Brain & Language 144 (2015) 16–25

Contents lists available at ScienceDirect

Brain & Language

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

Seriality of semantic and phonological processes during overt speech in Mandarin as revealed by event-related brain potentials

Xuebing Zhu [b,d], Markus F. Damian [c], Qingfang Zhang [a,b,*]

[a] Department of Psychology, Renmin University of China, Beijing, China
[b] Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China
[c] School of Experimental Psychology, University of Bristol, UK
[d] Institute of Linguistic Studies, Shanghai International Studies University, Shanghai, China

ARTICLE INFO

Article history:
Received 23 June 2014
Accepted 19 March 2015
Available online 13 April 2015

Keywords:
Speech production
Semantic interference
Phonological facilitation
Electroencephalography

ABSTRACT

How is information transmitted across semantic and phonological levels in spoken word production? Recent evidence from speakers of Western languages such as English and Dutch suggests non-discrete transmission, but it is not clear whether this view can be generalized to other languages such as Mandarin, given potential differences in phonological encoding across languages. The present study used Mandarin speakers and combined a behavioral picture–word interference task with event-related potentials. The design factorially crossed semantic and phonological relatedness. Results showed semantic and phonological effects both in behavioral and electrophysiological measurements, with statistical additivity in latencies, and discrete time signatures (250–450 ms and 450–600 ms after picture onset for the semantic and phonological condition, respectively). Overall, results suggest that in Mandarin spoken production, information is transmitted from semantic to phonological levels in a sequential fashion. Hence, temporal signatures associated with spoken word production might differ depending on target language.

1. Introduction

Speaking, as a highly skilled behavior in daily life, is marked by astonishingly high-speed retrieval from the mental lexicon and low error rates (Levelt, 1992). One of the key requirements of spoken production is to select an appropriate target at a given time and to focus the execution of goal-directed articulation. Over the past few decades, the speech production system has been envisaged as a system of interrelated layers of mental representations, such as semantic, syntactic and phonological codes (Caramazza, 1997; Dell, 1986; Levelt, Roelofs, & Meyer, 1999). A central theoretical issue concerns how information flows within this cognitive system and its underlying neural implementation in speech production (Abdel Rahman & Sommer, 2003; Caramazza, 1997; Starreveld & La Heij, 1995; Starreveld & La Heij, 1996a, 1996b). Serial-discrete models (Levelt et al., 1999) argue that only a single selected lexical-semantic/syntactic node ("lemma") spreads its activation to the phonological level, and semantic processing must be completed before phonological processing. Non-serial models dispute some of these assumptions: cascaded models (Humphreys, Riddoch, & Quinlan, 1988; Morsella & Miozzo, 2002) propose that multiple lexical-semantic candidates which are co-activated during retrieval of the target word transmit activation to the phonological level. Interactive models (Dell, 1986) additionally assume that transmission of activation between semantic and phonological encoding is bidirectional. In both cascaded and interactive but not in serial models, phonological processing can begin on the basis of early partial information provided by semantic processes.

Recent empirical findings provide support against a strictly serial view, and for some degree of cascadedness. For instance, in a task in which two line drawings of objects are superimposed and one is to be named based on its color, a facilitation effect is observed when target and distractor objects overlap in their phonemes (Meyer & Damian, 2007; Morsella & Miozzo, 2002; but see Jescheniak et al., 2009). Similarly, when the color of a line drawing is named while ignoring the object, priming is found when target color and object names overlap in their form (Kuipers & Heij, 2009; Navarrete & Costa, 2005; note that henceforth, we use the term "form" to refer to surface properties – sound or spelling – of lexical items). These findings suggest that multiple lexical candidates are phonologically activated, which contradicts a central tenet of the seriality view. At the same time, the existing evidence suggests that cascading is not "universal" such that all activated units at higher level necessary transmit activation to lower levels. For instance, Kuipers and Heij (2009; see also Dumay & Damian,

* Corresponding author at: Department of Psychology, Renmin University of China, 59 Zhongguancun Street, Haidian District, Beijing 100872, China.
E-mail address: qingfang.zhang@ruc.edu.cn (Q. Zhang).

http://dx.doi.org/10.1016/j.bandl.2015.03.007
0093-934X/

2011) have suggested a principle of "limited cascadedness" according to which properties associated with the identity of the target dimension (such as an object's name) will cascade to the form level, whereas modifying dimensions (such as its color or size) will not. Cascadedness might additionally be modulated by factors such as attention and task demands (Mädebach, Jescheniak, Oppermann, & Schriefers, 2011). Nevertheless, overall the currently available evidence on phonological activation of multiple lexical codes contradicts a strictly serial view of lexical access in spoken production, and suggests that at least under some circumstances, multiple lexical candidates can transmit activation to the phonological level.

A different way to tackle the issue of activation transmission between semantic and phonological stages in word production is to employ one of the most widely used paradigms in speech production, namely the picture–word interference (hereafter PWI) task. In this task, participants are instructed to name a target picture while ignoring a distractor word which is either visually superimposed on the target, or presented in spoken format. A semantic relationship between a context word such as "dog" and a target picture such as "cat" slows naming relative to an unrelated word (e.g., "table"), whereas a phonological relationship (e.g., "key") speeds up latencies (Glaser & Düngelhoff, 1984; Schriefers, Meyer, & Levelt, 1990; Starreveld & La Heij, 1995). These two phenomena have been termed "semantic interference" and "phonological facilitation", and numerous studies with PWI have shown that those effects provide important constraints on models of speech production. For instance, the prominent WEAVER++ model (Word-form Encoding by Activation and VERification; Roelofs, 1992, 1997) postulates that semantic interference arises at a processing stage of lexical-semantic retrieval, where co-activated entries ("lemmas") engage in competition with one another. By contrast, phonological facilitation arises mainly at the segmental level (with the possibility of weaker priming also arising at the morpheme level): distractor words activate corresponding segments, and therefore partially pre-activate the segments which form the target response, resulting in faster encoding for related than unrelated distractors.

The PWI task can be used to explore how semantic and phonological processing stages relate to each other, via employing not only the semantically and phonologically related distractors, but additionally, by including "mixed" distractors which are semantically as well as form-related (e.g., picture: "cat"; distractor: "calf"). Factorially crossing semantic and phonological relatedness allows to determine whether the two variables have statistically additive or interactive effects. Based on "additive factors logic" (Sternberg, 1969) the idea is that if the two experimental variables exert statistically additive effects, then they affect different and separate processing stages, with strictly serial information transmission between the stages. By contrast, if the two variables show non-additive effects, then either they act on a single processing stage, or they affect two processing stages but these two stages are themselves closely related in terms of processing, for instance, via cascaded transmission, or interactivity (i.e., feedback). The currently available results clearly demonstrate a statistical interaction between semantic and phonological relatedness in PWI tasks (Damian & Martin, 1999; Starreveld & La Heij, 1995, 1996b; Taylor & Burke, 2002). More specifically, the pattern that is typically found is that the semantic interference effect is attenuated when a distractor is also form-related to the target name; hence, "rabbit-rat" acts predominantly as a form-related pair whereas the semantic effect which should arise from shared category membership is much diminished. This general pattern has been interpreted as supporting non-discrete models of word production (Damian & Martin, 1999).

1.1. Event-Related Potentials (ERPs) and spoken word production

The bulk of evidence concerning spoken word production has traditionally come from chronometric studies. However, response latencies merely index the "end point" in a cascade of mental processes which precede initiation of a response. Hence questions associated with the time course of various types of mental activities (i.e., how processing stages unfold over time) are difficult to address with chronometry. A complementary approach is to employ electroencephalography (EEG). By tracking electrical activity along the scalp, brain responses to specific sensory, cognitive, or motor events can be assessed millisecond-by-millisecond as they unfold.

The EEG approach is well-established in various areas of language research. However, until relatively recently it was assumed that EEG could not be measured for spoken responses because artefacts from muscular activity associated with articulation distort the signal (Wohlert, 1993). Hence, many empirical studies used manual responses as a substitute for spoken ones (Van Turennout, Hagoort, & Brown, 1997; Zhang & Zhu, 2011). Yet, it has recently become clear that the problems associated with overt articulation are tractable, and a number of studies have combined spoken production tasks with EEG (e.g., Blackford, Holcomb, Grainger, & Kuperberg, 2012; for reviews, see Ganushchak, Christoffels, & Schiller, 2011; Indefrey, 2011; Strijkers & Costa, 2011) and MEG (magnetoencephalography, Levelt, Praamstra, Meyer, Helenius, & Salmelin, 1998; Maess, Friederici, Damian, Meyer, & Levelt, 2002; Salmelin, Schnitzler, Schmitz, & Freund, 2000). In these EEG studies, classical ERP components have been replicated during overt naming. For instance, the N400 complex, first reported by Kutas and Hillyard (1980) in semantic violations, is widely interpreted as a central index of lexical and semantic processing (for review, see Lau, Phillips, & Poeppel, 2008) and phonological processing (i.e., Chen, Lee, Kuo, Hung, & Cheng, 2010; Valdes-Sosa et al., 1993) in language comprehension. Importantly, this negative-ongoing waveform apparently also reflects phonological processing in spoken production (Blackford et al., 2012; Dell'Acqua et al., 2010), and hence indicates priming resulting from the convergence of phonological processing from pictures and distractors in the PWI task. Moreover, Dell'Acqua et al. (2010) used ERPs combined with a subtraction technique to explore the time course of activation of semantic and phonological representations in the PWI task. Difference ERP waveforms were generated in the semantic condition and in the phonological condition by subtracting ERP waveforms in the unrelated condition. In the time window of 250–450 ms, they found significant differences on mean amplitude for both semantic and phonological relatedness. Furthermore, the peak latencies of semantically related distractors (320 ms) coincided temporally with those of phonologically related distractors (321 ms). These estimates are difficult to reconcile with a strictly serial information transmission model (see previous section) which would predict a more sequential pattern.

We should note that in the still limited literature on EEG studies exploring spoken production, it is at present typical to focus on ERP differences between experimental and baseline conditions, rather than (or sometimes in addition to) identifying components such as N400, etc., which are associated with specific particular mental processes. Undoubtedly, this is the case because EEG research on production is relatively less well developed than corresponding research on comprehension.

1.2. Cross-linguistic differences in phonological encoding?

Much of our understanding of how speakers plan and produce words is based on evidence from Indo-European languages such