Listening with great expectations:
An investigation of word form anticipations in naturalistic speech

M. Bentum[1], L. ten Bosch[1,2], A van den Bosch[1,3], M. Ernestus[1,2]

[1]Center for Language Studies, Radboud University, Nijmegen, the Netherlands
[2]Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands
[3]KNAW Meertens Institute, Amsterdam, the Netherlands

{m.bentum, l.tenbosch, a.vandenbosch, m.ernestus} @let.ru.nl

Abstract

The event-related potential (ERP) component named *phonological mismatch negativity* (PMN) arises when listeners hear an unexpected word form in a spoken sentence [1]. The PMN is thought to reflect the mismatch between expected and perceived auditory speech input. In this paper, we use the PMN to test a central premise in the predictive coding framework [2], namely that the mismatch between prior expectations and sensory input is an important mechanism of perception. We test this with natural speech materials containing approximately 50,000 word tokens. The corresponding EEG-signal was recorded while participants (n = 48) listened to these materials. Following [3], we quantify the mismatch with two word probability distributions (WPD): a WPD based on preceding context, and a WPD that is additionally updated based on the incoming audio of the current word. We use the between-WPD cross entropy for each word in the utterances and show that a higher cross entropy correlates with a more negative PMN. Our results show that listeners anticipate auditory input while processing each word in naturalistic speech. Moreover, complementing previous research, we show that predictive language processing occurs across the whole probability spectrum.

Index Terms: phonological mismatch negativity, speech perception, statistical language models, predictive coding, electroencephalography

1. Introduction

Human listeners easily perceive words when listening to continuous speech in normal circumstances. This apparent ease hides the profound difficulty of extracting words from a speech stream, which is for example attested by the challenge of developing artificial speech recognition systems with human-like performance. The details of the human speech processing system are still contentious. A long-standing debate concerns the timing and importance of top-down and bottom-up processing. Autonomous models of word recognition (e.g. Shortlist [4,5]) claim that early phases of speech processing are exclusively bottom-up and that top-down information can only exert influence at the lexical level. In contrast, interactionist models allow for top-down influence to affect lower-level acoustic processing (e.g. TRACE [6]). We investigate the influence of top-down expectations on the processing of low-level auditory speech input with an event related potential (ERP) called the phonological mismatch negativity (PMN).

The PMN (also referred to as N200, N250 or phonological mapping negativity) was first reported by [1,7,8]. They presented spoken sentences to participants while recording the electroencephalography (EEG) signal. The sentences were highly constraining, for example, "*the gambler had a streak of bad …*" and the final word either initially matched the expected word (e.g. *luggage*, when *luck* is expected), or, alternatively, mismatched the expected word. A mismatching word showed an early negativity, around 200 milliseconds from word onset compared to the initially matching word.

The PMN is interpreted by [1] as reflecting a mismatch between expected word forms (based on the context) and observed auditory input (for a different interpretation of the PMN see [9,10]). As noted by [11], this interpretation is closely related to an important claim in the predictive coding framework [2], namely, that higher-level cognitive processes generate predictions about low-level perceptual input. The mismatch between these predictions and the perceived input results in an error signal, useful for generating new expectations. For speech perception, this could mean that listeners generate word form expectations based on the preceding context, and violations of these expectations incur a processing cost.

In the current study, we aim to investigate predictive language processing. As [12] notes, many experiments investigating predictive language processing only test extremes of predictability (see also [13,14]). For example, the N400, an ERP component thought to reflect word predictability ([15]; but see [10] for a different interpretation) is typically based on a comparison of very likely versus very unlikely words. The aforementioned PMN is for example elicited with highly constraining sentences as the example above. This leaves open the question of whether language processing normally involves prediction, or only in these extreme cases.

Recently, several studies (e.g. [16,17]) have shown that the whole spectrum of word predictability can be investigated by utilizing information-theoretic measures. For example, [16] used word surprisal, estimated with a statistical language model, to successfully predict the amplitude of the N400 measured while participants were reading sentences. Their approach is not based on the dichotomy of likely versus unlikely words, but instead uses the whole range of word probabilities. Furthermore, they used a large set of naturalistic language materials (see also [18]), improving the ecological validity of their findings. These studies provide stronger evidence for prediction during normal language processing.

In the current study we expand on this type of research with a mismatch measure inspired by the predictive coding framework. The measure developed in [3] quantifies the mismatch between expected and actual sensory speech input. To implement this mismatch measure, we need to quantify the mismatch between top-down expectations and bottom-up

observations. The top-down expectations are estimated with a statistical language model (SLM). The SLM estimates the probability P of a word W_i given the preceding words W_{i-n} ... W_{i-1} and thus captures the top-down expectations (see Eq. 1).

$$P(W_i) = P(W_i|W_{i-n}, ..., W_{i-1}) \qquad (1)$$

The bottom-up speech input will be represented with an auditory fragment of the initial part of the current word. This audio fragment needs to be processed in such a way that it can update word expectations. To achieve this, we use automatic speech decoding techniques used in speech recognition software and adapt these to estimate the probability of a phoneme sequence given the partial auditory fragment of the current word. We use the resulting phoneme sequence probabilities to update the top-down expectations (i.e. word probabilities estimated with the SLM). In this manner we compute two word probability distributions (WPD): one prior WPD based only on the SLM output (which is based on the previous words), and one post WPD, which is the prior WPD updated with the phoneme sequence probabilities based on the audio fragment of the current word.

The post WPD differs only from the prior WPD in the added auditory information. We therefore propose that the cross entropy between prior and post WPD quantifies the mismatch between high-level expectations (based on previous word context) and auditory input. The cross entropy can be computed according to Equation 2, whereby H denotes cross entropy, p the prior WPD, q the post WPD and X the WPD word list.

$$H(p, q) = - \sum_{x \in X} p(x) \log q(x) \qquad (2)$$

In the current study we test whether we can predict the amplitude of the PMN with the cross entropy between prior and post WPDs. Based on the predictive coding framework and the EEG literature, we hypothesize that with increasing cross entropy a listener incurs a higher processing cost (i.e. the sensory speech input is more surprising), which is reflected in a more negative amplitude in the 200 millisecond latency range. In the following sections, we will describe the EEG experiment and the methods used to test our hypothesis, followed by the results, a discussion and a conclusion.

2. Method

2.1. Participants

Forty-eight neurologically unimpaired right-handed native speakers of Dutch (18-29 years, mean age = 21.7 years), 14 men and 34 women, participated in the three sessions of EEG recordings. All participants gave informed consent to participation.

2.2. Materials

For the experimental stimuli, we used materials from two corpora: the Spoken Dutch Corpus [19] and IFADV [20]. These corpora contain audio recordings of Dutch speech. We extracted stretches of speech from these corpora, varying in duration from 4 to 15 minutes). The extracted speech stretches contain 50,277 word tokens (see Table 1). This subset, henceforth called Speech Corpus, consists of annotated speech from three speech registers (see Table 1). The different registers were selected for a different experiment and will not be relevant

for the current study. For the estimation of the cross entropy (see Section 2.3) we used NLCOW14, henceforth COW [21,22], which is a large collection of web-crawled Dutch texts (4,7 billion words).

Table 1: *Overview of the materials in the Speech Corpus.*

speech style	word tokens (word types)	average word duration (ms)
spontaneous dialogues	21,718 (2,435)	206
read-aloud stories	13,209 (2,349)	256
news broadcast	15,350 (3,526)	289
total	50,277 (5,866)	245

2.3. Computing cross entropy

We computed the cross entropy (as detailed in [3]) for all words longer than 60 and shorter than 700 milliseconds (46,734 word tokens, 5,254 word types). All subsequent analyses are performed on this subset of the Speech Corpus. To compute the cross entropy, we need a word probability distribution (WPD) at the start of a word and a WPD after the auditory update. Therefore, we estimated for each word one WPD *prior* and one WPD *post* auditory update. These WPDs consists of a list of approximately 200,000 word types with associated probabilities. We created the word type list by selecting the most frequent word types in the COW corpus. We use the term word type to refer to the surface form of a word, i.e. *boy* and *boys* are two different word types.

The prior WPD is an estimation of word probabilities at the start of a word, given the preceding words (see Eq. 1). For example, consider the phrase *he played the guitar,* the prior WPD for the word *guitar* consists of the 200,000 word types with corresponding probabilities, given the preceding words *he played the.* To estimate the probabilities, we trained a 4th order Markov SLM on the COW corpus with the SRILM [23] toolkit (for smoothing we used Kneser-Ney discounting [24]). We used this SLM to create a prior WPD for each word in the Speech Corpus.

We created the post WPD by updating the prior WPD with an auditory fragment of part of the current word (i.e. *guitar*). To perform this update, we transformed the audio fragment into probabilities of phoneme sequences. To estimate these probabilities, we extracted an audio fragment from word onset. By testing different durations for the auditory update, [3] showed that a fragment of 190 milliseconds, is the optimal duration for the cross entropy computation. We analyzed audio material by using KALDI [25] as a speech decoding framework. We provided KALDI with a dedicated decoding lexicon whereby each entry was a sequence of $1 - 8$ phonemes. We limited the set of phoneme sequences ($n \approx 400,000$) to those that are found in Dutch words. The KALDI analysis resulted in a 50-best list of phoneme sequences with corresponding probabilities. The phoneme sequences were matched with the word types in the prior WPD and the probabilities were adjusted accordingly by the conventional Bayes rule, with the post WPD as result.

2.4. Procedure

Participants visited the lab on three occasions. Consecutive visits were separated by at least a week. Participants were fitted with the correct size electrode cap and seated in a sound-attenuated booth. They were asked to sit still and keep eye-movement and blinks to the minimum. During each visit, participants listened to approximately 90 minutes of speech,

270 minutes in total. The speech materials were presented over in-ear headphones (Etymōtic ER1) on a comfortable listening level (tested with a short audio fragment). The speech materials were presented in blocks of approximately 15 minutes, followed by a short break. During breaks in the experiment, yes-no comprehension questions were visually presented and participants responded via a button box.

2.5. EEG recording

We placed 26 cap-mounted silver-chloride electrodes according to the 10 - 20 international system (Fp2, Fz, F3, F4, F7, F8, FC1, FC2, FC5, FC6, Cz, C3, C4, T7, T8, P3, Pz, P4, P7, P8, CP1, CP2, CP5, CP6, O1, O2). We used four additional electrodes to monitor eye-related artifacts (eye-movements and blinks), placed at the outer left and right canthi, and below and above the left eye (converted off-line to horizontal and vertical EOG signals). Two additional electrodes were placed on the left and right mastoid. All electrodes were referenced to the left mastoid electrode and all electrode impedances were below 15 kΩ before recording started. The EEG-data was amplified with an Easycap system, band-pass filtered with 0.01 and 100 Hz cut-off frequencies, and digitized at a 1000 Hz sample frequency.

2.6. Preprocessing

The EEG-signal was re-referenced off-line to the left and right mastoid channels and filtered with a 5th order Butterworth bandpass filter with cut-off frequencies at 0.05 and 30 Hz. We removed artefacts from the EEG data semi-automatically, whereby all suggested artefacts were manually checked. We determined per block whether EEG channels with poor signal quality should be removed from the dataset. The Fp2 channel was completely removed from all recordings, due to poor overall signal quality.

Subsequently, we removed activity related to blinks and eye movements from the EEG signal with the aid of independent component analysis (ICA). Following [26], the ICA was computed on data bandpass filtered with cut-off frequencies at 1 – 30 Hz, with aid of the MNE toolkit [27,28]. We visually inspected the resulting components and selected those that were related to blinks and eye movement. We recomposed the EEG-data, bandpass filtered at 0.05 – 30 Hz, without the selected components.

The cleaned EEG-signal was time-locked to the words in the Speech Corpus. We extracted an epoch from 300 milliseconds before to a 1000 milliseconds after word onset for each word. All word epochs exceeding ± 75 µV on any channel were excluded from the dataset. We excluded the data of 9 (of the 48) participants because of poor signal quality (i.e. less than 40% of the data remaining after artefact removal). This resulted in a dataset of 1,172,894 word epochs (52.3% of all data).

3. Results

Based on previous literature (e.g. [1,9,11]), we expected a divergence of the grand average ERPs for low, middle and high cross entropy (i.e. we split the data into terciles based on cross entropy) at around 200 milliseconds from word onset at frontal sites. To select a latency range and a set of channels for analysis, we also inspected the grand average ERPs for all channels (see Figure 1 for a subset and Figure 2 for topographic plot of the relevant time window). We computed the average over the 150 – 350 millisecond time window and the following channels: F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7 and T8. The result of

this averaging was one value for each word epoch. Following [16], we did not subtract the baseline from the ERP. Instead, we used the baseline as predictor in our statistical model. We computed the baseline by averaging over the same channel set for the -150 – 0 millisecond time window.
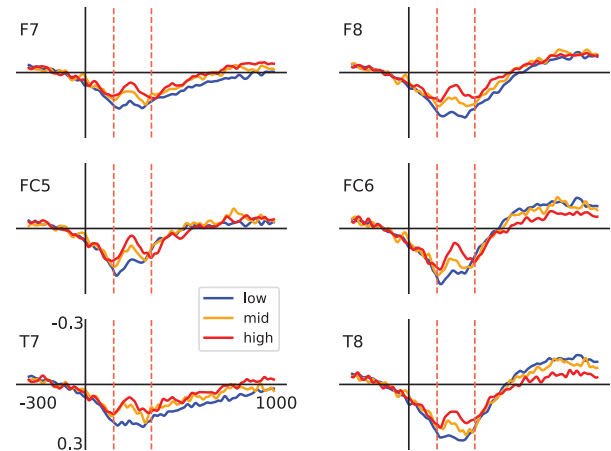

Figure 1: *Grand average ERPs for words with low (blue), middle (yellow) and high (red) cross entropy. The x-axis shows time in milliseconds and the y-axis shows amplitude in µV (negative is plotted upwards). The vertical dashed lines indicate the window between 150 and 350 milliseconds.*

We analysed the data with linear mixed effects models [29] in R [30], with the per-word ERP amplitude as dependent variable and cross entropy as predictor of interest. The standardized covariates are the aforementioned baseline, the surprisal of the word, the entropy of the prior WPD, log frequency of the word in the COW corpus, the duration of the word, the word number in the sentence, and the word number in the block. Furthermore, we added participant and word as random effects. We considered a random slope for cross entropy by participant but did not include it in the final model, because it resulted in a convergence error.

Table 2: *Overview of the fixed effects in the linear mixed effect model with the PMN as dependent variable. The variable names, the beta (B) value, the standard error (SE B) and the t-value (t) are reported.*

name	B	SE B	t
intercept	-0.23	0.04	6.5
entropy	0.11	0.01	16.3
surprisal	-0.02	0.01	-1.9
baseline	6.02	0.01	1145.9
log frequency	0.07	0.02	3.8
duration	-0.06	0.01	-6.7
word in sentence	-0.04	0.01	-6.4
word in block	-0.05	0.01	-8.5
cross entropy	**-0.02**	**0.01**	**-3.6**

We computed two models: a simple model without the predictor of interest, cross entropy, and a second model with this predictor. Table 2 lists the fixed effects of the second linear mixed effect model. Model comparison reveals that the model with cross entropy significantly improves compared to the simple model $\chi = 4.86$, $p < .05$. The PMN is more negative with increasing values of cross entropy.
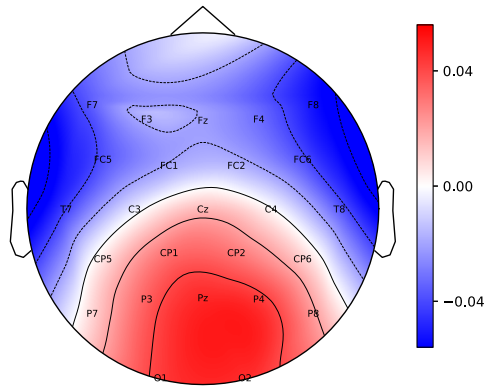
Figure 2: *Topographic difference plot between words with a high cross entropy versus words with a low cross entropy, averaged over the time window 150 - 350 milliseconds after word onset.*

4. Discussion

According to the predictive coding framework [2], the mismatch between prior expectations and sensory input is an important mechanism for perception. We tested this claim for speech perception with an event related potential (ERP) component named the phonological mismatch negativity (PMN). This component is thought to reflect the mismatch between expected and actual auditory word forms. Following [3], we quantified the mismatch between expectations and sensory input as the cross entropy between two-word probabilities distributions (WPD), one based on preceding words, and an updated version based additionally on the auditory input. We found that in line with our hypothesis, the cross entropy has a negative correlation with the PMN amplitude, i.e. higher cross entropy corresponds with a more negative amplitude of the PMN (see Figure 1 & 2 and Table 2). We propose that speech processing involves a comparison between high-level expectations and the auditory speech input. When the input mismatches with the expectations, this incurs a processing cost.

We recorded the electroencephalography (EEG) signal while participants listened to naturalistic speech. Following [18], the speech materials were extracted from corpora and represented normal language. Participants listened to long stretches of speech (approximately 15 minutes), instead of individually presented sentences or words. This has the distinct advantage that more relevant EEG data can be acquired. However, the downside is an increase of artefacts, because it is not possible for participants to sit perfectly still or refrain from blinking for a block of 15 minutes. As a result, more materials needed to be removed (~ 50% of the word epochs) than in more classical EEG experiments. However, because we analysed most words (93%) in the speech materials, we were able to create a dataset with approximately one million word epochs, orders of magnitude larger than analysed in classical EEG experiments.

The amount of data is important for this experiment. We investigated an ERP based on stimuli (i.e. words) while we could not control for sensory form or the preceding context, i.e. the target stimuli were the spoken words occurring in naturalistic speech. ERPs are sensitive to these kinds of differences [31] and it is only by averaging over very many tokens, which we had available, that this diversity averages out in the EEG signal.

Since we investigate many words in many contexts, we could investigate predictive language processing not just in highly constraining or artificially (un)likely contexts. As [12] noted, most experimental evidence for predictive language processing is based on experiments using these kinds of artificial language input. Recent studies (e.g. [16,17,32]) have been using a new approach, whereby language processing costs are predicted based on information theoretic measures, investigating a wide range of the probability spectrum. We extended these findings by using the mismatch measure (developed in [3]) to predict processing costs of word forms during natural language perception. Our study shows that listeners do indeed anticipate word forms over the whole range of predictability, in line with the idea of *graded predictions* (see [13]).

We propose that our finding is best explained by top-down feedback. This explanation is at odds with autonomous models of word recognition (e.g. Shortlist [4,5]), which claim that early speech perception consists of bottom-up-only processing and do not allow for top-down feedback. Norris et al. [14] defend the idea of no feedback by stating that processing of the acoustic signal is already optimal (by means of Bayesian inferencing), and therefore cannot be improved by feedback. However, feedback can be highly informative as an error-detection device, informing the listener whether the current word priors are on point (i.e. how well do they explain the current input?). If the overall error increases, then the generative model needs to be adapted. The difference between the expected and perceived speech input (the error signal) provides a mechanism to dynamically adapt perceptual processing (and a built-in learning system to boot). This 'error signal'-based reasoning is in line with predictive coding, which proposes that early stages of sensory input processing involves propagating the mismatch between the expected and actual sensory input.

5. Conclusions

We used a novel experimental approach in which participants listened to naturalistic speech while their EEG signal was recorded. Based on one million EEG word epochs, we showed that the ERP named the PMN has a negative correlation with cross entropy, which quantifies the mismatch between expected and perceived auditory input. We showed that naturalistic speech stimuli can be used in an EEG experiment, and that it is possible to analyse most words (93%) in these speech materials. Furthermore, we extended research using an information-theoretic measure to predict processing costs of word forms, and provided additional evidence for extensive predictive language processing.

6. Acknowledgements

7. References

[1] J. F. Connolly and N. A. Phillips, "Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences," *Journal of Cognitive Neuroscience,* vol. 6, no. 3, pp. 256-266, 1994.

[2] K. Friston, "A theory of cortical responses," *Philosophical transactions of the Royal Society B: Biological sciences,* vol. 360, no. 1456, pp. 815-836, 2005.

[3] M. Bentum, L. ten Bosch, A. van den Bosch and M. Ernestus, "Quantifying expectation modulation in human speech processing," in *INTERSPEECH 2019 – 20th Annual Conference of the International Speech Communication Association, September 15-19, Graz, Austria, Proceedings*, 2019.

[4] D. Norris, "Shortlist: A connectionist model of continuous speech recognition," *Cognition,* vol. 52, no. 3, pp. 189-234, 1994.

[5] D. Norris and J. M. McQueen, "Shortlist B: A Bayesian model of continuous speech recognition," *Psychological review,* vol. 115, no. 2, pp. 357-395, 2008.

[6] J. L. McClelland and J. L. Elman, "The TRACE model of speech perception," *Cognitive psychology*, vol. 18, no. 1, 1986.

[7] J. F Connolly, S. H. Stewart, and N. A. Phillips, "The effects of processing requirements on neurophysiological responses to spoken sentences," *Brain and language*, vol. 39, no. 2, pp. 302-318, 1990.

[8] J. F. Connolly, N. A. Phillips, S. H. Stewart, and W. G. Brake, "Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences," *Brain and language*, vol. 43, no. 1, pp. 1-18, 1992.

[9] D. van den Brink, C. M. Brown, and P. Hagoort, "Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects," *Journal of cognitive neuroscience,* vol. 13, no. 7, pp. 967-985, 2001.

[10] P. Hagoort, "The fractionation of spoken language understanding by measuring electrical and magnetic brain signals," *Philosophical Transactions of the Royal Society B: Biological Sciences,* vol. 363, no. 1493, pp. 1055-1068, 2007.

[11] A. Brunellière and S. Soto-Faraco, "The speakers' accent shapes the listeners' phonological predictions during speech perception," *Brain and language,* vol. 125, no. 1, pp. 82-93, 2013.

[12] F. Huettig, "Four central questions about prediction in language processing," *Brain research,* vol. 1626, pp. 118-135, 2015.

[13] J. J. van Berkum, C. M. Brown, P. Zwitserlood, V. Kooijman, and P. Hagoort, "Anticipating upcoming words in discourse: Evidence from ERPs and reading times," *Journal of Experimental Psychology, Memory, and Cognition,* vol. 31, no. 3, pp. 443-467, 2005.

[14] D. Norris, J. M. McQueen, and A. Cutler, "Prediction, Bayesian inference and feedback in speech recognition," *Language, cognition and neuroscience*, vol. 31, no. 1, pp. 4-18, 2016.

[15] M. Kutas, K. A. DeLong, and N. J. Smith, "A look around at what lies ahead: Prediction and predictability in language processing," in *Predictions in the brain: Using our past to generate a future*, New York: Oxford University Press, pp. 190-207, 2011.

[16] S. L. Frank, L. J. Otten, G. Galli, and G. Viliocco, "The ERP response to the amount of information conveyed by words in sentences," *Brain and language,* vol. 140, pp 1-11, 2015.

[17] N. J. Smith and R. Levy, "The effect of word predictability on reading time is logarithmic," *Cognition*, vol. 128, no. 3, pp. 302-319, 2013.

[18] R. M. Willems, (Ed.), *Cognitive neuroscience of natural language use.* Cambridge, Cambridge University Press, 2015.

[19] N. Oostdijk, "The design of the Spoken Dutch Corpus," *Language and Computers*, vol. 36, no. 1, pp. 105-112, 2001.

[20] R. van Son, W. Wesseling, E. Sanders, and H. van den Heuvel, "The IFADV Corpus: A Free Dialog Corpus," *LREC*, pp. 501-508. Marrakech: ELRA, 2008.

[21] R. Schäfer, "Processing and querying large web corpora with the COW14 architecture," *Proceedings of the 3rd Workshop on Challenges in the Management of Large Corpora (CMLC-3),* pp. 28-34. Mannheim: Institut für Sprache, 2015.

[22] R. Schäfer and F. Bildhauer, "Building Large Corpora from the Web Using a New Efficient Tool Chain," *Proceedings of the Eight International Conference on Language Resources and Evaluation,* Istanbul: ELRA, pp. 486-493, 2012.

[23] A. Stolcke, "SRILM-an extensible language modelling toolkit," *Proceedings of the International Conference on Spoken Language Processing.* International Speech Communication Association, 2002.

[24] S. F. Chen and J. Goodman, "An empirical study of smoothing techniques for language modeling," *Computer Speech & Language*, vol. 13, no. 4, pp. 359-393, 1999.

[25] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, and K. Vesely. "The Kaldi Speech Recognition Toolkit," *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding.* The IEEE Signal Processing Society, 2011.

[26] I. Winkler, S. Debener, K. R. Müller, and M. Tangermann, "On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP." In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, IEEE, 2015.

[27] A. Gramfort, M. Luessi, E. Larson, D. Engemann, D. Strohmeier, C. Brodbeck, L. Parkkonen, M. Hämäläinen, "MNE software for processing MEG and EEG data," *NeuroImage*, vol. 86, no. 1, pp. 446-460, 2014.

[28] A. Gramfort, M. Luessi, E. Larson, D. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, M. Hämäläinen, "MEG and EEG data analysis with MNE-Python", *Frontiers in Neuroscience*, vol. 7, 2013.

[29] B. Douglas, M. Maechler, B. Bolker, S. Walker, "Fitting Linear Mixed-Effects Models Using lme4," *Journal of Statistical Software,* vol. 67, no. 1, pp. 1-48, 2015.

[30] R Core Team, "R: A language and environment for statistical computing," R Foundation for statistical Computing, Vienna, http://www.R-project.org/, 2015.

[31] S. J. Luck, *An introduction to the event-related potential technique.* Cambridge: MIT press, 2005.

[32] R. M. Willems, S. L. Frank, A. D. Nijhof, P. Hagoort, A. van den Bosch, "Prediction during natural language comprehension," *Cerebral Cortex,* vol. 26, no. 6, pp. 2506-2516, 2015.