Speech Communication 85 (2016) 1–7

Contents lists available at ScienceDirect

Speech Communication

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

Effects of spoken Thai word-durations on brain recognition processing: An auditory event-related potential study

Chailerd Pichitpornchai*, Kittipun Arunphalungsanti

Department of Physiology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand

ARTICLE INFO

Article history:
Available online 13 October 2016

Keywords:
Word-duration
Recognition
Auditory event-related potential
P300, N400
Reaction time

ABSTRACT

The present study investigated the effects of spoken Thai word-durations on auditory event-related potentials (aERPs) and reaction time (RT) in a recognition task. Thirty healthy graduate students were recruited. Stimuli were monosyllabic words recorded at short (421.1 ± 82.6 ms), normal (635.8 ± 99.8 ms), and long (847.0 ± 130.3 ms) durations. Participants listened to ten blocks of word stimuli that included a study phase and test phase in each block. In each study phase, participants were instructed to memorize six randomly presented words (the targets; three short-duration and three long-duration). In the test phase, they were instructed to indicate whether words matched the targets from the study phase. Stimuli in the test phase were presented at normal word duration and included the 6 test-phase targets and 14 new randomly presented words (non-targets). Reaction time and aERPs (P300 and N400 components) were analyzed. Performance accuracy and P300 amplitude were higher for target words with long durations than for those with short durations. However, there were no significant differences between short-target and long-target words for either reaction time or the N400 component. Additionally, we found that long-target words elicited high P300 and N400 amplitudes in the fronto-central and frontal areas, respectively. This study demonstrated that a longer than normal word duration can facilitate word processing and recognition. Additionally, our results indicate that ERPs can be used to develop valuable clinical tests that evaluate frontal lobe function and help improve instructional media for speech recognition.

1. Introduction

Speech recognition is the ability to match words that we hear with words that we heard in the recent past. Recognition includes two fundamental stages of processing. The first is a pre-lexical stage in which an abstract description of a given utterance is generated. The second is a lexical stage involving the activation of many candidate words (McQueen, Cutler, and Norris, 2003). Speech recognition is typically studied with recognition tasks that contain a study phase and a test phase (Campeanu, Craik, and Alain, 2013; Curran, 1999, 2000; Van Strien, Hagenbeek, Stam, Rombouts, and Barkhof, 2005; Zaske, Volberg, Kovacs, and Schweinberger, 2014). Individuals are asked to remember target words in the study phase and memory performance is evaluated in the test phase using the target words and previously unheard (non-target) words. Combined with analysis of brain electrical potential, the test phase is intended to assess the differential brain activity in response to targets and non-targets that occurs between 300 ms and 800 ms after test-phase words are presented. This “old/new” effect has been interpreted as the representation of recollection processes that occur when people try to remember the stimuli presented in the study phase, combined with the representation of intentional retrieval processes that occur when they consciously compare or discriminate stimuli in the test phase (Campeanu et al., 2013; Curran, 1999, 2000; Van Strien et al., 2005; Zaske et al., 2014). Speech recognition is necessary for understanding the meaning of words in syntactic contexts, and is one aspect of speech comprehension, a process that requires the timely integration of phonology, semantics, syntax, and prosody (Li, Chen, and Yang, 2011; Rodd, Davis, and Johnsrude, 2005). Understanding the meaning of individual words is a part of speech comprehension, which requires some sort of recognition. The speech recognition task in the present study used meaningful words (semantic content) spoken at different durations (prosodic content) as stimuli.

Behavioral and electrophysiological studies of speech processing typically measure reaction time (RT) and auditory event related potentials (aERPs). Reaction time is a behavioral measurement of how fast a response to a stimulus is (Miller and Low, 2001; Skurvidas, Mickevichiene, Cesnavichiene, Gutnik, and Nash, 2012). aERP analysis is used to find associations between brain activity and specific auditory events (Pilling, 2009; Polich, 1996) at a high temporal resolution. aERPs represented an immediate neural

* Corresponding author. Fax: +66 2 411 5009.
E-mail address: Chailerd.Pic@mahidol.ac.th (C. Pichitpornchai).

http://dx.doi.org/10.1016/j.specom.2016.10.003
0167-6393/

response recorded within 1000 ms after stimulus onset, whereas RT represented a behavioral response occurred between 700-900 ms. The combination of RT and aERP techniques have been widely used in the field of speech recognition (Bonte and Blomert, 2004; Desroches, Newman, and Joanisse, 2009; Dumay et al., 2001; Friedrich and Kotz, 2007; Friedrich, Schild, and Roder, 2009; Liu, Shu, and Wei, 2006; Zhao, Guo, Zhou, and Shu, 2011), In the present study, precise timing of brain activity (aERP) was needed for determining what happens in the brain after perceiving sound stimuli.

Previous ERP studies have found that a positive component occurs between 250 – 400 ms after the stimuli onset (Friedman, Nessler, and Johnson, 2007). The functional interpretation of the P300 component is that it reflects updating memories and a combination of processes that vary by task and situation, including active stimulus discrimination and response preparation. It is associated with concentrating on identifying the stimuli or with recognizing them. P300 latency is assumed to reflect the duration of stimulus evaluation. It has also attracted attention in clinical studies (Friedman et al., 2007). A study using a spoken word recognition-task found that the P300 component was modulated by the age of the participants (Tainturier, Tamminen, and Thierry, 2005).

In addition to P300, studies have found a negative component occurring between 250 – 500 ms after stimulus onset (the N400) that is associated with word recognition and language comprehension (Liu et al., 2006; Renoult and Debruille, 2011). N400 component was initially discovered by Kutas and Hillyard in 1980, which was the first report of an N400 response (Kutas and Hillyard, 1980). The amplitude of the N400 is most sensitive to semantic relations between individual words, or between words and their sentence and discourse context. A number of studies suggest that N400 amplitude is markedly reduced when stimuli are repeated or when words that match the semantics of a sentence are presented (Duncan et al., 2009; Martin, Tremblay, and Korczak, 2008; Renoult and Debruille, 2011; Toivonen and Rama, 2009). N400 has been shown to be elicited by a semantic mismatch between the meaning of a word and the semantic specification of its sentence context (Liu et al., 2006). Its amplitude has been shown to be smaller for words that are used frequently in daily speech than for those that are used infrequently (Hagoort, 2008). Furthermore, whether the N400 component reflecting integration or anticipation processes is still a debated issue. In the integration view, the word meaning can be integrated very rapidly within the prior sentence context using all available information (Hagoort, 2008; Van Petten and Luka, 2012). In the anticipation view, the prior context pre-activates the word representation in memory before it is actually presented so that upon presentation it's processing is facilitated through spreading of activation (DeLong, Urbach, and Kutas, 2005; Federmeier, 2007).

Sound duration has been shown to affect the timing and amplitude of aERPs. For example, duration of tone stimuli (200 – 300 ms) has been shown to correlate with peak N200 amplitude in both newborns and adults (Kushnerenko, Ceponiene, Fellman, Huotilainen, and Winkler, 2001). Additionally, a study on the effect of audible word length (depending on the number of syllables) showed that P200 amplitude increases with word length (Murphy, Roodenrys, and Fox, 2006). While these studies revealed ERPs related to sound length, neither experiment dealt with the meaning of speech. In contrast, brain functions represented by the P300 and N400 components include perception of audible sound characteristics, spoken-word recognition, and decision-making, making them reasonable targets when investigating aERPs in response to different speeds of speech.

In everyday life, fast speech is more difficult to understand than slow speech. The speed of speech can be modified by changing the duration with which words are spoken; shorter word duration generates faster speech and vice versa. A previous study about the effect of speaking rates on word recognition in young and elderly listeners showed that the younger and older subjects had decreased speech recognition in the fast speaking rate of monosyllabic words presented compared to medium or slow speaking rate conditions (Sommers, Nygaard, and Pisoni, 1994). It showed that the faster speaking rate can reduce the intelligibility in normal subjects. Moreover, another previous clinical study about the effect of speaking rate on word recognition in Parkinson's disease and normal aging showed that subjects in all three groups had lower identification scores for words presented in the fast rate condition than in the normal or slow rate conditions (Forrest, Nygaard, Pisoni, and Siemers, 1998). However, these studies did not demonstrate brain electrical activities occurring during word cognitions. Thus, studying the effect of word duration on brain activity can help us understand the neural mechanism underlying the effects of speaking quickly or slowly at the whole-sentence level.

The aim of this study was to examine whether differences in the durations of spoken Thai have any effect on aERPs or RT in a word recognition task. Specifically, we investigated the effects of monosyllabic word duration on recognition processing by evaluating differences in RT and the P300 and N400 components. We hypothesized that long duration of spoken Thai words in the recognition task would correlate with large aERP amplitudes and shorter latencies, and with shorter RTs.

2. Materials and Methods

2.1. Participants

Thirty right-handed graduated students (15 women and 15 men, aged 20 – 30 years) with normal hearing thresholds were recruited from the graduate school of the Faculty of Medicine, Siriraj Hospital, Mahidol University. No participant had a history of neurological illness, drug addiction, musical training, language-related disorders, or neuromuscular disorders of the hands or fingers. Before the experiment, each participant signed an informed consent form explaining the study details. In the data analysis, data from three subjects were excluded because of large number of artifacts contamination. The protocol was approved by the ethics committee of the Siriraj Institutional Review Board, Faculty of Medicine, Siriraj Hospital, Mahidol University.

2.2. Stimuli preparation

200 monosyllabic Thai content words were daily-life used words chosen from the Royal Institute Dictionary. They were all mid-lexical tone and the mean word frequency from Google search engine was 11.36 million items (SD 7.61 million items). The 200 chosen words were spoken by a native Thai 35 year-old female. Each word was originally recorded in three durations and slightly modified by using Adobe audition software (Adobe Systems Incorporated, USA) resulting in: normal (635.8 ± 99.8 ms), short (421.1 ± 82.6 ms), and long (847.0 7 ± 130.3 ms) duration words. The durations of these words were modified within 10 % of the originally pronounced words resulting in closed to natural heard words and they did not sound strange.

A one-way ANOVA showed that the three categories significantly differed in duration ($F(2, 118) = 241.87$, $p \leq 0.001$). All stimuli were adjusted to be 60 dB, and the inter-stimulus interval was 2000 ms.