Brain and Cognition 111 (2017) 1–12

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

Switching of auditory attention in "cocktail-party" listening: ERP evidence of cueing effects in younger and older adults

Stephan Getzmann *, Julian Jasny, Michael Falkenstein

Leibniz Research Centre for Working Environment and Human Factors, D-44139 Dortmund, Germany

ARTICLE INFO

Article history:
Received 7 April 2016
Revised 28 June 2016
Accepted 13 September 2016
Available online 1 November 2016

Keywords:
Aging
Speech perception
Attention
Cueing
Event-related potentials

ABSTRACT

Verbal communication in a "cocktail-party situation" is a major challenge for the auditory system. In particular, changes in target speaker usually result in declined speech perception. Here, we investigated whether speech cues indicating a subsequent change in target speaker reduce the costs of switching in younger and older adults. We employed event-related potential (ERP) measures and a speech perception task, in which sequences of short words were simultaneously presented by four speakers. Changes in target speaker were either unpredictable or semantically cued by a word within the target stream. Cued changes resulted in a less decreased performance than uncued changes in both age groups. The ERP analysis revealed shorter latencies in the change-related N400 and late positive complex (LPC) after cued changes, suggesting an acceleration in context updating and attention switching. Thus, both younger and older listeners used semantic cues to prepare changes in speaker setting.

1. Introduction

Speech comprehension in the presence of noise or concurrent talkers is one of the most challenging tasks for the auditory system. Especially older adults have difficulties in the so-called cocktail-party situation (Burke & Shafto, 2008), and often failed to understand what has been said in a multi-speaker conversation (for review, Wingfield & Stine-Morrow, 2000). These deficits are partly based on age-related changes in peripheral hearing (e.g., presbycusis) and in central auditory processing (Humes & Dubno, 2010). In addition, speech perception deficits have been related to declines in general cognitive abilities, such as working memory capacity, inhibitory control, and information processing speed (Van der Linden et al., 1999; for review, see Schneider, Pichora-Fuller, & Daneman, 2010). According to the inhibition deficit hypothesis (Hasher & Zacks, 1988), for example, reduced inhibitory control might lead to attentional deficits, resulting in a higher distraction by concurrent auditory stimuli in older, than younger, adults.

In general, successful speech perception in a cocktail-party situation depends on auditory scene analysis, consisting of auditory object formation and segregation (Bregman, 1994), and on focusing of auditory attention on the speaker of interest (for review, see Shinn-Cunningham, 2008). Auditory scene analysis and focusing of attention appear to be (at least partly) object-based (for review, see Alain & Arnott, 2000). According to the object-based hypothesis of auditory attention, an auditory object is defined as the coherent whole of a group of sounds emanating from a single source. Thus, using perceptual grouping principles (Bregman, 1994), we can selectively focus our attention to perceptual objects derived from the auditory scene. There is evidence that the processes underlying auditory scene analysis and attentional orienting are affected by aging. In particular, concurrent stream segregation, i.e., the ability to segregate speech sounds from concurrently presented sounds, appears to be impaired (for review, see Alain, Dyson, & Snyder, 2006), whereas sequential auditory scene analysis is obviously preserved in older adults (Snyder & Alain, 2007). Using behavioral and electrophysiological measures, Snyder and Alain (2005) found age-related differences in concurrent vowel segregation, indicating that older adults were less able than younger ones to use vocal differences between talkers to separate their speech streams. Older adults needed more time for stream segregation in the presence of concurrent speech than younger ones (Getzmann & Näätänen, 2015). Furthermore, using event-related potential (ERPs) measures and a dynamic multi-talker speech perception task, it could be demonstrated that attentional control (i.e., the switching of attention between different speakers) was delayed in older adults (Getzmann, Wascher, & Falkenstein, 2015a).

A number of imaging studies aimed at the neural substrates of auditory attention in speech processing. These studies revealed that the initiation of an auditory attentional shift resulted in

* Corresponding author at: Leibniz Research Centre for Working Environment and Human Factors, Ardeystraße 67, D-44139 Dortmund, Germany.

E-mail address: getzmann@ifado.de (S. Getzmann).

http://dx.doi.org/10.1016/j.bandc.2016.09.006
0278-2626/

cortical activity changes mainly in superior parietal lobus (SPL), in temporo-parietal junction (TPJ), and some regions of prefrontal cortex (PFC) (e.g., Larson & Lee, 2014; Shomstein & Yantis, 2006). Notably, these regions are also involved in goal-directed attentional control in vision (Corbetta & Shulman, 2002; Serences & Yantis, 2006), suggesting a supramodal mechanism of attention shifting (Corbetta, Patel, & Shulman, 2008). Interestingly, different neural mechanisms of attentional control (i.e., the search of a task-relevant auditory object in a complex auditory scenery) and attentional selection (i.e., the focussing of attention toward that object of interest and the inhibition of concurrent sound) have been observed: While attentional control activated a left-dominated, fronto-parietal network, mainly comprising inferior frontal gyrus (IFG), SPL, and intraparietal sulcus (IPS), attentional selection recruited a network of cortical areas that include bilateral superior temporal gyrus (STG), superior temporal sulcus (STS), and IPS (Alho, Rinne, Herron, & Woods, 2014; Hill & Miller, 2010). In particular, the left IPS is assumed to be an integrative, multi-modal center that coordinates attention. Comparing brain networks processing (uncued) involuntary and (cued) voluntary auditory attention shifts revealed that areas most selectively involved with cued shifting included the superior/posterior IPS, SPL, and some precentral areas, indicating these regions' involvement in top-down/voluntary attentional control (Huang, Belliveau, Tengshe, & Ahveninen, 2012; Rossi, Huang, Furtak, Belliveau, & Ahveninen, 2014). Taken together, these neurophysiological findings indicate a distributed cortical network of attention shifting with the involvement of different areas depending on stimuli and task-relevant factors.

There is evidence that auditory scene analysis and focusing of attention are time-consuming (Cusack, Decks, Aikman, & Carlyon, 2004; Larson & Lee, 2013a; Shinn-Cunningham & Best, 2008; for electrophysiological evidence, see Kerlin, Shahin, & Miller, 2010; for a review, see Fritz, Elhilali, David, & Shamma, 2007). Using ERP measures, it has been demonstrated that the attention-dependent process that integrates successive tones within streams takes several seconds to build up (Snyder, Alain, & Picton, 2006). Accordingly, continuity in the auditory scenery resulted in a more efficient object selection and focussing of attention toward the features of a relevant source, whereas changes in the scenery (e.g., switches in speaker locations and voices) require renewed object formation/segregation and attentional re-focussing (Best, Ozmeral, Kopco, & Shinn-Cunningham, 2008). Moreover, deliberate switches from one speaker of interest to another, previously unattended speaker have been found to produce switch costs, indicated by higher error rates and reaction times after a change, relative to the pre-change level (Koch, Lawo, Fels, & Vorländer, 2011).

This could also be demonstrated in two recent studies, in which listening costs associated with shifts in spatial attention were tested in dynamic multi-speaker environments. Using conversational turn-takings in a sentence recall task, a pronounced decrease in word recall was found when a target speaker switched location while talking, relative to a non-switching condition (Lin & Carlile, 2015). Significant correlations between switching performance and individual cognitive measures were found, indicating that conversational tracking requires cognitive resources. In particular, working memory, required for maintaining task-relevant information and extracting meaning in adverse listening conditions (Baddeley, 2003; Engle & Kane, 2004), appeared to be essential for speech perception in conversational turn-takings. Using electrophysiological measures, we recently investigated the effect of changes in speaker settings on speech perception of younger and older adults (Getzmann, Hanenberg, Lewald, Falkenstein, & Wascher, 2015b). In a simulated stock-market scenario, sequences of short words (combinations of company names and values) were simultaneously presented by four speakers at different locations in space, and the participants responded to the value of a target company. While continuity in the target speakers' voice and location resulted in improved performance, an unexpected and unpredictable change in target speaker caused an increase in error rates. This holds even more so in older participants, who needed more time to re-achieve the pre-change level of performance.

The analysis of the event-related potentials (ERPs) before and after a change revealed a $N400_{diff}$ and a late positive complex (LPC_{diff}) over parietal areas in both age groups. A similar pattern of activation has recently been reported in a multi-talker word-pair semantic categorization task in which younger and middle-aged adults responded to an attended stream of words, while ignoring competing speech from a different location (Davis & Jerger, 2014; Davis, Jerger, & Martin, 2013). These components have been interpreted as correlates of increased evaluation of stimulus meaning and context updating as well as attention switching, elicited by a change of the target speaker (Davis & Jerger, 2014; Getzmann, Hanenberg et al., 2015b). Also, similar pattern of N400 and LPC are usually observed after language switching, i.e., when speech information switches from one language to another different language, which has been related to in increased processing at the level of the lexico-semantic system and updating or reanalysis process (e.g., Moreno, Federmeier, & Kutas, 2002; Van Der Meij, Cuetos, Carreiras, & Barber, 2011). Taken together, changes in speaker settings in multi-speaker environments appear to be critical events in speech perception for younger and older adults that are associated with increased cortical processing.

It has to be noted, however, that changes in speaker settings are usually not completely unexpected for the listener. In a realistic conversation, a change from one speaker of interest to another may be indicated by modulations in prosodic features and intonation as well as by the semantic content. For example, the structure of an interrogative sentence typically indicates a subsequent change between speakers, i.e., from the questioner to the potentially answering person. Likewise, the semantic end of a sentence might indicate that another speaker will probably continue the conversation. Hence, the question arises whether listeners actively use this information to prepare a switch in attention. This preparation has the potential to improve object re-selection and attentional re-focussing and may thereby reduce the switch costs after a change.

In the present study, we investigated effects of cueing of an upcoming change in speaker setting on speech perception of younger and older adults. We therefore modified our previous stock-price monitoring task and introduced a condition in which an incoming change of the target speaker was indicated by a cue. This verbal cue was specified by task-relevant semantic content, i.e., by a keyword consisting of a specific company value. By comparing the performance in speech perception after a cued vs. uncued change we tested whether cueing information helps managing dynamic multi-speaker situations, and whether younger and older adults may differ in the use of cueing information. The analysis of the ERPs focused on the change-related components observed in previous studies (Davis & Jerger, 2014; Davis et al., 2013; Getzmann, Hanenberg et al., 2015b). In particular, we tested whether cued changes are associated with differences in the $N400_{diff}$ and LPC_{diff} amplitudes and latencies. In addition, we analyzed possible effects of cueing and age on the ERPs that are typically elicited by speech onset. Here, the most pronounced effects were expected for the N2 component. The N2 is usually assumed to reflect inhibitory control and suppressing irrelevant information (Folstein & Van Petten, 2008), and has been found to be reduced in speech perception in older adults (Getzmann, Lewald, & Falkenstein, 2014; Getzmann, Wascher et al., 2015a). Finally, we investigated whether younger and older adults may differ in cue processing by analyzing the ERPs elicited by the cue stimulus.