Neuroscience Letters 609 (2015) 147–151

Contents lists available at ScienceDirect

Neuroscience Letters

journal homepage: www.elsevier.com/locate/neulet

Research paper

Turn-taking: From perception to speech preparation

Hendrik Wesselmeier*, Horst M. Müller

Experimental Neurolinguistics Group and Collaborative Research Center "Alignment in Communication" (SFB 673), Bielefeld University, Bielefeld, Germany

HIGHLIGHTS

- EEG language production study of a quasi-dialog.
- Readiness potential analysis of articulation start based on turn-anticipation.
- Syntactic stimuli structure modulates readiness potential onset intervals.
- Cognitive load modulates response time but not readiness potential.
- Readiness potential is not related to speech intention but to speech planning.

ARTICLE INFO

Article history:
Received 3 June 2015
Received in revised form 5 October 2015
Accepted 12 October 2015

Keywords:
Event-related potential
Readiness potential
Spoken language
Articulation preparation
Turn-taking
Human

ABSTRACT

We investigated the preparation of a spoken answer response to interrogative sentences by measuring response time (RT) and the response-related readiness potential (RP). By comparing the RT and RP results we aimed to identify whether the RP-onset is more related to the actual speech preparation process or the pure intention to speak after turn-anticipation. Additionally, we investigated if the RP-onset can be influenced by the syntactic structure (one or two completion points). Therefore, the EEG data were sorted based on two variables: the cognitive load required for the response and the syntactic structure of the stimulus questions. The results of the response utterance preparation associated event-related potential (ERP) and the RT suggest that the RP-onset is more related to the actual speech preparation process rather than the pure intention to speak after turn-anticipation. However, the RP-onset can be influenced by the syntactic structure of the question leading to an early response preparation.

1. Introduction

Turn-taking is the change of an interlocutor from listener to speaker in a natural spoken dialog. It has been studied for several decades, and a good understanding of its properties has been achieved. An universal property of a dialog is that it proceeds efficiently, and most turn-taking occurs with temporal precision without overlaps and almost no obvious gaps [1,2]. However, when an interlocutor produces a response utterance, it takes approximately one second between the appearance of the mental concept and the actual motoric articulation [3]. This process leads to the assumption that the following speaker anticipates the turn-end of the interlocutor to prepare their response. Nevertheless, it is unclear at what point a passive listener becomes an active interlocutor and starts to prepare their response, it is also uncertain if syntactic information influences turn-anticipation. Most recent EEG studies investigating language production focused on single word production [4] rather than utterance production in the form of a sentence or even a turn consisting of an illocutonary act. To our knowledge, only one study has investigated the conceptual planning of complex utterances in overt speech, elicited by presenting two pictures [5]. Each picture was associated with an action (e.g., book = reading) whereas a cue indicated if the associated action should be described in a chronological or reversed order. In that EEG experiment, they found a P300-related effect caused by the task complexity in linearization. In most EEG experiments investigating turn-taking, participants were prompted to indicate the anticipated turn-end with a button-press [6] or a simple verbal cue "ja" (yes) [7]. In this experiment, participants were prompted to utter a complete answer with one or more sentences.

The EEG readiness potential (RP, Bereitschaftspotential) is an event-related potential (ERP) in the form of a negative deflation which is related to selective motor preparation [8]. Previous EEG experiments comparing limb movements-RP to speech-RP [9] and experiments comparing the RP of vocal onset to RP of a button-press

* Corresponding author at: Experimental Neurolinguistics Group – SFB 673, Bielefeld University, P.O. Box 100131, 33501 Bielefeld, Germany.
E-mail address: hendrik.wesselmeier@uni-bielefeld.de (H. Wesselmeier).

http://dx.doi.org/10.1016/j.neulet.2015.10.033
0304-3940/

during anticipation [7] have shown that the response-locked RP of vocal onset is similar to RP of a button-press. This evidence means that a voice-related RP-onset can be used for motor preparation in the same manner as it is used for button-press and is not affected by breathing in or other preparation factors for verbalization. In an earlier EEG experiment [10] on anticipation, the button-press-related RP in turn-end detection was influenced by syntactic or semantic violations in the stimulus sentence: participants had to press a button exactly when the sentence ended, and some sentences contained a syntactic or semantic violation. The results showed that both types of violations generated a shorter RP compared to intact sentences, whereas the RT was similar. Nonetheless, in responses after a syntactic violation, an argument for a shorter RP was that it could indirectly influence turn-anticipation by disturbing semantic integration. Additionally, recent behavioral studies [11] manipulated the presence and absence of syntactic information by low-pass filtering open- and closed-class words in turns; these studies confirm that syntactic information provides anticipation cues, even though semantic information is a more important cue for anticipation.

In this study, we investigated the influence of cognitive load, which has been shown to influence speech planning in turn-taking [12], for response preparation on the RT and RP-onset. Therefore, the stimulus questions were divided into two groups: those that could be answered without delay (low cognitive load) and those that could be answered with a short delay (high cognitive load). Furthermore, we investigated the influence of syntactic structure (one or two completion points) using two different types of stimulus questions with different types of syntactic complexity, whereas semantic processing is unaffected. To accomplish a more natural turn-taking situation, participants were asked to respond with an articulated answer. The aim of this study was to gain insight into the temporal aspects of the transition from speech perception to speech production and its preparation in turn-taking. Therefore, at first, we tested if the RP-onset was more related to the general intention to speak or more time related to the actual speech planning process after the decision of how to respond has already been made. Then, we tested if the syntactic structure of the stimulus questions influenced response preparation. The aim was to see whether syntax influences turn-anticipation rather than only indirectly influencing turn-anticipation by disturbing semantic integration. To test these hypotheses, we ran an EEG experiment in which participants were acoustically presented with questions to which the participants had to respond with a brief answer.

2. Material and methods

2.1. Participants

Thirty students (17 f, 20–35 years, mean 24.5) from Bielefeld University participated in our experiment, which lasted about one hour. All participants were native German speakers and right-handed with a lateralization quotient of 88.9 according to the Edinburgh Handedness Inventory [13]. According to their own accounts, they did not suffer from either auditory or motor restrictions or diseases that could have influenced the results. Written consent was obtained from all participants, and the study was approved by the Ethics Committee of Münster University.

2.2. Stimuli

The acoustically presented stimuli consisted of 25 interrogative clauses that varied from 1300 ms to 6643 ms in length with a mean of 3989 ms and a SD of 1421 ms. Questions with two syntactic completion points were significantly longer compared to questions with only one syntactic completion point, $F(1,23) = 11.65$, $p < .005$. For the 12 sentences with a possible completion point within the sentence, the time interval from the first possible syntactic completion point to the utterance end ranged from 1194 ms to 2637 ms (mean = 2048 ms, SD = 508 ms). The duration of the stimuli between the two groups selected based on the RT were not significantly different at the .05 level. All sentences in the experiment were spoken at a speed of 400–450 syllables per minute by a professional female speaker with natural intonation and were recorded in a sound studio.

Stimulus examples:
The "#" indicates the first possible syntactic completion point.
Question, to be answered without delay:
Finden Sie die Mietpreise in Bielefeld angemessen?
(Do you consider the rental prices in Bielefeld as appropriate?)
Question, to be answered with a delay:
Sehen Sie im neuen Bachelor/Master-System eine Verbesserung?
(Do you consider the new Bachelor/Master-system an improvement?)
Question, with the only possible syntactic completion point at the end of the sentence:
Müssen Sie neben dem Studium arbeiten gehen?#
(Do you need to hold a job in addition to your study?#)
Question, with a first possible syntactic completion point within the sentence:
Wohnen Sie alleine# oder in einer WG?
(Do you live alone# or in a shared flat?)

2.3. Rating of the stimuli

A four-point scale for the 25 stimulus sentences was used to verify that the measured RT difference depends on the cognitive load necessary to answer the question rather than the complexity of the sentence and the difficulty of understanding. To distinguish a question that could be answered quickly or with a short delay to think, 58 students (55 f, mean age 23.7, SD = 4.3) who were all native German speakers participated in this rating.

2.4. Procedure

Our experiment was conducted in an electromagnetically shielded and sound-proof booth. Each trial started with a fixation cross presented in the middle of the screen. After the fixation cross appeared, the spoken stimulus sentence started after a random inter-trial interval (range 1000–2500 ms) so that the participants could not anticipate the sentence onset time after several trials. The fixation cross was continuously shown until 6000 ms after the spoken sentence ended. Participants were instructed to give a quick and short answer to the interrogative clause. The participants were informed that the responses would not be recorded or judged. After a short practice block with five different sentences matched in complexity and length, all participants became comfortable with the task. The mean stimulus intensity ranged between 55 and 60 dB, which matches a normal face-to-face conversation.

2.5. EEG recording

EEG was continuously recorded from 32 active scalp electrodes (ActiCap, Brain Products) placed at locations based on the International 10/20 system [14] with the reference at FCz. Signals were sampled at 1000 Hz, amplified with a bandpass of 0.16–80 Hz and a 50 Hz notch filter by amplifiers (QuickAmp, Brain Products) and recorded with BrainVision Recorder software (Version 1.20). The impedance remained below 5 kΩ for all channels prior to recording.

Table 1
Descriptive statistics of the overall RT and divided into two groups for fast and slow RTs as well as without or with a first possible syntactic completion point within the sentence.

Condition		*N*	Mean [ms]	Min [ms]	Max [ms]	SD
Overall		705	815	23	1994	449
Cognitive load	Undelayed	379	677	23	1994	403
	Delayed	326	975	78	1993	446
Syntactic structure	One cp	336	803	23	1993	482
	Two cp	369	825	69	1994	416

2.6. Data analysis

2.6.1. Behavioral data

RT data were recorded by voice key, and analyses were performed with SPSS (Version 20) using Linux. In the RT dataset, 6% of the events were excluded from the RT analysis due to detection errors or outliers that exceeded the two times standard deviation from the mean. In this within subject design, twice two groups of questions were formed independently. Two groups of questions were formed according to whether the mean RT of the questions was above or below the overall mean RT of 815 ms. Another two groups of questions were formed: questions with a possible syntactic completion point within the question and questions with the only syntactic completion point at the end of the sentence. An ANOVA and a repeated-measure ANOVA were conducted for both levels of the two factors on the corresponding RT values from all trials.

2.6.2. Event-related potentials

ERP analyses were conducted to the same groups as in the RT analysis using EEGLAB [15] in MATLAB (Linux). The raw data were re-referenced to an average of all cortical electrodes and segmented in 2000 ms epochs including 1800 ms pre-response and 200 ms post-response onset. For artifact pre-selection, epochs were automatically screened for artifacts of signal values exceeding four standard deviations using a probability function built into EEGLAB [16]. Epochs were then subject to independent component analysis (ICA) to remove ocular and other muscle artifacts [17,18]. The EEG waveforms displayed in Section 3 were digitally smoothed with a 6 Hz low pass-filter. To detect RP-onset, we used the criteria based method [19]: the RP-onset is identified at the first point in time that the RP exceeds some arbitrary value (here $-1\ \mu V$). Since it is even possible to analyze single RP events [20], the statistical significance was assessed with the jackknife procedure [21]. This was done by averaging trials with sentences as analytical unit and entering the values in an ANOVA and participants as analytical unit in a repeated-measure ANOVA, with the *F*-value corrected [22].

3. Results

3.1. Rating and reaction time data

A Spearman's correlation was run to determine the relationship between the rating and RT dependent grouping of the 25 questions. There was a positive correlation between rating and $r_s(25) = .442$, $p = .027$).

Across all participants, 94% of the RTs were within two standard deviations from the mean and were analyzed separately for each group. Table 1 presents the RT descriptive statistics for each group.

Only the RTs sorted for the mean reaction time of the respective cognitive load question was significantly different for sentences $F(1,23) = 57.99$, $p < .001$ and participants $F(1,28) = 55.61$, $p < .001$. A first possible syntactic completion point within the sentence had no significant influence on the RT, for sentences $F(1,23) = 0.83$, $p = .371$ and participants $F(1,28) = 4.07$, $p = .053$, and participants were 22 ms slower in responding to sentences with the only possible syntactic completion point at the end of the question.

3.2. Event-related potentials

After artifact removal, the RP was plotted for the electrode Cz. The trials rejected because of artifacts (18%) were distributed almost equally across all conditions.

Fig. 1A shows a typical RP slow negative waveform peaking during the response onset. The waveform from the average of all responses shows a visible negative deflection at approximately −800 ms and crosses the threshold of $1\ \mu V$ at 375 ms before the response onset. As shown in Fig. 1B, the RP of the responses with undelayed RT reached the threshold at 370 ms before the response-onset, and the RP of the delayed responses reached the threshold at RT 380 ms before for the response-onset. The 10 ms difference between the two RPs was not significant for sentences ($F_{corrected}(1,23) = 0.026$, $p > .1$) and participants ($F_{corrected}(1,28) = 0.278$, $p > .1$). In responses to questions with short mean RT, the RP reached the threshold 297 ms after the question offset, whereas in responses to questions with a long mean RT, the RP reached the threshold 605 ms after the question offset by factoring in the RT. As shown in Fig. 1C, the RP of responses to questions with a syntactic completion point only at the end reached the threshold at 250 ms before the response-onset, and the RP of responses to questions with a possible syntactic completion point within the question reached the threshold 535 ms before the response-onset. The 285 ms difference between the two RPs was significant for sentences ($F_{corrected}(1,23) = 56.562$, $p < .01$) and participants ($F_{corrected}(1,28) = 24.786$, $p < .01$). By factoring in the RT, in responses to questions with a syntactic completion point only at the end, the RP reached the threshold 553 ms after the question offset. In responses to questions with a possible syntactic completion point within the question, the RP reached the threshold 290 ms after the question offset. No RP-onset had an amplitude above the threshold of $1\ \mu V$ at the first syntactic completion point within the question. Taken together, the RP-onset to response-onset interval in spoken responses seems to be related to the language preparation process rather than the decisional process of what to say. However, the interval can be modulated by changes in the syntactic structure of the stimulus question. Fig. 2 illustrates the deviating RT of the two cognitive load conditions with the same RP onset to response onset interval, whereas in the two syntactic structures, the RT was the same but was accompanied with different RP onset to response onset intervals.

4. Discussion

The aim of this study was to provide information on the process of response-utterance preparation and its temporal aspects in turn-taking. The first manipulated variable was the cognitive load for response preparation; the second was the syntactic structure of the stimulus questions leading to an earlier or delayed response preparation. For all stimulus questions, participants had to perform the same degree of language comprehension, involving the retrieval of

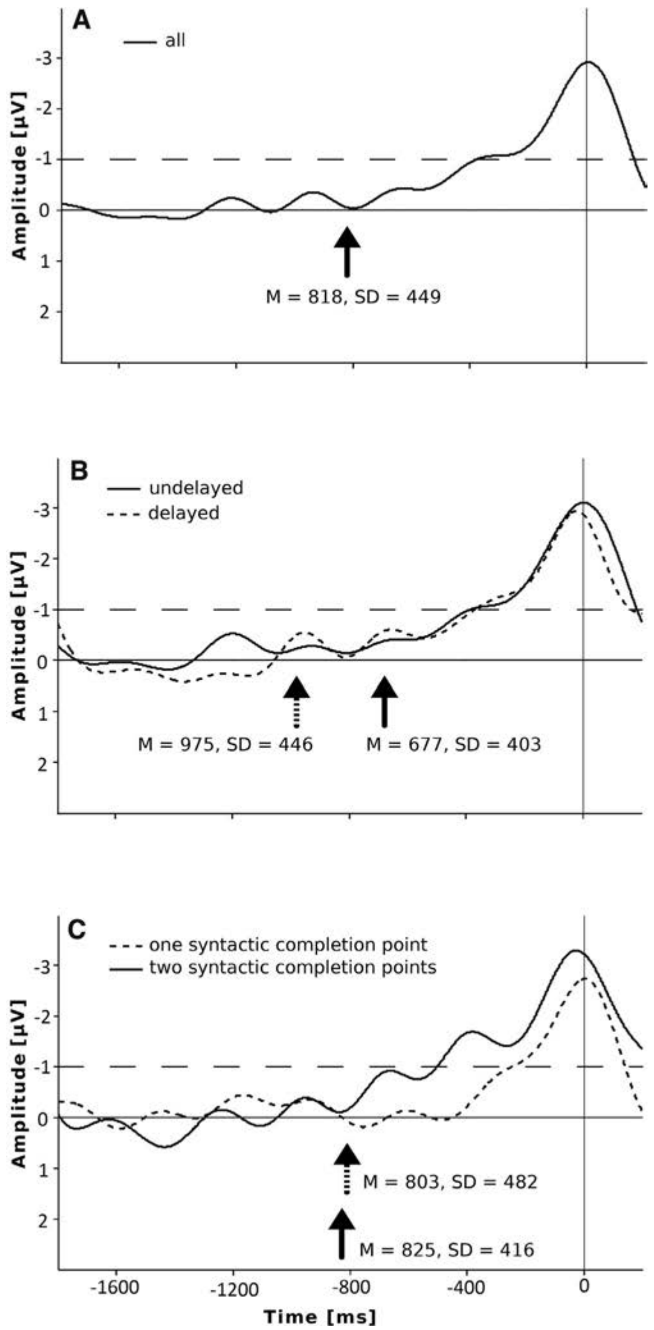

Fig. 1. Response-locked RPs at the electrode position Cz. In each diagram, time 0 indicates the articulatory response onset and the upper horizontal line represents the selected threshold of 1 μV. Below the waveforms are the mean (M) question ends with standard deviation (SD) indicated by solid and dashed arrows, which are respective to the waveform plotting. (A) Grand average of the RP of all responses to all questions. (B) Cognitive load: Grand average of the RP of responses divided into fast and slow RT. (C) Syntactic information: Grand average of the RP of responses divided into questions that had one or two syntactic completion points.

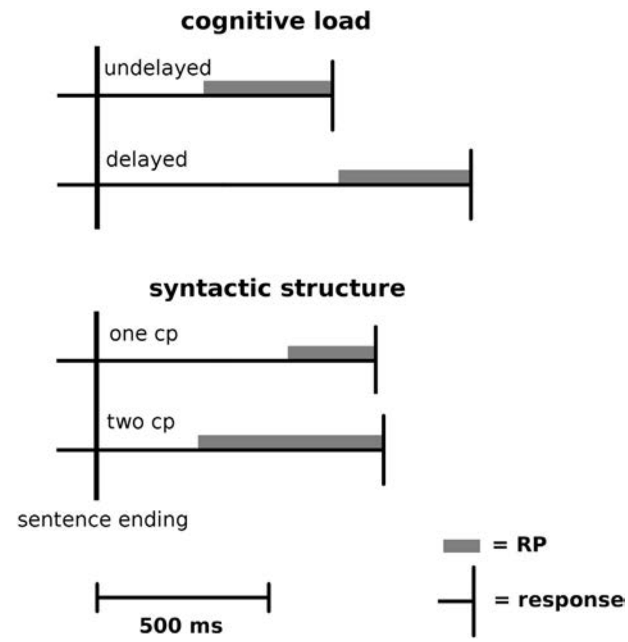

Fig. 2. Illustration of schematic temporal sequences of the RP-onset and the response onset after stimulus sentence ending. Time-locked to the stimulus sentence ending at the continuous vertical lines, with the RP onset indicated by the grey bar and response onset indicated by black vertical lines.

lexical information to get a coherent representation out of the utterance. In using electroencephalography in language production, it is important to apply comparable overt output due to the variety of morphological speech artifacts associated with phonetic properties of the utterance beginning [1]. Nonetheless, an advantage of the current study is the investigation of psychophysiological correlates time-locked to vocal onset with equally distributed phonetic variation of the response onset by using a voice activated trigger. This approach gives us the opportunity to access the time interval of interest which is located before the response onset which is evidenced by the progression of the RP to the response utterance onset. The number of events for each condition was low for signatures of higher cognitive processes (e.g. N400, P600) but was sufficient for the detection of activities of motor preparation, especially in conjunction with a relatively high number of participants. A RP lateralization effect has not been found due to bilateral motor preparation in speech production compared to single limb movements.

It seems that this artificial question-answer-situation with a mean RT of 815 ms does not correspond with Heldner and Edlund's meta-analysis of natural conversational corpora [23] on turn-taking; their study showed different results with a mean RT of −460 ms to 644 ms. This deviation may have been caused by the experimental paradigm in which participants had to answer multiple varied questions instead of participating in a conversation with a constant context and subject as a global semantic factor. Additionally, the situation that the participants had to answer to audio-recorded questions impacted the RT. In this situation, the question cannot be repeated in case of interruption such as with an interlocutor. Therefore, the participants had to make sure to hear the questions completely.

The mean RT of the response interacted significantly with the preliminary rating of the stimulus sentences. Overall, the questions asking for simple facts (low cognitive load) were answered without delay, whereas questions asking for a personal opinion or demanding more thought to answer (high cognitive load) were answered with a short delay. However, in undelayed and delayed answers, the response locked RP-onset was similar, and no significant difference of the RP could be found (see Fig. 1B). In both conditions, if the response is predetermined (low cognitive load) and in answers in which participants have to think about what to say (high cognitive load), the RP onset to speech onset interval was the same. Considering the latency of the speech production processes and because the comprehension of the questions and speech preparation after the response decision were the same in both conditions and only higher cognitive processes of planning a response were different, the RP-onset must be correlated with the time-point when the participants decided and began to prepare their response.

Moreover, our behavioral data showed no significant difference in the RT between the responses to sentences containing a first possible syntactic completion point within the question and the responses to questions with the only syntactic completion point at the end. Yet, the ERP results do show a significant earlier RP-onset in answers to questions with a first possible syntactic completion point within the question (see Fig. 2). In manipulating only the question comprehension process with deviating syntactic structure, a premature decision of what to say, gives more time for speech preparation. This process is caused by an earlier completion point and indicated by an earlier RP-onset. This result leads us to the assumption that variation in syntactic information can lead to different speeds of speech comprehension. Although there was an early possible syntactic completion point within the ques-tion, no RP-onset was found at that moment. This finding indicates that participants did not intend to respond at this early syntactic completion point. Although anticipation plays an important role in comprehension, it prosodic characteristics of the question could have signaled that the question was not completed at this point [24].

In exploring the two results of this study in combination with previous research on the RP development turn-taking [7], response preparation of formulated sentences delays the RP-onset compared to the simple verbal cue "Ja" (yes) as a response. However, both short and long RTs caused the same RP-onset to response interval, indicating different time intervals for the response decision, but the same intervals for response preparation (lexical selection, lemma and phonological retrieval).

5. Conclusion

We investigated whether response utterance preparation-related RP-onset was more related to the pure intention to respond or to the actual preparation of the response and whether it could be influenced by syntactic information of the stimulus question. Unde-layed and delayed answers are accompanied by the same RP-onset time to response-onset interval, suggesting that the RP-onset is more related to the actual speech preparation process than the pure intention to speak. The RP-onset of the response to a question with the only syntactic completion point at the end, however, was sig-nificantly delayed compared to that of responses to questions with two possible syntactic completion points. This evidence suggests an earlier response preparation due to earlier turn-anticipation.

Acknowledgments

Supported by the Deutsche Forschungsgemeinschaft (SFB 673). We thank Lorenz Sichelschmidt and two anonymous reviewers for their comments.

References

[1] H. Sacks, E.A. Schegloff, G. Jefferson, A simplest systematics for the organization of turn-taking for conversation, Language 50 (1974) 696–735.
[2] T. Stivers, N.J. Enfield, P. Brown, C. Englert, M. Hayashi, T. Heinemann, et al., Universals and cultural variation in turn-taking in conversation, Proc. Natl. Acad. Sci. 106 (2009) 10587–10592.
[3] P. Indefrey, The spatial and temporal signatures of word production components: a critical update, Front. Psychol. 2 (2011).
[4] L.Y. Ganushchak, I.K. Christoffels, N.O. Schiller, The use of electroencephalography in language production research: a review, Front. Psychol. 2 (2011).
[5] B. Habets, B.M. Jansma, T.F. Münte, Neurophysiological correlates of linearization in language production, BMC Neurosci. 9 (2008) 77.
[6] L. Magyari, M.C.M. Bastiaansen, J.P. de Ruiter, S.C. Levinson, Early anticipation lies behind the speed of response in conversation, J. Cogn. Neurosci. 26 (2014) 2530–2539.
[7] S. Jansen, H. Wesselmeier, J.P. de Ruiter, H.M. Mueller, Using the readiness potential of button-press and verbal response within spoken language processing, J. Neurosci. Methods 232 (2014) 24–29.
[8] H.H. Kornhuber, L. Deecke, Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale, Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere. 284 (1965) 1–17.
[9] J.J. McArdle, Z. Mari, R.H. Pursley, G.M. Schulz, A.R. Braun, Electrophysiological evidence of functional integration between the language and motor systems in the brain: a study of the speech bereitschaftspotential, Clin. Neurophysiol. 120 (2009) 275–284.
[10] H. Wesselmeier, S. Jansen, H.M. Müller, Influences of semantic and syntactic incongruence on readiness potential in turn-end anticipation, Front. Hum. Neurosci. 8 (2014).
[11] C. Riest, A.B. Jorschick, J.P. de Ruiter, Anticipation in turn-taking: mechanisms and information sources, Front. Psychol. 6 (2015).
[12] M.J. Sjerps, A.S. Meyer, Variation in dual-task performance reveals late initiation of speech planning in turn-taking, Cognition 136 (2015) 304–324.
[13] R.C. Oldfield, The assessment and analysis of handedness: the edinburgh inventory, Neuropsychologia 9 (1971) 97–113.
[14] H. Jasper, The ten–twenty electrode system of the international federation, Electroen. Clin. Neurophysiol. 10 (1958) 371–375.
[15] A. Delorme, S. Makeig, EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis, J. Neurosci. Methods 134 (2004) 9–21.
[16] A. Delorme, T. Sejnowski, S. Makeig, Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis, Neuroimage 34 (2007) 1443–1449.
[17] T.P. Jung, S. Makeig, C. Humphries, T.-W. Lee, M.J. Mckeown, V. Iragui, et al., Removing electroencephalographic artifacts by blind source separation, Psychophysiology 37 (2000) 163–178.
[18] T.P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, T.J. Sejnowski, Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects, Clin. Neurophysiol. 111 (2000) 1745–1758.
[19] F.T.Y. Smulder, J.L. Kenemans, A. Kok, Effects of task variables on measures of the mean onset latency of LRP depend on the scoring method, Psychophysiology 33 (1996) 94–205.
[20] D. Bryce, D. Sz ˝ucs, F. Soltész, D. Whitebread, The development of inhibitory control: an averaged and single-trial lateralized readiness potential study, Neuroimage 57 (2011) 671–685.
[21] J. Miller, T. Patterson, R. Ulrich, Jackknife-based method for measuring LRP onset latency differences, Psychophysiology 35 (1998) 99–115.
[22] R. Ulrich, J. Miller, Using the jackknife-based scoring method for measuring LRP onset effects in factorial designs, Psychophysiology 38 (2001) 816–827.
[23] M. Heldner, J. Edlund, Pauses, gaps and overlaps in conversations, J. Phonetics 38 (2010) 555–568.
[24] H. Koiso, Y. Horiuchi, S. Tutiya, A. Ichikawa, Y. Den, An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs, Lang. Speech 41 (1998) 295–321.