Acoustic cue variability affects eye movement behaviour during non-native speech perception

Jessie S. Nixon[1]*, Catherine T. Best*[2]

[1]University of Tübingen, Germany
[2]The MARCS Institute, Western Sydney University, Australia
jessie.nixon@uni-tuebingen.de, C.Best@westernsydney.edu.au

Abstract

A fundamental question in speech research is how listeners use continuous (non-discrete) acoustic cues to discriminate between discrete alternative messages. An important factor is the statistical distribution of acoustic cues in speech. Previous research has shown that when native speakers listen to speech with high within-category variability in the discriminative cue dimension, perceptual uncertainty increases, resulting in increased looks to competitor objects. The present study investigated effects of within-category acoustic variability on eye movements during acquisition of a non-native acoustic dimension, namely English speakers acquisition of lexical tone.

All participants heard a bimodal distribution of stimuli, with distribution peaks at the prototypical pitch values for Cantonese high and mid level tones; however, presentation frequency differed between conditions: high-variance vs. low-variance. Based on previous research, we expected lower uncertainty and better learning in the low-variance condition.

GAMM models showed that towards the end of the experiment, fixations were closer to the target object in the low-variance, compared to the high-variance condition. This suggests that within-category acoustic variability not only increases uncertainty for native listeners, but may also initially hinder learning of acoustic cues during non-native language acquisition.

Index Terms: speech perception, statistical learning, second language acquisition, Cantonese lexical tone, visual world eyetracking, generalised additive mixed models (GAMMs)

1. Introduction

The organisation of acoustic cues varies substantially across languages. Cue dimensions that are lexically contrastive in one language may not be contrastive in another. Therefore, acquisition of a new language often involves learning to substantially adjust cue weights (i.e. to adjust the degree to which various cues in the signal are utilised, consciously or unconsciously) for lexical contrasts. In some cases, this can pose significant challenges. Expert knowledge of statistical regularities in one's native language can lead to expectations that hinder non-native speech perception [1, 2]. Statistical properties that seem to play a role in shaping cue perception include the number of distribution peaks along a cue dimension [3, 4, 5], acoustic distance between peaks in a bimodal distribution [6] and within-category acoustic variance [7, 8, 9].

Many recent studies have emphasised the role of variability in shaping and reshaping native and non-native sound systems. For example, early first language acquisition [10] and second language acquisition [11, 12] seem to benefit from multiple speakers, compared to a single speaker in the training input. When there are multiple speakers, this increases variability in *non-contrastive* indexical dimensions, which seems to have the effect of highlighting the relative invariance of the contrastive dimensions. This is consistent with learning models, which posit that learning not only involves acquisition of knowledge, but also learning to ignore irrelevant cues [13, 14]. Interestingly, one recent study suggests that exposure to a non-native discriminative dimension improves not only discrimination between categories, but also perception of within category acoustic information within that dimension [5].

An aspect that has received less attention is variability *within* contrastive dimensions. Acoustic studies suggest that this variability may be an important factor in adjusting cue weights for discriminating sound contrasts. For example, for native English speakers, the third formant (F3) is generally the most reliable cue to the /l/ and /r/ contrast [15]. F3 values cluster around the /l/ and /r/ productions with relatively little spread or overlap. While there is some difference in the distribution of the second formant (F2) between /l/ and /r/, it is highly variable, with a high degree of overlap [16]. This variance means that, while F2 seems to play a role as a secondary cue, it is not as reliable as F3 and is not relied on as much by native English listeners. In addition, in recordings of native Japanese productions of English /l/ and /r/ [16], the F3 values are highly variable and largely overlapping for /l/ and /r/ productions. This increased variance and category overlap corresponds to reduced effectiveness of the cue for discriminating between the /l/ and /r/ tokens produced by these speakers.

Nixon and colleagues [8] investigated the temporal dynamics of perceptual uncertainty during native speech perception using the'visual world' eyetracking paradigm. In this paradigm, participants see four pictures on the screen, hear a word and are instructed to click on the picture corresponding to the word. Effects emerged very early, in the first fixations of the trial. As variability increased and speech cues became less reliable, listeners looked around more, presumably in search of further support for partially activated candidates. The idea that listeners were seeking additional evidence in the high-variability condition seems to suggest the appropriate conditions for adjusting cue weights, and perhaps increasing weights of previously downweighted cues.

What is not yet known is whether such within-category acoustic variance also affects acquisition of a new acoustic dimension in a non-native language. The present study addressed this question by examining the effect of within-category acoustic variance on eye movements during native English speakers' acquisition of a pitch cue (fundamental frequency; *f0*) in a Cantonese lexical tone contrast. English does not use f0 as a lexical contrast, and tone can be notoriously difficult for beginning learners of non-tonal languages [17]. Based on previous studies [7, 8, 9], we expected greater weighting of the pitch cue over the course of the experiment - that is, better learning - in the

low-variance, compared to the high-variance condition.

2. Method

2.1. Participants

Thirty-seven native English-speaking students from the University of Western Sydney who had not previously studied any tone language were recruited for the experiment for course credit[1]. Participants were tested individually in a sound-attenuated booth.

2.2. Experiment design and stimuli

Visual stimuli were black-on-white line drawings of eight common objects. Auditory stimuli were four minimal pairs of single-character mid- and high-tone words (e.g. *gun_mid* 'can' and *gun_high* 'crown'). All auditory stimuli were produced by a male native speaker of Hong Kong Cantonese. Stimuli were then resynthesised into a 14-step pitch continuum (e.g. *gun_mid* to *gun_high*) using PRAAT [18]. One half of the continuum corresponded to the mid tone and one half to the high tone.

The number of times participants heard each token of the continuum followed a bimodal distribution, with the two peaks of the distribution corresponding to the prototypical f0 for the mid- and high-tone stimuli, respectively. All participants heard the same number of tokens; but the number of times they heard each token differed between conditions, with greater spread from the mean (statistical variance) in the *high-variance* versus the *low-variance* distribution (see Figure 1). The experiment consisted of 240 experimental trials, divided into six blocks of 40 trials, with breaks between the blocks. The order of presentation was pseudo-randomised for each participant.

2.3. Procedure

Participants sat 60 cm from a computer screen equipped with an SR Research Eyelink 1000 remote eyetracker with a chinrest and headrest. Stimulus presentation and data acquisition were conducted using SR Research Experiment Builder with a sampling rate of 1000 Hz. The session began with ten practice trials. None of the images or auditory stimuli from the experimental block appeared in the practice block. Each experimental trial began with a brief (1000 ms) presentation of four pictures, one in each quadrant of the screen. The purpose of the preview was to reduce noise in the data by reducing the time and likelihood of participants scanning the images at the beginning of the trial. The display always contained a target, a competitor and two distractor items. The target and competitor had the same segmental syllable, but differed in tone. The location of each picture condition on the screen and their location relative to each other were randomised to avoid strategic effects. The preview disappeared, followed by a gaze-contingent fixation cross to ensure participants were fixating the centre of the screen at the beginning of the critical trial period. The pictures then reappeared simultaneously with presentation of the auditory stimulus. Participants were instructed to select the picture corresponding to the word they heard by clicking on it with the mouse, and to guess if they did not know. They were given feedback ('correct'/'wrong') after each trial. Participants were told that this was a language-learning task, but were not informed about the pitch or tone manipulation or the target language.

[1] Participants were not explicitly asked whether they had studied a tone language, as this might influence the experiment results. Instead, they were asked to list all languages they spoke or had studied, and were screened if they did not meet this criterion.

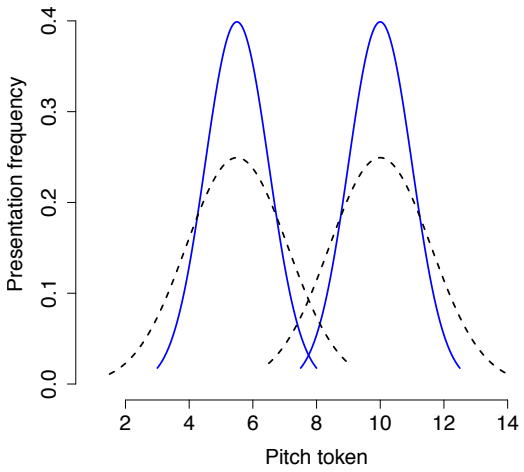

Figure 1: *Illustration of the presentation frequency distributions in the high-variance (black lines) and low-variance conditions (blue lines).*

3. Analysis

Eye movement data were analysed using *Generalised Additive Mixed Models* [19, 20, GAMM] using the `mgcv` package (version 1.8.17) in R [21, version 3.4.0]. Generalised Additive Models (GAMs) are a type of Generalised Linear Model that use smooth functions to model nonlinear effects of continuous predictors. The 'mixed' in GAMMs refers to the inclusion of random effects in addition to fixed effects.

GAMM is a well-established method of analysis that is increasingly being used in the cognitive and language sciences [22, 23, 24, 25], articulography [26, 27], acoustic analysis [28], temporal clustering of sociolinguistic variants [29] and dialectology [30]. Recently, it has also been applied to visual world paradigm eye movement data [8, 9, 31] and pupilometry [32].

GAMMs are a valuable method for analysing visual world fixation data for several reasons, including their ability to capture nonlinear changes in eye movements over the course of the trial and/or over the course of the experiment, the inclusion of random effects to deal with taking repeated measures from the same participants and items, and methods for dealing with autocorrelation [8]. An important aspect of eyetracking data is how fixations change over time. In experimental data sets, and especially time series data, autocorrelation can occur between data points [33]. In the `mgcv` package, functions have been implemented to deal with autocorrelation in GAM models.

Eye movement data were modelled as a continuous predictor of Euclidean distance of fixations from the centre of the target image. Because this gradient measure of distance includes data points that have not reached the target image or fall between images, it is more likely to pick up on uncertainty effects, such as undershooting, hesitant or inaccurate oculomotor movements due to low activation or competing activations, compared to a categorical measure of within or outside the interest area [8]. All predictors of interest were entered into the initial model, and predictors that did not contribute to model fit were removed. Model comparison was conducted by means of χ^2 tests of fREML scores, using the `compareML` function in the `itsadug` package [34, version 2.2] in R. Because we were

interested in the time course of fixations over the course of the trial, a continuous predictor of *time* was included. Data were downsampled to 50 Hz to reduce autocorrelation between data points. A 3200 ms window was selected for analysis, from 200 ms prior to to 3000 ms after auditory stimulus presentation. To test whether there was a learning effect over the course of the experiment, the model included a continuous predictor of trial, centred around 0 (*centred trial*). To determine whether participants were using pitch as a cue to distinguish between target and competitor images, the model included a continuous predictor of pitch, also centred around zero (*centred pitch*). The centred values ranged from -5.5 to 5.5, with the distribution peaks at -3 and 3. Distribution variance was modelled as a two-level factor, low-variance and high-variance. Previous research with the visual world paradigm has shown that the location of the target object on the screen significantly affects eye movement behaviour [35, 8]. Therefore, a smooth for *target position* over time was included as a control variable, a factor with four levels: top-left, top- right, bottom-left and bottom-right. A random smooth for subject by item over time was included to account for differences in individual participants and items.

The initial model included intercepts for the two factor variables, variance condition and target position, and smooths (for each of the main effects) and nonlinear regression lines[2] (for each two- and three-way interaction) for each level of condition. A smooth was also included for each level of target position. Random effects were modelled with shrunk factor smooths. After running the model, the model residuals were examined to check for autocorrelation. An AR(1)[3] model was included to account for autocorrelation in the residuals.

4. Results

Model comparisons showed that model fit was improved by including smooths for centred trial by condition ($p < .001$); target position over time ($p < .001$); and nonlinear regression smooths for time by trial ($p < .01$); trial by pitch by condition ($p < .001$); time by pitch by condition ($p < .01$); trial by time by pitch ($p < .001$); and trial by time by pitch by condition ($p < .001$).

The difference between the high- and low-variance conditions over the course of the experiment is shown in Figure 2. Where the line is above zero, the eyes were closer to the target in the low-variance condition than in the high-variance condition. The vertical dotted red lines indicate areas of significant difference between conditions. The effect emerged later (2800ms) and more distally for the high tone (positive centred pitch values 4.5 and 3.5) than the low tone (2400ms; negative centred pitch values -3.5, -2.5, -1.5). At the beginning of the experiment, the distance from the target is greater in the low-variance condition than the high-variance condition for some pitch values. However, as the experiment progresses, the distance gets smaller, and the distance from the target becomes significantly smaller in the low-variance condition. This effect emerges late for the high tone and just after halfway through for the mid tone.

The summed effects are shown in Figure 3. The figure shows a topographic plot of the Euclidean distance of fixations from the target object in the low-variance (left panels) and high-variance conditions (right panels). Time (ms) is on the horizontal axis. Centred pitch is on the vertical axis: positive pitch values correspond to the high tone and negative values to the mid tone. Distance from the target (in pixels) is on the z-axis and is colour-coded. Higher values (warmer colours) indicate fixations were further from the target image; lower values (cooler colours) indicate fixations were closer to the target image. The key in the top right corner indicates the corresponding values and z-limits. The panel rows show snapshots of trials throughout the experiment.

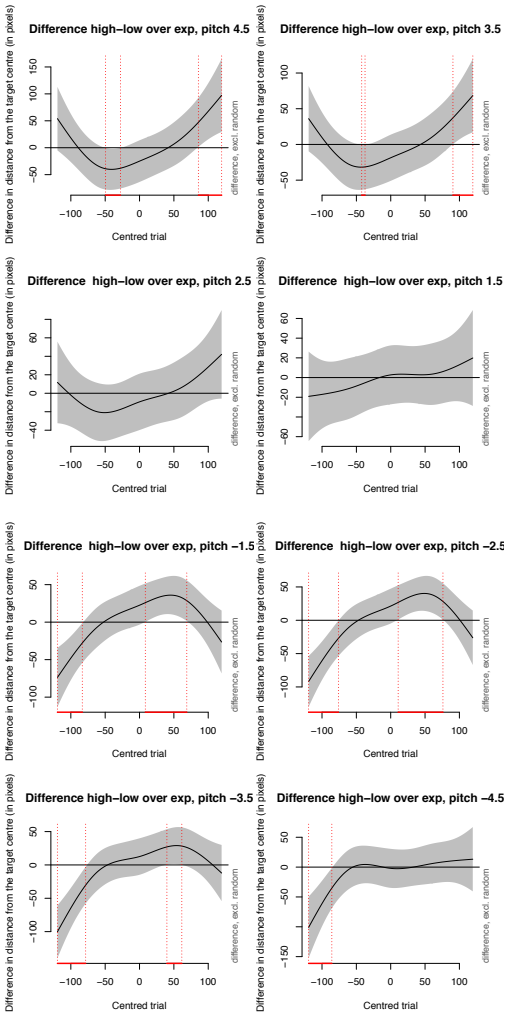

Figure 2: *Smooth of the difference between the high-variance and low-variance conditions over the course of the experiment for high tone (top 4 panels) and mid tone (bottom 4 panels). Vertical dotted lines indicate significant difference. Trial is on the x-axis and is centred around 0. The difference (in pixels) between conditions (high minus low) is on the y-axis. Time is set to the peak of the effect for high (2800 ms) and mid tones (2400ms). Random effects are removed from these plots.*

At the beginning of the experiment, fixations are further from the target in the low-variance condition, compared to the high-variance condition. This effect lessens over the following trials, and has disappeared by trial 100. For the remainder of the experiment, fixations become gradually closer to the target in the low-variance condition. This becomes significant earlier in the low pitch values, as seen in trial 170. Over time, fixations become significantly closer to the target in the low-

[2]The partial effects tensors are modelled with the ti() function in the `mgcv` package.

[3]For data points in a series (in this case trials), the AR(1) is a measure of the current error as a proportion of the preceding error (plus Gaussian noise).

variance condition for both the low and high pitch values.

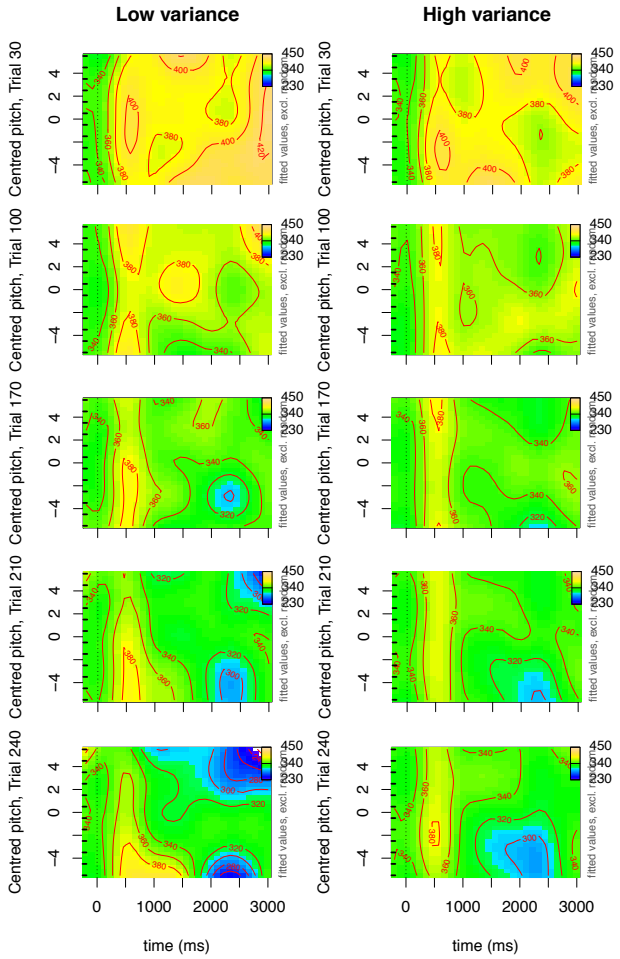

Figure 3: *Topographic maps of the model estimates for Euclidean distance from the target object in the low-variance (left panels) and high-variance conditions (right panels). Time (ms) is represented on the x-axis. Pitch is on the y-axis. Pitch is centred around 0, the category boundary. Positive values correspond the the high tone, negative values to the mid tone. Distribution peaks were at 3 and -3. Distance of fixations from the target object is plotted on the z-axis and is colour coded. Higher values (warmer colours) indicate greater distance; lower values (cooler colours) indicate a smaller distance. The key in the top-left corner shows the corresponding distance (in pixels) and z-limits. Random effects are excluded from this plot.*

5. Discussion

The present study investigated the effects of within-category acoustic variance on non-native acquisition of a new acoustic cue dimension, that is, pitch (f0) in a lexical tone contrast. Participants saw pictures of common objects and heard minimal word pairs, differing only in lexical tone. The tones were based on two Cantonese level tones, which are distinguished by pitch height. Auditory stimuli were sampled from pitch continua corresponding to the words. Stimuli were sampled according to a bimodal distribution. The critical manipulation was the statistical variance of the distribution, i.e. the amount of acoustic variability within the critical contrastive dimension, pitch. Participants heard either a *high-* or a *low-variance* distribution. Based on literature investigating effects of variance in native speech processing [7, 8], we predicted that acquisition of the pitch cue would be better in the low-variance condition. GAMM models of eye movements showed that the Euclidean distance between fixations and the centre of the target picture reduced over the course of the experiment in both conditions. In addition, by the end of the experiment, distance was lower in the low-variance condition, compared to the high-variance condition. Interestingly, the effect seemed to emerge earlier in the mid tone than the high tone, in terms of both the time point in the trial and the trial in the experiment. This suggests the participants begin to associate the mid tone with its target picture more quickly, perhaps because it is closer to the English prosodic range.

The present results provide new evidence that within-category acoustic variance shapes nonnative acoustic cue acquisition. Previous studies have shown that acoustic variance affects native speech perception, with increased variance leading to increased perceptual uncertainty [7, 8]. The present study shows that the same mechanism can also help shape acquisition of a new acoustic dimension not present as a lexical contrast in the native language.

Cue variance has been investigated previously in native Japanese listeners' learning of the English /l/-/r/ contrast [36]. Many native Japanese listeners have trouble attending to the third formant (F3) cue - which native English listeners tend to use to distinguish /l/ and /r/ - and rely instead on the less reliable second formant (F2). Using video game training over several days, Lim and Holt found that by presenting stimuli with high variability in the F2 dimension and low variability in the F3 dimension, participants' cue weighting shifted towards F3 and categorisation accuracy significantly increased. While this innovative study demonstrates the potential of variability to adjust cue weighting, it differs from the present study in several respects. Firstly, the present study directly compared effects of high- vs. low-variance; the Lim and Holt study compared effects of training to a control condition that did not involve English exposure. Secondly, participants in the Lim and Holt study were proficient English speakers. They had been studying English for at least 12 years and had lived in an English-speaking environment for up to 2.5 years. The present study investigated acquisition of a new cue dimension, not encountered before in a lexical contrast. Rather than improving an already partially acquired contrast, participants in the present study were experiencing both the language and the tonal contrast for the first time. Thirdly, Lim and Holt used flat distributions - four steps of F2 and two steps of F3 - presented at equal frequency, whereas the present study used approximately Gaussian distributions. Therefore the present study makes an important contribution by directly testing effects of the degree of distributional variance on acquisition of a new acoustic dimension.

While several recent studies have emphasised the facilitative effect of variability on learning [11, 12, 10], it is important to distinguish between within-category acoustic variance in the critical dimension and variability in non-contrastive dimensions. Variability can lower cue weighting. If that variability is in a contrastive dimension, it may hinder discrimination.

6. Acknowledgements

This research was made possible by an Endeavour Research Fellowship (Grant Number ERF RDDH 114001) to the first author.

7. References

[1] C. T. Best, "A direct realist view of cross-language speech perception," in *Speech perception and linguistic experience: Issues in cross-language research*, W. Strange, Ed. Timonium, MD: York Press, 1995, pp. 171–204.

[2] J. E. Flege, "Second language speech learning: Theory, findings, and problems," *Speech perception and linguistic experience: Issues in cross-language research*, pp. 233–277, 1995.

[3] J. Maye and L. Gerken, *Learning phonemes without minimal pairs*. Proceedings of the 24th Annual Boston University Conference on Language Development, 2000.

[4] K. Wanrooij, P. Boersma, and T. L. van Zuijen, "Fast phonetic learning occurs already in 2-to-3-month old infants: an ERP study," *Frontiers in Psychology*, vol. 5., 2014.

[5] J. S. Nixon, N. Boll-Avetisyan, T. O. Lentz, S. van Ommen, B. Keij, Ç. Çöltekin, L. Liu, and J. van Rij, "Short-term exposure enhances perception of both between- and within-category acoustic information," in *Proceedings of the 9th International Conference on Speech Prosody 2018*, June in press.

[6] P. Escudero, T. Benders, and K. Wanrooij, "Enhanced bimodal distributions facilitate the learning of second language vowels," *The Journal of the Acoustical Society of America*, vol. 130, no. 4, 2011.

[7] M. Clayards, M. K. Tanenhaus, R. N. Aslin, and R. A. Jacobs, "Perception of speech reflects optimal use of probabilistic speech cues," *Cognition*, vol. 108, no. 3, pp. 804–809, 2008.

[8] J. S. Nixon, J. van Rij, P. Mok, R. H. Baayen, and Y. Chen, "The temporal dynamics of perceptual uncertainty: eye movement evidence from Cantonese segment and tone perception," *Journal of Memory and Language*, vol. 90, pp. 103–125, 2016.

[9] ——, "Eye movements reflect acoustic cue informativity and statistical noise," in *ExLing 2015: Proceedings of the International Conference of Experimental Lingustics*, A. Botonis, Ed., 2015, pp. 54–57.

[10] G. C. Rost and B. McMurray, "Speaker variability augments phonological processing in early word learning," *Developmental Science*, vol. 12, no. 2, pp. 339–349, 2009.

[11] J. S. Logan, S. E. Lively, and D. B. Pisoni, "Training Japanese listeners to identify English/r/and/l: A first report," *The Journal of the Acoustical Society of America*, vol. 89, no. 2, 1991.

[12] R. A. Yamada, "Effect of extended training on /r/ and /l/ identification by native speakers of Japanese," *The Journal of the Acoustical Society of America*, vol. 93, no. 4, pp. 2391–2391, 1993.

[13] R. A. Rescorla, "Pavlovian conditioning: It's not what you think it is." *American Psychologist*, vol. 43, no. 3, p. 151, 1988.

[14] A. Wagner and R. Rescorla, "Inhibition in pavlovian conditioning: Application of a theory," *Inhibition and learning*, pp. 301–336, 1972.

[15] J. D. O'Connor, L. J. Gerstman, A. M. Liberman, P. C. Delattre, and F. S. Cooper, "Acoustic cues for the perception of initial /w, j, r, l/ in english," *Word*, vol. 13, no. 1, pp. 24–43, 1957.

[16] A. J. Lotto, M. Sato, and R. L. Diehl, "Mapping the task for the second language learner: The case of japanese acquisition of /r/ and /l/," *From sound to sense*, vol. 50, no. 2004, pp. C381–C386, 2004.

[17] C. Kiriloff, "On the auditory perception of tones in Mandarin," *Phonetica*, vol. 20, no. 2-4, pp. 63–67, 1969.

[18] P. Boersma and D. Weenink, "Praat (version 5.5)," 2014.

[19] X. Lin and D. Zhang, "Inference in generalized additive mixed models using smoothing splines," *Journal of the Royal Statistical Society*, vol. 61, no. 7, p. 381, 1999.

[20] S. Wood, *Generalized additive models: an introduction with R*. Boca Raton: CRC press, 2006.

[21] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2017. [Online]. Available: https://www.R-project.org/

[22] J. S. Nixon, *Sound of Mind: electrophysiological and behavioural evidence for the role of context, variation and informativity in human speech processing*. Doctoral dissertation, University of Leiden, Netherlands, 2014.

[23] C. de Cat, E. Klepousniotou, and H. Baayen, "Representational deficit or processing effect? An electrophysiological study of noun-noun compound processing by very advanced L2 speakers of English," *Frontiers in Psychology*, vol. 6, p. 77, 2015.

[24] J. S. Nixon, J. van Rij, X. Q. Li, and Y. Chen, "Cross-category phonological effects on ERP amplitude demonstrate context-specific processing during reading aloud," in *ExLing 2015: Proceedings of the International Conference of Experimental Lingustics*, A. Botonis, Ed., 2015, pp. 50–53.

[25] A. Tremblay and A. Newman, "Modelling non-linear relationships in ERP data using mixed-effects Regression with R examples," *Psychophysiology*, vol. TBA, pp. 1–16, 2014.

[26] D. Arnold, P. Wagner, and H. Baayen, "Using generalized additive models and random forests to model German prosodic prominence," *Proceedings of Interspeech 2013*, pp. 272–276, 2013.

[27] F. Tomaschek, M. Wieling, D. Arnold, and R. H. Baayen, "Word frequency, vowel length and vowel quality in speech production: an EMA study of the importance of experience." in *Interspeech*, 2013, pp. 1302–1306.

[28] S. Kawase, *Examination of the role of native speech rhythm in non-native speech production and its perception*. Doctoral dissertation, University of Western Sydney, Australia, 2017.

[29] M. Tamminga, C. Ahern, and A. Ecay, "Generalized additive mixed models for intraspeaker variation," *Linguistics Vanguard*, vol. 2, no. s1, 2016.

[30] M. Wieling, S. Montemagni, J. Nerbonne, and R. H. Baayen, "Lexical differences between tuscan dialects and standard italian: Accounting for geographic and sociodemographic variation using generalized additive mixed modeling," *Language*, vol. 90, no. 3, pp. 669–692, 2014.

[31] J. van Rij, B. Hollebrandse, and P. Hendriks, "Children's eye gaze reveals their use of discourse context in object pronoun resolution." in *Empirical perspectives on anaphora resolution: Information structural evidence in the race for salience*, A. Holler, C. Goeb, and K. Suckow, Eds. Berlin, Walter de Gruyter, 2016.

[32] K. Lõo, J. van Rij, J. Järvikivi, and R. H. Baayen, "Individual differences in pupil dilation during naming task," in *Proceedings of the 38th Annual Conference of the Cognitive Science Society*, A. Papafragou, D. Grodner, D. Mirman, and J. Trueswell, Eds. Austin, TX: Cognitive Science Society, 2016, pp. 550–555.

[33] H. Baayen, S. Vasishth, R. Kliegl, and D. Bates, "The cave of shadows: Addressing the human factor with generalized additive mixed models," *Journal of Memory and Language*, vol. 94, pp. 206–234, 2017.

[34] J. van Rij, M. Wieling, R. H. Baayen, and H. van Rijn, "itsadug: Interpreting time series and autocorrelated data using GAMMs," 2016, R package Version 2.2.

[35] D. Dahan, M. K. Tanenhaus, and A. P. Salverda, "The influence of visual processing on phonetically-driven saccades in the 'visual world' paradigm." in *Eye movements: A window on mind and brain*, 2007, pp. 471–486.

[36] S.-j. Lim and L. L. Holt, "Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization," *Cognitive science*, vol. 35, no. 7, pp. 1390–1405, 2011.