Toddlers' sensitivity to within-word coarticulation during spoken word recognition: Developmental differences in lexical competition

Tania S. Zamuner *, Charlotte Moore, Félix Desmeules-Trudel

Department of Linguistics, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada

A B S T R A C T

To understand speech, listeners need to be able to decode the speech stream into meaningful units. However, coarticulation causes phonemes to differ based on their context. Because coarticulation is an ever-present component of the speech stream, it follows that listeners may exploit this source of information for cues to the identity of the words being spoken. This research investigates the development of listeners' sensitivity to coarticulation cues below the level of the phoneme in spoken word recognition. Using a looking-while-listening paradigm, adults and 2- and 3-year-old children were tested on coarticulation cues that either matched or mismatched the target. Both adults and children predicted upcoming phonemes based on anticipatory coarticulation to make decisions about word identity. The overall results demonstrate that coarticulation cues are a fundamental component of children's spoken word recognition system. However, children did not show the same resolution as adults of the mismatching coarticulation cues and competitor inhibition, indicating that children's processing systems are still developing.

© 2016 Elsevier Inc. All rights reserved.

* Corresponding author.
  E-mail address: tzamuner@uottawa.ca (T.S. Zamuner).

Introduction

To understand speech, listeners need to be able to decode the speech stream. As native speakers of a language, this process happens so naturally and easily that in normal contexts it feels as though words are being provided to us in perfect condition. This sensation is a testament to the power of human speech perception because the phonetic input a listener receives is extremely variable, yet listeners are able to cope with this variability. Each individual sound in an utterance is affected by speaker characteristics such as speech rate, pitch, and environment (Lindblom, 1990). In addition, phonemes, when produced in speech, are affected by their surrounding sounds. Each phoneme's articulatory gesture temporally overlaps with the gestures of adjacent phonemes (Browman & Goldstein, 1989). Coarticulation causes phonemes to differ based on their context, but in most cases it does not cause a difference in phonemic category (Gow & McMurray, 2007). For example, the /k/ in *took* is articulated further back in the mouth than the /k/ in *teak*. Coarticulation can be progressive (i.e., carryover coarticulation), meaning that one segment influences a following sound, as well as regressive (i.e., anticipatory coarticulation), which is when a following segment influences a preceding sound (Flege, 1988). Vowel nasalization in English is a case of regressive coarticulation as a following nasal consonant influences a preceding vowel. Because coarticulation is an ever-present component of the speech stream, it seems plausible that listeners would exploit these cues to identify words. This research examines listeners' sensitivity to these cues and the development of the speech recognition system. In an eye-tracking study using a looking-while-listening paradigm, adults and children were tested on coarticulatory cues that either matched or mismatched the target. The goal was to observe participants' attempts at resolving these miscues and to observe the time course of word recognition as the system potentially predicts upcoming phonemes based on anticipatory coarticulation to make decisions about word identity.

Coarticulation is a by-product of the speech mechanism. Because the speech stream is continuous, coarticulation occurs between words as well as within them. The vocal tract positions itself for the next phoneme before the preceding one is completed. Although coarticulation can be attributed to physical constraints of the vocal tract in production, some models of spoken word recognition recognize that the variability it introduces into the speech stream is not noise to be filtered out. It can rather provide a source of information about the upcoming segments, exploitable by the listener, although how useful these cues may be or the relative importance of a given cue with regard to other phonetic cues (Miller & Dexter, 1988) remains an open question. Numerous studies with adults have found that listeners use subphonemic coarticulation cues to quickly make predictions about upcoming phonemes before they have been completely articulated, despite the fact that these coarticulation cues do not create changes that cross category boundaries in the phonemic system. Listeners exploit these cues both across word boundaries (Gow & McMurray, 2007; Salverda, Kleinschmidt, & Tanenhaus, 2014) and within words (Beddor, McGowan, Boland, Coetzee, & Brasher, 2013; Dahan, Magnuson, Tanenhaus, & Hogan, 2001; McQueen, Norris, & Cutler, 1999; Whalen, 1991). For example, Dahan et al. (2001) cross-spliced minimal pair CVC (consonant–vowel–consonant) words to create matching or mismatching coarticulatory cues. As one example, the word *bat* was in one of three forms: ba_tt, ba_gt, or ba_pt. The subscript letter indicates the coarticulated consonant, that is, the consonant that was originally pronounced in the word before cross-splicing and for which the characteristics can be found on the vowel (e.g., formant transitions). In a visual world paradigm with eye-tracking, participants reacted to tokens like ba_gt as if the subphonemically cued word (*bag*) had been presented instead of the target (*bat*), at least at the beginning of the word recognition process. This slowed down response times due to lexical competition between the subphonemically cued word (*bag*) and the phonemically cued word (*bat*). Similarly, Beddor et al. (2013) found that English-speaking adults treated the presence of vowel nasalization as an indication that an upcoming segment would be nasal. An event-related potential (ERP) study with adults by Archibald and Joanisse (2011) showed that both phonemic and non-contrastive mismatches created the same neurological pattern characteristic of unexpected phonemic information (see also Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005). These studies provide support for the view that coarticulation cues are parts of the words' phonological representations and that these cues are used during on-line spoken word recognition.