cortex 93 (2017) 28–40

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Research Report

Integrating speech in time depends on temporal expectancies and attention

Mathias Scharinger [a,b,c,*], *Johanna Steinberg* [b,d] *and Alessandro Tavano* [b,c]

[a] *Phonetics Research Group, Department of German Linguistics, University of Marburg, Pilgrimstein 16, 35032 Marburg, Germany*
[b] *Institute of Psychology, Leipzig University, Germany*
[c] *Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany*
[d] *Institute for Special Education, Leibniz Universität Hannover, Germany*

ARTICLE INFO

Article history:
Received 13 October 2016
Revised 11 January 2017
Accepted 1 May 2017
Action editor Peter E. Turkeltaub
Published online 19 May 2017

Keywords:
Mismatch negativity
Prediction
Omission
Speech
Temporal integration

ABSTRACT

Sensory information that unfolds in time, such as in speech perception, relies on efficient chunking mechanisms in order to yield optimally-sized units for further processing. Whether or not two successive acoustic events receive a one-unit or a two-unit interpretation seems to depend on the fit between their temporal extent and a stipulated temporal window of integration. However, there is ongoing debate on how flexible this temporal window of integration should be, especially for the processing of speech sounds. Furthermore, there is no direct evidence of whether attention may modulate the temporal constraints on the integration window. For this reason, we here examine how different word durations, which lead to different temporal separations of sound onsets, interact with attention. In an Electroencephalography (EEG) study, participants actively and passively listened to words where word-final consonants were occasionally omitted. Words had either a natural duration or were artificially prolonged in order to increase the separation of speech sound onsets. Omission responses to incomplete speech input, originating in left temporal cortex, decreased when the critical speech sound was separated from previous sounds by more than 250 msec, i.e., when the separation was larger than the stipulated temporal window of integration (125–150 msec). Attention, on the other hand, only increased omission responses for stimuli with natural durations. We complemented the event-related potential (ERP) analyses by a frequency-domain analysis on the stimulus presentation rate. Notably, the power of stimulation frequency showed the same duration and attention effects than the omission responses. We interpret these findings on the background of existing research on temporal integration windows and further suggest that our findings may be accounted for within the framework of predictive coding.

* *Corresponding author*. Phonetics Research Group, Department of German Linguistics, University of Marburg, Pilgrimstein 16, 35032 Marburg, Germany.
E-mail address: mathias.scharinger@staff.uni-marburg.de (M. Scharinger).
http://dx.doi.org/10.1016/j.cortex.2017.05.001

1. Introduction

Human perception involves the transformation of sensory information into more abstract, perceptual units (Massaro, 1987). Acoustic inputs with considerable temporal extent such as speech require efficient neural mechanisms to obtain an optimal chunking of a continuous acoustic signal (Giraud & Poeppel, 2012; Luo & Poeppel, 2012). During speech perception, chunking may occur on several levels (e.g., relating to sounds, syllables, words, or sentences (Everaert, Huybregts, Chomsky, Berwick, & Bolhuis, 2015) and consequently will require differently sized temporal windows for successful integration at different time scales. In speech, two distinct window sizes have been established: a short one (20–40 msec), corresponding to phonemes, and a long one (125–250 msec), corresponding to syllables (cf. Poeppel, 2003).

Neural evidence for chunking mechanisms has been provided by studies with a focus on the brain's oscillatory dynamics, as measured by frequency-domain analyses of the ongoing electroencephalographic (EEG) signal during auditory processing (Giraud & Poeppel, 2012; Hickok, Farahbod, & Saberi, 2015; Luo & Poeppel, 2012; Peelle & Davis, 2012). This research has shown that temporal windows of integration establish a mechanism of recurrently sampling auditory information in time (Luo & Poeppel, 2012; Poeppel, 2003). Importantly, attention to auditory information capitalizes on this mechanism and tightens the link between the (quasi)-periodic auditory signal and the corresponding periodic electric activity of the brain: neural stimulus tracking is stronger under attention (Lakatos et al., 2008).

The importance of the long, i.e., 125–250 msec integration window for auditory processing, has also been established by means of the brain's automatic change detection response, the mismatch negativity (MMN; Näätänen et al., 1997). The MMN is elicited by deviances in acoustic stimulation and indexes the violation of a prediction generated by the repeated presentation of a so-called standard stimulus. Importantly, the MMN is also elicited when the deviance corresponds to stimulus omission (*omission MMN* [oMMN], Winkler, Takegata, & Sussman, 2005). Yabe, Tervaniemi, Reinikainen, and Näätänen (1997) found that the oMMN response to omissions in tone sequences depended on the temporal separation between tone onsets. The omission of tones elicited a significant oMMN when tone onsets were separated by 125 msec or less. However, the omission of tones elicited no oMMN when tone onsets were separated by more than 125 msec. This result was taken as evidence for a temporal window of integration constraining the encoding of two successive tones as one-unit object, with a maximal window size of about 125–150 msec. Further studies suggest a more flexible window size depending on experimental task, stimulus material and expertise of the participants (e.g., Lee and Noppeney, 2014; Wang, Lin, Zhou, Pöppel, & Bao, 2015; Weise, Grimm, Trujillo-Barreto, & Schröger, 2014).

Recently, Bendixen et al. (2014) showed that speech sound omissions, commonly observed in natural speech (e.g., Zimmerer, Scharinger, & Reetz, 2011), also elicited oMMNs. Bendixen et al. (2014) examined the omission of word-final consonant clusters, temporally separated by 202 msec from word-onset, in the German nouns *Lachs* [laks] (salmon) and *Latz* [lats] (bib). Given that the authors obtained significant oMMNs with a temporal separation between word onset and consonant cluster of more than 125 msec, it seems that integration windows in speech must be more flexible and expandable compared to non-speech processing of simple tones. This interpretation is corroborated by the observation that speech sound, syllable, and word durations are inherently variable and additionally depend on speaking rate (Loukina et al., 2011; Reinisch et al., 2011). Hence, it remains to be empirically tested whether a more substantial elongation of words and the corresponding temporal separation of their sound onsets would still result in significant oMMNs when a specific speech sound is omitted.

Importantly, the size of the temporal integration window was established by means of passive oddball paradigms, during which participants' attention is not controlled (Yabe et al., 1997, 1998; Yabe, Koyama, et al., 2001; Yabe, Winkler, et al., 2001). The constraints of temporal integration windows might change with differing degrees of attention, also considering that previous research showed attention-related effects on the MMN (e.g., Auksztulewicz & Friston, 2015). Furthermore, attention usually increases auditory evoked responses (Näätänen, 1990; Picton and Hillyard, 1974). Therefore, attention on stimuli might either relax the constraining effects of temporal integration windows or, by contrast, sharpen the expectation that the integration window be not violated. Altogether, attention might be a moderator of the window-size effect on the oMMN in speech processing.

Here, we tested the effects of word duration (and temporal speech sound separation) as bottom-up, and attention as top-down factor in modulating the amplitude of the oMMN. We contrasted word-final consonant cluster omissions occurring 200 msec after word-onset with those occurring 400 msec after word-onset. We refer to the 200 msec separation as *natural* word duration [same as used in Bendixen et al. (2014)], where the omission occurs 150 msec after preceding speech sound onset. The 400 msec separation is termed *prolonged* word duration, where the omission occurs 300 msec after preceding speech sound onset, i.e., clearly exceeding a 125–150 msec window. Given that speech sound onsets in words are usually not clearly separated from preceding sounds, we assume that it is most likely that integration windows are “opened” at the beginning of each word presentation in our study. Effectively, then, we attempt to test whether the further separation of two sound sequences beyond the one reported in Bendixen et al. (2014) would still yield oMMNs if the second sequence was omitted.

Stimuli with omissions in natural-duration and prolonged-duration contexts were either actively processed or not. Active versus passive processing was intended to modulate attention, with relatively more attention deployed during active than during passive processing. Given the results of Bendixen et al. (2014) that clearly suggest a more flexible and expandable window than the one proposed by Yabe and colleagues, we hypothesized that an increase in word duration would yield a decrease, but not necessarily a lack of the oMMN, while attention would result in an increase of the oMMN, paralleling previous MMN findings (e.g., Auksztulewicz & Friston, 2015), and furthermore interact with the manipulation of word duration.