1 × F2 trajectory approximated a *closing* vowel, e.g., /iː/, AusE listeners detected a difference between this vowel and an identical vowel without spectral change; conversely, when the F1 × F2 trajectory approximated a *centering* vowel, e.g., /ɪə/ or /ɪ/, the change was not reliably detected. This was contrary to our prediction that a difference would be detected regardless of the vowel's F1 × F2 trajectory direction, since this is an acoustic difference which distinguishes AusE close front vowels in speech production (cf., Figure 1). When the F1 × F2 trajectory length was exaggerated (3.9 ERB), no such asymmetry in pre-attentive sensitivity was found, as both closing and centering vowels were strongly detected as different from an identical vowel without spectral change (Figure 3).

We suspect that AusE listeners failed to detect a change from a vowel without spectral change to a vowel with a 1.5 ERB but closing F1 × F2 trajectory because all four Deviant types had the same durations. It has been reported that AusE /ɪə/ has variants pronounced more like [ɪː], possibly due to sociophonetic reasons, e.g., gender or sound change [25]. It may be that, for /ɪə/ and /ɪ/, spectral change is less relevant because they are primarily distinguished from one another perceptually by duration, i.e., /ɪə/ is long and /ɪ/ is short. Without the availability of durational differences, a short [ɪə]-like vowel may therefore be perceived as /ɪ/, or conversely, a long [ɪ]-like vowel may be perceived as /ɪə/. Only when F1 × F2 trajectory length is strongly exaggerated do spectral change differences become perceptually prominent for centering vowels. On the other hand, a closing F1 × F2 trajectory is distinguished from a vowel without spectral change, regardless of whether its trajectory length is exaggerated or not, suggesting that this property of AusE /iː/ is perceptually very prominent. Further testing with a wider range of stimuli that also vary in duration will confirm whether this is the case.

Finally, a theory on vowel perception may shed some light on the present results. Polka and Bohn [26]'s Natural Referent Vowel (NRV) framework posits that there is a language-universal perceptual bias favoring vowels that fall closer to the periphery of the F1 × F2 vowel space because they act as "reference points" for vowels within this space. Although the NRV framework does not explicitly consider the spectrally dynamic properties of vowels, it may still be relevant. For instance, the F1 × F2 trajectory of AusE /iː/ begins more central and moves towards the periphery of the F1 × F2 vowel space, whereas the F1 × F2 trajectories of /ɪ/ and /ɪə/ begin more peripheral and move towards the center (Figure 1). If spectral differences between the onsets and offsets of individual vowels are assumed to be important within the NRV framework, then the bias towards peripheral vowels would mean closing vowels are perceptually more prominent than centering vowels, explaining some of the present results (for further discussion, see [27] and [28]).

5. Conclusion

Dynamic spectral information is clearly a pertinent property in vowel discrimination, as demonstrated by AusE listeners' pre-attentive sensitivity to it in close front vowels. Importantly, this sensitivity is not uniform: spectral change, as in AusE /iː/, is perceptually more prominent than that in AusE /ɪə/ or /ɪ/. Thus, the perceptual relevance of spectral change is likely to be vowel-category dependent because other acoustic information, such as duration, also aids discrimination.

6. Acknowledgements

This research was supported by an Endeavour Research Fellowship (4699_2015) and the ARC Centre of Excellence for the Dynamics of Language (CE140100041). We are grateful to Alba Tuninetti and research assistants at the MARCS Institute for experiment coordination.

7. References

[1] J. M. Hillenbrand, "Static and dynamic approaches to vowel perception," in G. S. Morrison and P. F. Assmann (eds), *Vowel Inherent Spectral Change*, Berlin-Heidelberg: Springer, 2013, pp. 9–30.

[2] G. E. Peterson and H. L. Barney, "Control methods used in a study of the vowels," *Journal of the Acoustical Society of America*, vol. 24, pp. 175–184, 1952.

[3] J. M. Hillenbrand and T. M. Nearey, "Identification of resynthesized /hVd/ utterances: effects of formant contour," *Journal of the Acoustical Society of America*, vol. 105, pp. 3509–3523, 1999.

[4] J. Elvin, D. Williams, and P. Escudero, "Dynamic acoustic properties of monophthongs and diphthongs in Western Sydney Australian English," *Journal of the Acoustical Society of America*, vol. 140, pp. 576–581, 2016.

[5] D. Williams and P. Escudero, "A cross-dialectal acoustic comparison of vowels in Northern and Southern British English," *Journal of the Acoustical Society of America*, vol. 136, pp. 2751–2761, 2014.

[6] W. Strange and J. J. Jenkins, "Dynamic specification of coarticulated vowels," in G. S. Morrison and P. F. Assmann (eds), *Vowel Inherent Spectral Change*, Berlin-Heidelberg: Springer, 2013, pp. 87–115.

[7] T. M. Nearey and P. F. Assmann, "Modeling the role of inherent spectral change in vowel identification," *Journal of the Acoustical Society of America*, vol. 80, pp. 1297–1308, 1986

[8] K. Chládková, S. Hamann, D. Williams, and S. Hellmuth, "F2 slope as a perceptual cue for the front-back contrast in Standard Southern British English," *Language and Speech*, vol. 60, pp. 377–398, 2017.

[9] B. H. Repp, A. F. Healy, and R. G. Crowder, "Categories and context in the perception of isolated steady-state vowels," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 5, pp. 129–145, 1979.

[10] P. A. Hallé, Y.-C. Chang, and C. T. Best, "Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners," *Journal of Phonetics*, vol. 32, pp. 395–421, 2004.

[11] M. Hisagi, V. L. Shafer, W. Strange, and E. S. Sussman, "Perception of a Japanese vowel length contrast by Japanese and American English listeners: Behavioral and electrophysiological measures," *Brain Research*, vol. 1360, pp. 89-105, 2010.

[12] R. Dadwani, V. Peter, K. Chládková, A. Geambașu, and P. Escudero, "Adult listeners' processing of indexical versus linguistic differences in a pre-attentive discrimination paradigm," in The Scottish Consortium for ICPhS 2015 (ed.) *Proceedings of the 18th International Congress of Phonetic Sciences August 10–14*, Glasgow: University of Glasgow, 2015.

[13] R. Näätänen, A. Lehtokoski, M. Lennes, M. Cheour, M. Huotilainen, A. Iivonen, M. Vainio, P. Alku, R. J. Ilmoniemi, A. Luuk, J. Allik, J. Sinkkonen, and K. Alho, "Language-specific phoneme representations revealed by electric and magnetic brain responses," *Nature*, vol. 385, pp. 432–434, 1997.

[14] D. H. Klatt and L. C. Klatt, "Analysis, synthesis and perception of voice quality variations among male and female talkers," *Journal of the Acoustical Society of America*, vol. 87, pp. 820–856, 1990.

[15] P. Boersma and D. Weenink, *Praat: Doing phonetics by computer*, Version 6.0.13, http://www.praat.org/, 2016.

[16] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics," *Journal of Neuroscience Methods*, vol. 134, pp. 9–21, 2004.

[17] *MATLAB and Statistics Toolbox*, Release 2016a, Natick, MA: The MathWorks, Inc., 2016.

[18] K. Chládková, P. Escudero, and S. C. Lipski, "When "AA" is long but "A" is not short: speakers who distinguish short and long vowels in production do not necessarily encode a short-long contrast in their phonological lexicon," *Frontiers in Psychology*, vol. 6, #438, 2015.

[19] J. R. Folstein and P. C. Van Petten, "Influence of cognitive control and mismatch on the N2 component of the ERP: a review," *Psychophysiology*, vol. 45, pp., 152–170, 2008.

[20] D. V. M. Bishop and M. J. Hardiman, "Measurement of mismatch negativity in individuals: A study using single-trial analysis," Psychophysiology, vol. 47, pp. 697–705, 2010.

[21] R. Frömer, M. Maier, and R. A. Rahman, "Group-level EEG-processing pipeline for single trial-based analyses including linear mixed models," *Frontiers in Neuroscience*, vol. 12, #48, 2018.

[22] D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily, "Random effects structure for confirmatory hypothesis testing: Keep it maximal," *Journal of Memory and Language*, vol. 68, pp. 255–278, 2014.

[23] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models Using *lme4*," *Journal of Statistical Software*, vol. 67, pp. 1–48, 2015.

[24] R Core Team, *R: A Language and Environment for Statistical Computing*, Version 3.3.3, Vienna: R Foundation for Statistical Computing, https://www.R-project.org/, 2017.

[25] F. Cox, "/hVd/ vowels in the speech of some Australian teenagers," *Australian Journal of Linguistics*, vol. 26, pp. 147–179, 2006.

[26] L. Polka and O.-S. Bohn, O.-S. "Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development," *Journal of Phonetics*, vol. 39, pp. 467–478, 2011.

[27] P. Escudero, K. E. Mulak, J. Elvin, and N. M. Traynor, "'Mummy, keep it steady': phonetic variation shapes word learning at 15 and 17 months," *Development Science*, #e1260, 2017.

[28] B. P. Kriengwatana and P. Escudero, "Directional asymmetries in vowel perception of adult nonnative listeners do not change over time with language experience," *Journal of Speech, Language, and Hearing Research*, vol. 60, pp. 1088–1093, 2017.