Neural response development during distributional learning

Natalie Boll-Avetisyan[1], Jessie S. Nixon[2], Tomas O. Lentz[3], Liquan Liu[4,5], Sandrien van Ommen[6], Çağrı Çöltekin[2], Jacolien van Rij[7]

[1]University of Potsdam, Germany
[2]University of Tübingen, Germany
[3]University of Amsterdam, The Netherlands
[4]Western Sydney University, Australia
[5]Centre of Excellence for the Dynamics of Language, Australian Research Council, Australia
[6]Laboratoire Psychologie de la Perception, Université Paris Descartes, France
[7]University of Groningen, The Netherlands

`nboll@uni-potsdam.de`, `jessie.nixon@uni-tuebingen.de`, `j.c.van.rij@rug.nl`

Abstract

We investigated online electrophysiological components of distributional learning, specifically of tones by listeners of a non-tonal language. German listeners were presented with a bimodal distribution of syllables with lexical tones from a synthesized continuum based on Cantonese level tones. Tones were presented in sets of four standards (within-category tokens) followed by a deviant (across-category token). Mismatch negativity (MMN) was measured. Earlier behavioral data showed that exposure to this bimodal distribution improved both categorical perception and perceptual acuity for level tones [1]. In the present study we present analyses of the electrophysiological response recorded during this exposure, i.e., the development of the MMN response during distributional learning. This development over time is analyzed using Generalized Additive Mixed Models and results showed that the MMN amplitude increased for both within- and across-category tokens, reflecting higher perceptual acuity accompanying category formation. This is evidence that learners zooming in on phonological categories undergo neural changes associated with more accurate phonetic perception.

Index Terms: lexical tone, distributional learning, mismatch negativity, Generalized additive mixed-effects modeling

1. Introduction

If specific sounds (e.g., phonemes, lexical tones) are used contrastively in a language, then typically sounds with acoustic properties close to their prototypes have a higher likelihood of occurrence than sounds that are further away from the prototypes. Listeners have been found to be sensitive to these distributions, i.e., they have an intuition about which instances of a sound are a more prototypical exemplar of a category, in the language they speak. There is abundant evidence that phonemes are processed this way (since [2]). Evidence is also accumulating that similar procedures determine the perception of lexical tone [3]. The present study investigates the perception of non-native phonetic contrasts cued by acoustic information that is irrelevant in the native language: we explore the formation of sensitivity to tonal contrasts by native speakers of a language without lexical tone.

Generally, listeners have difficulties perceiving, identifying and learning non-native contrasts [4, 5]. However, adults' perception and categorization of speech sounds can be modulated based on brief exposure with novel distributional information [e.g., 6, 7].

The present study is the first to investigate the effect of distributional learning on listeners' perception of a non-native phonetic contrast *during exposure* in real time, using neurophysiological measures. Prior distributional learning studies have used variants of the same classical two-phase procedure: first, participants are familiarized with sound tokens sampled from an acoustic continuum, and the frequency of occurrence of these tokens will be organized in a uni- or bimodal distribution. After this, it is tested whether listeners are now more likely to perceive two sounds as belonging to one or two different categories [6, 8]. Here, rather than assessing effects of distributional learning in such a test phase, we apply Generalized Additive Mixed-Modeling (GAMM) to explore how perceptual sensitivity to tonal contrasts changes over the time course of the experiment. In [9] we used GAMM models to show that not only the number of Gaussians (uni- vs. bimodal), but even the distribution shape (statistical variance) can affect perceptual uncertainty. The present study extends this work by investigating online perceptual adaptations by listeners exposed to non-native acoustic cues presented in a bimodal distribution, to test the online development of non-native phonetic category acquisition.

Earlier distributional learning studies have largely focused on whether listeners' intuitions about category boundaries can be altered [e.g., 6]. However, we [1] recently demonstrated that distributional learning is broader: in addition to the learning of new categories, listeners showed enhanced sensitivity to small acoustic differences within these novel categories. We exposed German listeners to bimodally distributed tokens from a non-native lexical tone continuum and examined effects of exposure in two offline tasks that followed the exposure phase. Results suggested increased discrimination of both small acoustic pitch differences and lexical tone after exposure, compared to a no exposure control group. First, only the group that received exposure demonstrated effects of categorical perception in AX discrimination. Second, their sensitivity to fine-grained acoustic detail was enhanced along the continuum, as revealed by their performance in a pitch height estimation task. This finding is consistent with studies that demonstrated listeners' sensitivity to fine-grained within-category acoustic detail in their native language [e.g., 10].

While in [1] we focused on effects of distributional learning on behavior after exposure, for the present paper, we analyzed EEG data collected in the same study, during the exposure to the bimodally distributed stimuli. We explore changes in the auditory mismatch negativity (MMN) response over the time course of the experiment. The MMN [11; for a review see 12] is

Table 1: *Top row: Pitch continuum with pitch < 7 labeled as a and pitch > 7 labeled as b. Second row: number of tokens per block of type Ab (a-a-a-a-b). Third row: number of tokens per block of type Ba (b-b-b-b-a). Each block consisted of 42 sequences.*

mid tone (a)							high tone (b)					
1	2	3	4	5	6	7	8	9	10	11	12	13
1	5	15	15	5	1	0	4	20	60	60	20	4
4	20	60	60	20	4	0	1	5	15	15	5	1

a relative negativity obtained by subtracting the neural response to a perceived deviant in a series of repeated standards from that to those standards. The MMN occurs between 150 and 250 ms after change onset, in frontal and central regions. Prior studies have suggested that MMNs do not only reflect acoustic-phonetic sensitivity: if differences between sounds are contrastive for listeners, then their mismatch response shows a larger peak and lasts longer than their mismatch response to a purely acoustic-phonetic change [13, 14]. Hence, the MMN is a good measure for our current purposes.

We hypothesized that the difference in MMN response to standards and deviants would increase over the course of the experiment as a result of learning. Additionally, we expected an increase in MMN over the course of the experiment in response to a within-category change indicative of an increased sensitivity to pitch. In this case larger amplitude reflects a more precise neural representation of the memory trace of each pitch token.

2. Methods

2.1. Participants

Twenty right handed, native speakers of German (mean age: 26.5 years, sd=6.9, range: 19–45) were recruited for the experiment and received monetary compensation. After exclusion of data due to excessive noise in the recordings (mainly ocular drift), sixteen participants remained (mean age 26 years, range 21–42, sd=5.6, four male).

2.2. Stimuli

A female speaker with phonetics training produced the nonword /li:/ (duration = 285 ms),with neutral pitch (F0 = 258 Hz). The CV syllable /li/ was selected because its segments are good carriers of pitch information. Pitch was manipulated with Praat [15] to be level on the vowel. From this syllable, a 13-step pitch continuum was created ranging from 260 Hz to 287 Hz. The steps were spaced 0.14 semitones from each other, that is, the tones were relatively close: Adults usually cannot perceive a difference of 0.14 semitones, but it gets easier with growing distance.

2.3. Experiment design and procedure

Participants sat at a computer in a sound-attenuated booth. Stimuli were presented over Sennheiser HD 280 pro, 64Ω headphones. EEG was recorded from 32 electrodes as participants were exposed to the training stimuli. Eye movements were recorded with an Eyelink 1000 eye tracker in order to record and later remove ocular artefacts in the EEG data. Stimulus presentation and participant responses were conducted with SR Research Experiment Builder software. The exposure phase lasted approximately 25–30 minutes.
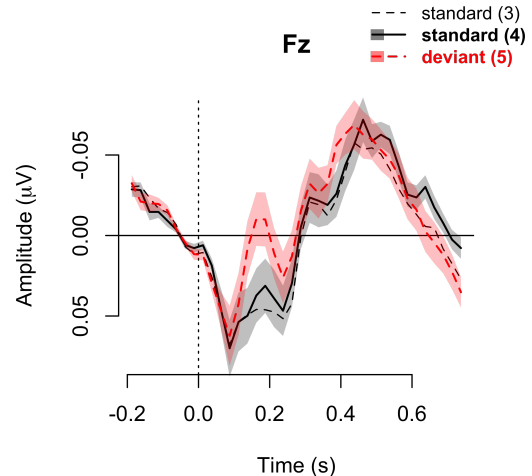

Figure 1: *Grand averages of the data for Fz, with ribbons representing the variation in participant averages ($\pm 1SE$).*

The presentation frequency of the 13 continuum steps followed a bimodal distribution (Table 1). Each trial consisted of a series of five stimuli. An *oddball* design was used, in which a series of four *standards* (tokens selected from the same tone category) was presented (randomly sampled from one of the two distributions, e.g., mid tone tokens 1–6), followed by one *deviant* (randomly selected from the other distribution, e.g., high tone tokens 8–13). The inter-stimulus interval was jittered (between 600–800 ms) to reduce expectancy waves in the EEG. Note that unlike traditional oddball experiments, the standard tokens are not identical, but only sampled from the same distribution (i.e., a trial presents e.g, steps 1-4-2-3-12 or 9-12-11-10-2). Participants were asked to focus on the screen. A progress bar showed the progression through the experiment.

The experiment consisted of eight blocks of 42 trials (336 in total). Within a block, a standard was either always a mid tone or always a high tone. The tone used for the standard alternated between blocks and the order (starting high or mid) was counterbalanced between participants. To make sure participants were alert, they were presented with one tone from the continuum at the end of each block and had to decide whether or not they had heard this token during the previous block. Following training, participants performed two behavioral tasks; results of these are reported in [1].

2.4. Preprocessing

After EEG recording, the data was preprocessed using the Field-Trip toolbox [16]. Trials were first visually inspected to exclude trials and/or channels with too much noise or movement artefacts. 1183 trial-channel combinations were excluded (4.4% of the total). Next, an Independent Component Analysis (ICA) was used to isolate components of the signal that are of interest from artefact-related components (for a detailed overview, see [17]), allowing deletion of e.g., blinks or eye movement components. The ICA was performed on unaveraged data; the components that had scalp distributions consistent with eye movements or blinks were excluded. The signal was then recompiled without the excluded components.

3. Results

Due to space limitations we present results for electrode Fz (where MMN effects are usually expected to occur) and a topo-

Single electrode model:

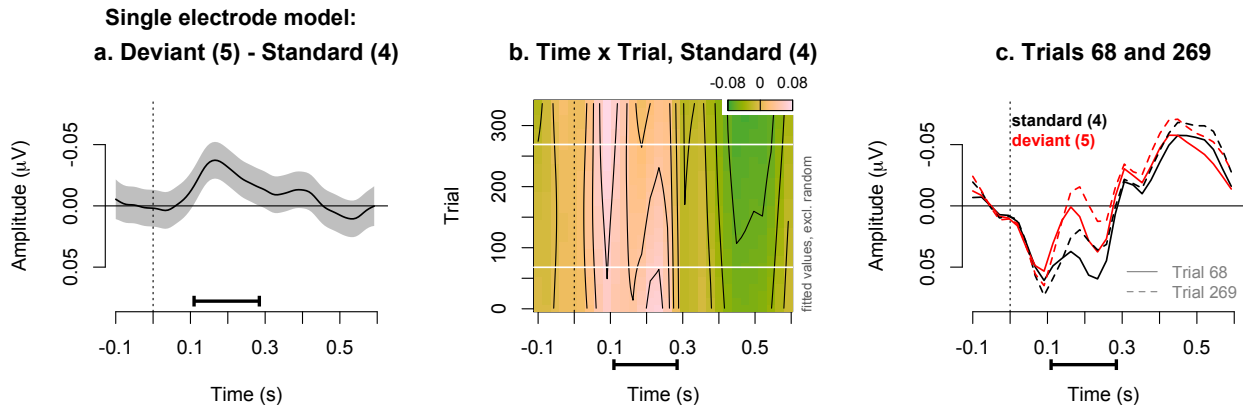| a. Deviant (5) - Standard (4) | b. Time x Trial, Standard (4) | c. Trials 68 and 269 |

Figure 2: *Visualizing the estimates of the single electrode model for Fz. Left: The difference between deviant and standard trials over time, with the horizontal thick line on the x-axis indicating where the difference is significant, based on the pointwise 95% confidence intervals (random effects set to zero). The difference is not modulated by Trial. Center: Interaction between Time and Trial for standard trials (summed effects, random effects set to zero); the peak around 100 ms gets more positive over the course of the experiment, and the modulations between 150–300 ms more negative. The white horizontal lines indicate trials 68 (25% of the experiment) and 269 (75%). Right: The time course of standard (black lines) and deviant (red lines) trials for trials 68 (solid lines) and 269 (dashed lines).*

graphic analysis only. Figure 1 shows the grand averages of the data for electrode Fz. The plot compares the last three stimuli of each trial sequence (two standards, one deviant). Visual inspection suggests an increased negative amplitude between 100–300 ms for the deviant stimuli in comparison with the two standards, and a more positive amplitude after 500ms.

The data were analyzed using GAMMs [18, 19] as implemented in the R package `mgcv` version 1.8-23 [19]. The package `itsadug` version 2.3 [20] was used for evaluation and visualization of the statistical models. GAMM is a nonlinear regression method that allows for nonlinear trends and interactions, which makes it particularly well suited for analyzing EEG data. In our analyses, we compared the last standard (stimulus 4) with the deviant (stimulus 5). Two types of analyses were performed: *a)* a time course analysis on single electrodes to determine the time window of the differences between the standard and deviant stimuli, and *b)* an analysis investigating the scalp distribution of the effect within the determined time period.

3.1. Single electrode analysis

We modeled the change in amplitude over Time (100 ms before until 600 ms after onset of stimulus) and Trial (position of stimulus sequence in experiment, ranging from 1–336), and the potentially nonlinear interaction between these two covariates. In addition, the interaction between Time, Trial, and IsDeviant (binary predictor capturing the difference between the standard and deviant stimuli) was added. We included a nonlinear random effect (factor smooth) for Event (interaction between Participant, Block, and IsDeviant) over Time, to capture the general pattern per participant per block for the two types of stimuli In addition, a nonlinear random effect was included for Trial by Participant to capture differences between participants in the general change in amplitude over the course of the experiment. Finally, an AR1 model was included to account for autocorrelation in the residuals that arises with time course analysis (see [21]). The same model was run for a number of electrodes, with comparable results in the region of interest. The analysis described here exemplifies a model of Fz. The scalp distribution of the effects will be addressed in the following section.

The model summary indicated that the amplitude change over time was significantly different for the two types of stimuli

$(F(15.434, 634373.967) = 2.48; p < .001)$, but the difference was not modulated by Trial: The interaction between Time and Trial significantly contributed to the model $(F(21.006, 634373.967) = 1.87; p < .01)$, but this interaction was the same for the standard and deviant stimuli $(F(1.018, 634373.967) = 0.03; p > .1)$. Figure 2 visualizes the model estimates. The left panel (a) plots the difference between standard and deviant stimuli, which was significant in the time window 100–300 ms, with more negative amplitudes for the deviant stimuli. The center panel (b) shows the interaction between Time and Trial, suggesting that the strongest effect of training exposure is found in the time window 50–300 ms, namely a positivity increase at 50–150 ms and a negativity increase at 150–300 ms. The right panel (c) illustrates the exposure effect by plotting the model estimates for the trials 68 (25% through experiment) and 269 (75%), suggesting more negativity around 150–300 ms at trial 269 than at trial 68. Interestingly, the difference between deviant and standard stimuli (as illustrated in the left panel) is not modulated by trial, but stays the same over the course of the exposure.

3.2. Topography analysis

To investigate the scalp distribution of the training effect, we calculated the mean amplitude per participant per trial in the time window 150–250 ms. A topographic GAMM analysis was run on the mean amplitudes in this time window, to investigate the training effect in more detail and to inspect the scalp distribution of the difference between the standard and deviant stimuli. The GAMM model included a nonlinear three-way interaction (and underlying main effects and interactions) between the X and Y coordinates of the electrodes and Trial. In addition, the interaction between X, Y, Trial and IsDeviant was added, and a random nonlinear smooth over Trial for each Electrode nested within Participants.

The topographic analysis revealed a difference in topographic distribution between the standard and deviant stimuli $(F(10.273, 303767.298) = 3.61; p < .001)$, and different effects of Trial for the standards and deviants $(F(9.562, 303767.298) = 18.00; p < .001)$. In addition, the interaction between Trial and the topographic distribution of the electrodes was found significant $(F(8.948, 303767.298) = 3.31; p < .001)$. However, this additional interaction ef-
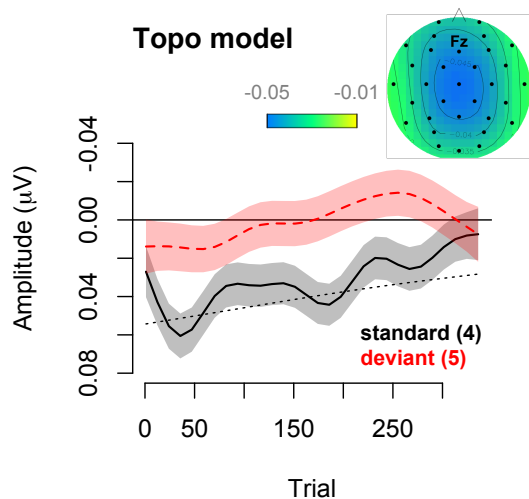
Figure 3: *Topography analysis of time window 150–250 ms. The graph visualizes the amplitude change in Fz over the course of the experiment for standard (black solid line) and deviant (red dashed line) stimuli with pointwise 95% confidence intervals (random effects set to zero). For comparison, the single electrode model's estimate of the training effect in time window 150–250 ms is added (dotted thin line). Topright the difference in amplitude at Trial 200 between standard and deviant stimuli in all electrodes is visualized.*

fect was the same for standard and deviant stimuli ($F(4.856, 303767.298) = 0.66; p > .1$). The training effect is visualized in Figure 3. In this time window, the amplitudes get more negative over the course of the experiment, in line with the single electrode analysis (represented by the thin dotted line). The standard and deviant stimuli differ significantly in amplitude most of the experiment, only after Trial 250 the difference disappears. The topographical distribution of the difference between deviant and standard is localized fronto-centrally and centrally as expected for an MMN [12].

To summarize, the difference between deviant and standard stimuli was significant between 100–300 ms after stimulus onset, with more negative amplitudes for the deviant stimuli. Although the single electrode analysis did not detect an effect of training on the difference in amplitude, our topographic distribution analysis found a significant effect of training: we found that the difference disappears at the end of the experiment, and that the difference is found in the fronto-central electrodes, pointing to an MMN effect.

4. Discussion and conclusion

The present study investigated the effects of distributional learning on the MMN response to a non-native cue dimension, namely pitch height in lexical tone. Stimuli were sampled from a bimodal distribution, where the peaks corresponded to Cantonese high and mid level tones. On each trial, a series of four stimuli (the standards) were sampled from one of the Gaussian distributions (e.g., mid tone), followed by a single stimulus sampled from the other Gaussian (the deviant; e.g., high tone). We were interested in the development of the MMN response to the standards and deviants over the course of the experiment.

Results showed a significantly greater amplitude in the negative-going waveform at 200 ms following the deviant, compared to the standard stimulus, consistent with previous studies

investigating the MMN component (see [12]). The difference in MMN between standards and deviants was present in the first block and remained until near the end of the experiment. This training effect on the MMN was found in the topographic analysis, but was smoothed out in the single electrode analysis.

Most interestingly, there was a significant effect of ongoing exposure for both the standard and the deviant stimuli. Because there was acoustic variation within the standard stimuli as well as between standards and deviants, this suggests an increased sensitivity to the acoustic pitch cue differences over the course of the experiment. The present results thus suggest that neural measures also reveal the increased acuity to within-category acoustic pitch differences following exposure, convergent with our [1] behavioral data.

Our exploration of changes in neural responses over the time course of the experiment revealed fluctuations that we were not expecting, such as the N1 positivity increase and the MMN negativity increase. Currently we hesitate to interpret these results, as we are not aware of observations of similar results in previous studies. In any case, it is interesting that voltage changes at the N1 versus MMN component go in opposite directions, highlighting that they are separate components. Moreover, while we found a constant MMN effect in the deviant compared to the last standard during most part of the exposure phase, i.e., until trial 250 of 336, but the effect disappeared afterwards (Figure 3). Currently we do not understand this result, and we are not aware of observations of similar results in previous studies. It might simply be a result of fatigue or habituation to the paradigm, but further investigation is required to explain this result.

In sum, we showed that GAMMs allow to test the neural processing of level tones during the distributional learning that ends up in category formation. Our results indicate that an MMN associated with a perceived change between a stimulus and the previous four standards can be elicited between categories even with standards that are not identical but fall within one category. The MMNs of participants that learned a category during exposure do not necessarily increase as the perceptual difference between the emerging categories increases [cf. 22]. However, the increasing perceptual acuity also made all stimuli appear more different from each other, not just the ones of the different freshly acquired categories. This effect may be enhanced by the fact that learners with a native non-tonal language started out not perceiving pitch as contrastive at the lexical level.

In conclusion, our study shows that, within a short period of time, learners zooming in on phonological categories undergo neural changes associated with more accurate phonetic perception. Moreover, our study shows that it is important to investigate the development of neural responses during speech perception experiments, as they are not constant over time. This study shows that EEG data gathered during phonological category acquisition is an important addition to behavioral studies as they help us gain insights into the dynamics of perceptual adaptation.

5. Acknowledgements

6. References

[1] 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 Speech Prosody 9*, 2018, in press.

[2] A. M. Liberman, K. S. Harris, H. S. Hoffman, and B. C. Griffith, "The discrimination of speech sounds within and across phoneme boundaries." *Journal of experimental psychology*, vol. 54, no. 5, p. 358, 1957.

[3] 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, no. 3, pp. 395–421, 2004.

[4] 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. Baltimore: York Press, 1995, pp. 171–204.

[5] A. Chen, L. Liu, and R. Kager, "Cross-domain correlation in pitch perception, the influence of native language," *Language, Cognition and Neuroscience*, vol. 31, no. 6, pp. 751–760, 2016.

[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, pp. EL206–EL212, 2011.

[7] J. H. Ong, D. Burnham, and P. Escudero, "Distributional learning of lexical tones: A comparison of attended vs. unattended listening," *PloS one*, vol. 10, no. 7, p. e0133446, 2015.

[8] J. Maye and L. Gerken, "Learning phonemes without minimal pairs," in *Proceedings of the 24th annual Boston university conference on language development*, vol. 2, 2000, pp. 522–533.

[9] 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.

[10] B. McMurray, M. K. Tanenhaus, and R. N. Aslin, "Gradient effects of within-category phonetic variation on lexical access," *Cognition*, vol. 86, no. 2, pp. B33–B42, 2002.

[11] R. Näätänen, A. W. Gaillard, and S. Mäntysalo, "Early selective-attention effect on evoked potential reinterpreted," *Acta psychologica*, vol. 42, no. 4, pp. 313–329, 1978.

[12] R. Näätänen, P. Paavilainen, T. Rinne, and K. Alho, "The mismatch negativity (MMN) in basic research of central auditory processing: a review," *Clinical neurophysiology*, vol. 118, no. 12, pp. 2544–2590, 2007.

[13] G. Dehaene-Lambertz and S. Baillet, "A phonological representation in the infant brain," *NeuroReport*, vol. 9, no. 8, pp. 1885–1888, 1998.

[14] R. Näätänen and K. Alho, "Mismatch negativity–the measure for central sound representation accuracy," *Audiology and Neurotology*, vol. 2, no. 5, pp. 341–353, 1997.

[15] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [computer program] version 5.3.74, retrieved 30 april 2014, from http://www.praat.org," 2014.

[16] R. Oostenveld, P. Fries, E. Maris, and J.-M. Schoffelen, "FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data," *Computational intelligence and neuroscience*, vol. 2011, p. 1, 2011.

[17] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of neuroscience methods*, vol. 134, no. 1, pp. 9–21, 2004.

[18] T. Hastie and R. Tibshirani, *Generalized additive models*. CRC Press, 1990.

[19] S. N. Wood, *Generalized Additive Models: An Introduction with R*, 2nd ed., ser. Chapman & Hall/CRC Texts in Statistical Science. CRC Press, 2017.

[20] J. van Rij, M. Wieling, R. H. Baayen, and H. van Rijn, "itsadug: Interpreting time series and autocorrelated data using GAMMs," 2017, r package version 2.3.

[21] R. H. Baayen, J. van Rij, C. de Cat, and S. Wood, "Autocorrelated errors in experimental data in the language sciences: Some solutions offered by generalized additive mixed models," in *Mixed-Effects Regression Models in Linguistics*, D. Speelman, K. Heylen, and D. Geeraerts, Eds. Springer, 2018, pp. 49–69.

[22] L. Liu, J. H. Ong, A. Tuninetti, and P. Escudero, "One way or another: Adult listeners learn difficult non-native tone contrasts pre-attentively," *Frontiers in Psychology*, vol. 9, p. 162, 2018.