

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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



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

ARTICLE INFO

Article history:

Received 24 March 2016

Revised 24 July 2016

Keywords:

Coarticulation

Spoken word recognition

Language processing

Eye-tracking

Lexical competition

Vowel nasalization

ABSTRACT

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_it*, *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.

Anticipatory coarticulation appears to be an important feature that adults use to make lexical decisions on the identity of words.

The picture is murkier when it comes to the developing speech recognition system. Researchers have shown that infants are able to perceive and discriminate subphonemic or subcategorical alternations. McMurray and Aslin (2005) demonstrated that 8-month-old infants could differentiate among prototypical tokens of /b/, prototypical tokens of /p/, and within-category variations closer to the category boundary. There is also evidence that infants make use of coarticulatory information from surrounding syllables when perceiving an ambiguous token. Fowler, Best, and McRoberts (1990) showed that, similar to adults, 4-month-old infants discriminated an ambiguous token midway between /da/ and /ga/ differently based on whether that token was preceded by /ar/, which causes a more back tongue position, or /al/, which results in a more fronted position. Research has also found that 7- and 8-month-old infants are able to use coarticulatory cues in word segmentation (Curtin, Mintz, & Byrd, 2001; Johnson & Jusczyk, 2001) and that at 12 months infants are able to use subphonemic acoustic cues associated with syllable stress in word segmentation (Johnson, 2008). This literature supports the claim that both prelinguistic infants and adults are able to detect subphonemic variation.

Although infants are able to perceive non-contrastive differences (McMurray & Aslin, 2005), there has been little research on whether young children can make use of coarticulation cues during spoken word recognition. While it is important to know how and when coarticulation cues are perceived during development, it is also important to know how and when these cues activate targets and inhibit competitors during lexical access, as in spreading activation type models of spoken word recognition such as TRACE (Mayor & Plunkett, 2014; McClelland & Elman, 1986). When adults are presented with misleading coarticulation cues, they are able to resolve the ambiguity based on the identity of the final consonant because the activation of the target is greater than that of the competitor (e.g., Dahan et al., 2001). Previous work has indicated that 5-year-old children have difficulty in resolving competitor activation (Huang & Snedeker, 2011). Swingley, Pinto, and Fernald (1999) found that 2-year-old children showed slower response latencies in fixations to a target image (e.g., *dog*) when it was paired with a phonologically similar competitor (*dog-doll*) than when the same target was paired with a distracter that did not phonologically overlap (*dog-tree*). No studies, however, have examined the time course of target and competitor activation and inhibition in toddlers. Young children have not fully developed their phonological and lexical systems, nor have they had the extensive experience with their native languages as adult speakers (Curtin, Byers-Heinlein, & Werker, 2011; Curtin & Zamuner, 2014). By studying how children differ from adults, researchers can examine the mechanisms of speech recognition as children develop more adult-like systems. The questions of interest with children are as follows: Do children perceive the subcategorical differences in the articulation of sounds within word boundaries? Do children make use of these cues for word recognition in a way similar to adults? To what extent are children able to resolve lexical competition?

A handful of studies have looked at how young children use coarticulatory information as they process speech across word boundaries, with conflicting results. Minaudo and Johnson (2013) examined whether 2-year-olds use coarticulatory information during the processing of sentences where a medial word either matched the coarticulation cues with a following word (“Look at the pretty_s strawberry”) or were mismatched (“Look at the pretty_s ladder”). Looking times did not differ between the matching and mismatching conditions, suggesting that children do not yet possess the ability to use or integrate subphonemic cues across word boundaries during on-line processing at 24 months of age. However, Minaudo and Johnson argued that children are likely sensitive to coarticulatory cues, which could be studied in the future by choosing stimuli with phonemes that have “low resistance” to coarticulation, that is, that are more likely to be realized with coarticulation. In contrast, Mahr, McMillan, Saffran, Weismer, and Edwards (2015) found that 18- to 24-month-olds could use coarticulation on the preceding article *the* to facilitate looking to the correct target. In their study, toddlers were presented with trials asking them to, for example, “Find the duck,” where *the* varied between conditions. In neutral trials, *the* contained no coarticulation with the first sound in the target word and was taken from the phrase “the hut,” where the oral cavity remains in the same configuration throughout [ə] and [h]. In facilitating trials, *the* was taken from a context that provided the same cues as would be present in unspliced speech, meaning that if the trial were “find the ball,” coarticulation on *the* would cue an upcoming bilabial. Toddlers looked to the correct target approximately 100 ms

faster when facilitating coarticulation cues were present on the article compared with trials where no coarticulation cues were available before the word onset. This indicates that toddlers do in fact make use of coarticulation cues on a preceding definite article in a facilitative way.

Thus, gaps remain in our understanding of how children perceive and make use of coarticulation cues. Although it is becoming clear that infants and young children can, at least task dependently, perceive coarticulation cues, the current study examined whether children are able to perceive and use subcategorical cues within words during spoken word recognition. Determining how within-word coarticulation is processed will enable researchers to better understand how both the phonological and lexical systems develop to an end state of adult-like recognition. To explore the effects of within-word coarticulation, we focused on the coarticulation of nasality on a preceding vowel. Nasality is a strongly coarticulated cue in English, making it a reliable anticipatory cue of upcoming nasal consonants (Beddor, 2009; Cohn, 1993). Vowel nasalization in English can be extensive throughout the whole vowel duration, and the amount varies depending on syllable structure and context, as well as between individual speakers and dialects. Adults have been shown to use coarticulatory vowel nasalization to anticipate upcoming nasal consonants (Beddor et al., 2013). Vowel nasalization is salient in English as compared with other shorter subphonemic cues such as voice onset time and formant transitions. Consequently, it constitutes a good phonetic phenomenon to examine whether children use coarticulation at all during spoken word recognition; null results would suggest that young children do not yet use coarticulation in a “meaningful” manner, whereas conclusive results would show that children possess the abilities to use coarticulation for word recognition (and potentially word learning), at least for salient cues. To date, there has been little developmental work on learners’ perception of nasal consonants and nasalized vowels, although the work that has been done found early sensitivity to these segments in studies of speech perception with infants (Hillenbrand, 1984; Narayan, Werker, & Beddor, 2010; Seidl, Cristià, Bernard, & Onishi, 2009). Thus, 2- and 3-year-old English-learning children are expected to be able to perceive nasalization cues; however, we further examined whether these cues activate word candidates. Testing adults and children with the same stimuli would allow us to examine the development of the spoken word recognition system.

To look at the correct image when hearing a phonetic form, a listener must activate the concept that matches the auditory label. In this research, CVC and CVN (consonant–vowel–nasal) minimal pairs (e.g., *boat* [bɔʊt]–*bone* [bɔʊn]) were spliced to create new tokens where vowel coarticulation cued either an oral consonant (same-splice; [bɔʊt]) or a nasal consonant (cross-splice; [bɔʊ̃t]) in order to examine the time course of word recognition rather than pure auditory processing. Experiment 1 first tested adult participants. We predicted that in same-spliced trials, adults would look faster to the target image, as found in Beddor et al. (2013) and Dahan et al. (2001). We also predicted that in the cross-spliced trials, adult listeners would be cued by the misleading vowel nasality. Thus, it was predicted that adults would first fixate on the competitor image due to the lexical activation created by coarticulation (mis)cues before moving their gaze to the target image.

Experiment 1

Method

Participants

Participants were 26 university students (14 men and 12 women, $M_{\text{age}} = 22$ years) who received partial course credit for participating. All adult participants were native speakers of English who completed a language background questionnaire. Of the 26 participants, 12 self-reported some basic knowledge of French but reported using it 10% or less of the time in their daily lives. All adult participants reported having normal or corrected-to-normal vision, normal hearing, and no history of language impairment. An additional 2 participants were tested but excluded due to a ruptured ear drum ($n = 1$) or for being a balanced French–English bilingual ($n = 1$). This posed a potential problem because French has phonemic nasal vowels, and research has shown that there are differences in how vowel nasalization is realized in English and French (Cohn, 1993; Desmeules-Trudel, 2015; Versteegh, Seidl, & Cristià, 2014).

Stimuli

Stimuli were composed of 6 pairs (*boat–bone*, *castle–candy*, *cloud–clown*, *kick–king*, *hat–hand*, and *bug–bunkbed*) and 9 filler pairs (*boots–carrot*, *star–keys*, *monkey–camel*, *frog–fish*, *dog–elephant*, *turtle–sandwich*, *chicken–kangaroo*, *doll–clock*, and *flower–sun*). An additional 3 experimental pairs were originally included (*duck–dumptruck*, *leg–lemon*, and *egg–M*); however, it was later noticed that these pairs have multiple coarticulation cues of nasalization and place of articulation. The purpose of the study was not to compare the number of coarticulation cues, and there was not enough data to analyze the subset of nasal and place cued pairs in a meaningful way. Therefore, the pairs with two sources of coarticulation cues were removed from the analyses. The results across the experiments were the same with or without these items included in the analyses (most relevant was the significant effect of splicing on the quadratic term when all items are included; *estimate* = -5.46 , *SE* = 0.56 , $p < .001$). Each word was depicted with a color image of similar size and visual interest. Animacy was held constant as much as possible across pairs. For example, an image of a *cloud* with a face was selected to be closer to its minimal pair *clown*.

Auditory filler and experimental stimuli were recorded on a SHURE Microflex MX392/0 omnidirectional microphone and a Blue Microphone Icicle XLR-to-USB adapter by a female native English speaker trained in phonetics. Stimuli were spliced to create the two experimental conditions. In the same-splice condition (W1W1, W = word), a token of the non-nasal carrier word was spliced with another token of the same word. The result was a new token with the initial consonant and syllable coda of the first token (C_C), with the vowel of the second token (_V_). As an example, for the *boat–bone* pair, one token of the word *boat* that we can refer to as *boat*₁, had its vowel removed and replaced with the vowel from a different token of *boat*, *boat*₂, to make the experimental stimulus for the same-splice condition (W1W1). The final product had the form $b_1o_a2t_1$, with each subscript number labeling the token from which the audio originated. For the W2W1 condition, to create a coarticulation mismatch, the same original token of *boat* was used, but the vowel from a token of *bone* replaced the original vowel. The final product here was $b_1o_Nt_1$, where the vowel followed by a subscript “N” represents a nasalized vowel. Crucially, the target for both conditions was always *boat*, but the difference was whether the vowel was oral (same-splice, W1W1) or nasalized (cross-splice, W2W1). Fillers were not cross-spliced.

A trained phonetician did the splicing with Praat software (Boersma & Weenink, 2010). All segments were deleted or inserted at zero crossings to avoid unnatural clicks or pops in the final product. The resulting words in both conditions sounded like natural speech with no obvious splice points. In the cross-splice (W2W1) condition for the pair *kick–king*, the initial consonant and vowel from *king* was used rather than just the nasalized vowel. All auditory materials were set to 70 dB amplitude. Acoustic measurements for the stimuli are provided in Table 1. Vowel nasalization was identified

Table 1

Acoustic measurements of timing information (in ms) for same-splice (W1W1) and cross-spliced (W2W1) stimuli.

Condition	Word	IPA	V onset	Nasalization onset	% V nasalized	Final C onset	Total word duration
W1W1	boat	[b ^o u ^t]	23			348	670
	bug	[bʌg]	18			323	597
	castle	[kæ ^s l]	198			401	937
	cloud	[kla ^u d]	238			649	857
	hat	[hæt]	175			462	885
	kick	[kɪk]	185			350	779
	Average (W1W1)			140			422
W2W1	bo _a n _t	[b ^o u ^t]	23	89	88	555	848
	bu _n g	[bʌg]	15	58	76	192	452
	ca _n stle	[kæ ^s l]	148	189	81	370	906
	clou _n d	[kla ^u d]	250	373	73	704	914
	ha _n t	[hæt]	275	275	100	666	1032
	ki _n ck	[kɪk]	206	240	87	463	873
	Average (W2W1)		153	204	84	492	838
Total averages			146			457	813

by examining the general loss in spectral energy (formant flattening), loss of spectral energy in the high frequencies, changes in the waveform during the vowel (reduced amplitude), and auditory confirmation (see intervals between vertical lines in Fig. 1). Vowels in the cross-splice condition were nasalized an average of 84% of the vowel duration. The spectrogram for the target *cloud* from the W2W1 stimulus can be seen in Fig. 1, which indicates the timing of the beginning of the vowel, vowel nasalization, and the beginning of the final consonant. In some pairs, as a result of preserving natural sounding speech, vowel onsets within a pair differed by up to 100 ms. We return to the timing of nasalization and final consonant in the discussion of looking to the target relative to the acoustic signal. The splicing methodology of using just the vowel isolates the coarticulation cues to just vowel nasalization. This differs from Dahan et al. (2001), where splicing was done by taking both the onset and nucleus of the vowel pair.

Design

The experiment was composed of a familiarization phase and a test phase. In familiarization, all of the images were shown one at a time and received a prerecorded auditory label. This was to ensure that participants knew the label for each image. At test, using a looking-while-listening paradigm (Fernald, Perfors, & Marchman, 2006; Johnson & Zamuner, 2010), participants saw two images on the screen side by side in silence for 1500 ms, and the position of the images on the screen (left or right) was counterbalanced. Each image corresponded to one of the words in the pairs (e.g., *boat* and *bone*). Then, the phrase “Look at the _____” played, followed by the target word. The images remained on the screen for 4 s. Participants were tested on 6 test trials (plus the additional 3 items that were removed from the analysis) and heard a mix of same-splice W1W1 and cross-splice W2W1 trials. Items were presented only one time and not repeated. For any one pair, participants heard either the W1W1 condition or the W2W1 condition. In other words, participants did not hear the same target word in both the W1W1 and W2W1 conditions. Trials were presented in a semi-random order, with approximately one filler trial between each test trial. There were four different lists for counterbalancing. Each trial lasted approximately 6 s.

Procedure

Participants were tested in a sound-attenuated booth and were seated 500 to 600 mm from a monitor. Eye movements to the target and the distracter were recorded using an EyeLink 1000 Remote eye-tracker (SR Research, Ottawa, Ontario, Canada), tracking participants’ dominant eye, with a sampling

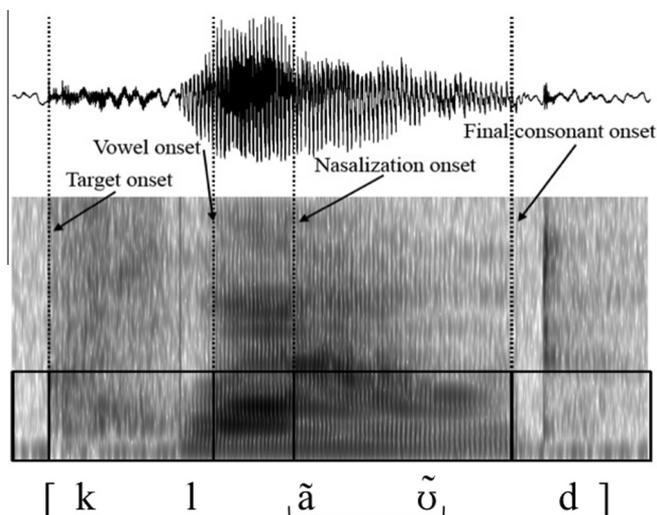


Fig. 1. Waveform and spectrogram for target word “cloud” from cross-splice (W2W1) condition.

rate of 500 Hz. Calibration was based on a three-point calibration grid. The experiment took approximately 5 minutes to complete.

Results

Data were extracted using DataViewer software in 100-ms time bins (SR Research). Analyses began 200 ms after the onset of the target because this is the time it typically takes for adults to program an eye movement (Matin, Shao, & Boff, 1993; Salverda et al., 2014). An inspection of the fixation data indicated that looking to the target in the mismatch condition peaked at 1400 ms after the target word onset, indicating that reference had been resolved. Although the images remained on the screen for an additional 600 ms, looking to the target in both conditions did not increase. This resulted in a window of analysis spanning from 200 to 1400 ms from the target word onset, with 12 time bins in total. Proportion of fixation to the target was calculated by comparing the time spent fixating on the target image (*boat*) compared with the time spent looking at both the target (*boat*) and the competitor image (*bone*).

The proportional data in the 200- to 1400-ms window were examined with a growth curve analysis (GCA; Mirman, Dixon, & Magnuson, 2008). GCA is a multilevel regression method allowing for the assessment of both the differences in time spent looking and in the steepness of a looking curve (i.e., in the rapidity of access to a target word over time). Adults' proportional looking to the target using GCA was done in R (R Core Team., 2013) using the lmer() function from the lme4 package (Version 1.1-7; Bates, Maechler, Bolker, & Walker, 2014). The empirical logit was calculated for each time bin as an approximation to log odds (Barr, 2008). The model contained splicing (same-splice W1W1, cross-splice W2W1) as the main predictor (within-participant variable), which was deviation coded. The model also included a main effect of time (captured by orthogonal polynomials) and a Splicing * Time interaction. Random by-participant effects were included in the model to account for by-participant variation in their slopes and intercepts. Item random effects were not included due to the small number of stimulus items.

Looking to the target over the time course was modeled with third-order (cubic) orthogonal polynomials. Results of the statistical analysis are presented in Table 2. Fixed effects of splicing and time were included as the main predictors of interest. There was a statistically significant effect of splicing ($estimate = 4.04$, $SE = 0.47$, $p < .001$) and also a significant effect of splicing on the linear term ($estimate = 2.86$, $SE = 0.73$, $p < .001$) and quadratic term ($estimate = -6.49$, $SE = 0.70$, $p < .001$). This indicates a difference in the shape of the looking curve between the conditions, as seen in Fig. 2; there was a rise in fixations in the W1W1 condition over time, and by contrast there was a fall and rise in the W2W1 condition over time.

In Experiment 1, on W1W1 trials, adults' fixations to the target image increased over time, starting 400 ms after word onset (Fig. 2). This difference was significant compared with chance starting at 500 ms after word onset (50% fixation to the target image) and extended throughout the rest of the

Table 2

Experiment 1 adults' GCA results: Empirical logit GCA for the effect of splicing on looking data for a window of analysis 200 to 1400 ms after the target word onset.

Predictor	Estimate	SE	t	p
Intercept	1.40	0.22	6.30	<.001
Splicing (W1W1 vs. W2W1)	4.04	0.47	8.66	<.001
Linear term	4.46	0.81	5.50	<.001
Quadratic term	0.86	0.47	1.81	.07
Cubic term	-0.07	0.39	-0.17	.86
Splicing * Linear Term	2.86	0.73	3.90	<.001
Splicing * Quadratic Term	-6.49	0.70	-9.31	<.001
Splicing * Cubic Term	0.59	0.68	0.86	.39

Note: Values represent parameter estimates. GCA structure in R: lmer(eiog ~ Splicing * (Linear + Cubic) + (1 + Splicing + Linear + Cubic|Participants). p-values were calculated in R using a normal approximation based on the assumption that the t-distribution converges to the z-distribution.

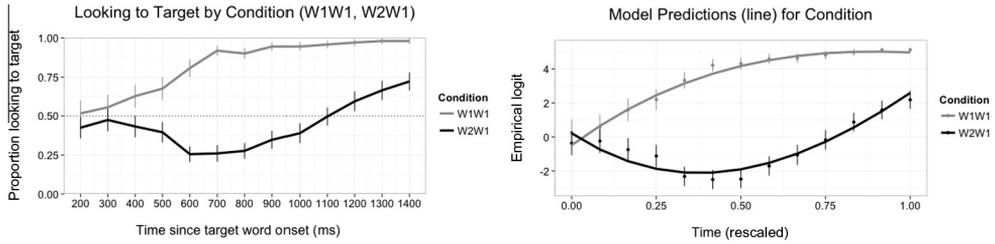


Fig. 2. Experiment 1: Adults' proportions of looking graph and model predictions. Left: Time course of proportion of looking to the target 200 ms after word onset for targets from W1W1 and W2W1 splicing conditions. The horizontal dotted line corresponds to chance (.50), and the vertical line throughout each point represents the standard error of the mean. Right: Model predictions for splicing.

trial (based on sliding two-tailed *t*-tests). In contrast, looking to the target on the W2W1 trials dropped at 400 ms and was significantly lower than chance from 600 to 900 ms. This is in line with the acoustic measurements of the stimuli, where we would expect to see evidence of listeners' sensitivity to nasalization in the W2W1 condition to begin 400 ms after word onset (nasalization occurred on average 204 ms after word onset, with an additional 200 ms for planning of eye movements). At 800 ms after word onset, participants shifted their gaze from the competitor image to the target image. This shift corresponds to the timing of the final consonant of the word, which began on average 700 ms after word onset (the average final consonant began 492 ms after the word plus 200 ms for eye movements). Starting at 1300 ms, adults looked significantly more to the target image until the end of the time period.

The results from Experiment 1 show that adults are sensitive to within-word coarticulatory nasalization during on-line spoken word recognition. Adult looking behavior in this study resembles that in previous studies with coarticulatory miscues (Beddor et al., 2013; Dahan et al., 2001). This supports a theory of phonological storage that preserves coarticulation cues. To determine whether this phonetic awareness is present earlier in development, in Experiment 2 we conducted the same study with 2- and 3-year-old children to determine whether they are able to perceive and make use of coarticulation cues in a way similar to adults. If children's spoken word recognition system uses the same mechanisms as adults, where coarticulation cues are used during on-line processing, children's looking behavior should appear similar to the adult results. If 2- and 3-year-olds are not yet making use of coarticulation cues, both trial types should show the same pattern.

Experiment 2

Method

Participants

Participants were 19 children (12 boys and 7 girls). The age range was 2;1 (years;months) to 3;10 ($M_{\text{age}} = 3;0$). All participants were native speakers of English with no more than 25% exposure to other languages. An additional 3 children were tested but excluded because of difficulty in tracking their eyes ($n = 1$) or not completing the experiment ($n = 2$). All children had normal hearing, normal vision, and no history of language impairment, as determined by parental questionnaire.

Stimuli

Stimuli were identical to those used in Experiment 1, with just the pairs with one source of coarticulation. As with Experiment 1, the results were the same when the pairs with two sources of coarticulation were included in the analyses (a significant effect of splicing on the quadratic term; $estimate = 2.07$, $SE = 0.73$, $p < .001$).

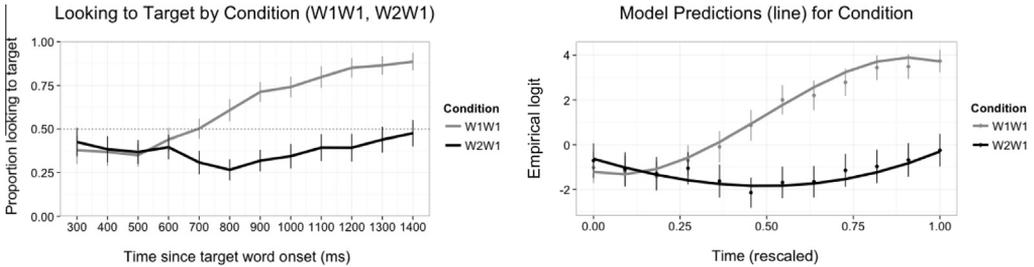


Fig. 3. Experiment 2: Children's proportions of looking graph and model predictions. Left: Time course of proportion of looking to the target 300 ms after word onset for targets from W1W1 and W2W1 splicing conditions. The horizontal dotted line corresponds to chance (.50), and the vertical line throughout each point represents the standard error of the mean. Right: Model predictions for splicing.

Procedure

The procedure was the same as in Experiment 1; however, children sat by themselves or on a parent's lap. Two thirds of the children were tested in a sound-attenuated booth at the Centre for Child Language Research (CCLR) at the University of Ottawa, and one third of the children were tested on the same eye-tracker that was transported to use in a sound-attenuated room at a local museum. There was no difference in the results for the children tested on campus versus those tested at the museum.

Results

Statistical analyses were the same as in Experiment 1. For children, the amount of time required to execute an eye movement in response to a stimulus is longer than the 200 ms used with adults, although there are varying standards within the field. We follow [Buckler and Fikkert \(2016\)](#), who used 300 ms as the beginning of the analysis window in their study with 38-month-olds. Children's fixations to the target in the mismatch cross-splice (W2W1) condition peaked at 1400 ms after target word onset and held constant throughout the remainder of the trial. This resulted in a window of analysis spanning from 300 to 1400 ms from the target word onset, with 11 time bins in total. [Fig. 3](#) shows the proportion of looks to the target for each of the splicing conditions as a function of time.

As with adults, children's looking to the target over the time course was modeled with third-order (cubic) orthogonal polynomials (see [Table 3](#)). There was a statistically significant effect of splicing overall ($estimate = 2.62, SE = 0.78, p < .001$) and a significant effect of splicing on the linear term ($estimate = 6.51, SE = 0.88, p < .001$) and quadratic term ($estimate = -1.79, SE = 0.86, p < .01$). As with the adults, there was a difference in the shape of children's looking curve between the conditions, as seen in [Fig. 3](#). In the W1W1 condition there is a rise in fixations to the target over time, and in the W2W1 condition there is a fall and shallow rise over time.

Children showed a strikingly similar pattern to adults on the same-splice W1W1 trials (cf. [Figs. 2 and 3](#)). In the W1W1 condition, children looked significantly more to the target compared with chance starting 900 ms after word onset, which remained above chance until the end of the trial (based on sliding two-tailed t -tests). On W2W1 trials between 700 and 1000 ms after word onset, children showed less than chance looking to the target, indicating that they initially looked to the competitor. If children had perceived the competitor as the target, we would expect that looking to the competitor would increase and continue throughout the trial. Instead, children then shifted their gaze from the competitor; however, the coarticulation miscues about the identity of the target were not fully overridden by subsequent information. Even though the words ended in final non-nasal consonants (e.g., the final /t/ in the cross-splice target /bo_nt/), children did not look more to the target toward the end of the trial. It is important to keep in mind that filler trials were not analyzed in the current experiment. Consequently, the claims made here concern the timing difference between matching and mismatching fine-grained coarticulatory cues of two competing lexical items. The interested reader should refer to [Swingley et al. \(1999\)](#) for an experiment on lexical competition using competitor and non-competitor items.

Table 3

Experiment 2 children's GCA results: Empirical logit GCA for the effect of splicing on looking data for a window of analysis 300 to 1400 ms after the target word onset.

Predictor	Estimate	SE	t	p
Intercept	0.0002	0.29	0.007	.99
Splicing (W1W1 vs. W2W1)	2.62	0.78	3.36	<.001
Linear term	3.52	0.93	3.80	<.001
Quadratic term	0.77	0.50	1.55	.12
Cubic term	−0.66	0.46	−1.44	.15
Splicing * Linear Term	6.51	0.88	7.41	<.001
Splicing * Quadratic Term	−1.79	0.86	−2.09	<.05
Splicing * Cubic Term	−1.50	0.86	−1.75	.08

Note: Values represent parameter estimates. GCA structure in R: $\text{Imer}(\text{elog} \sim \text{Splicing} * (\text{Linear} + \text{Cubic}) + (1 + \text{Splicing} + \text{Linear} + \text{Cubic}) | \text{Participants})$. *p*-values were calculated in R using a normal approximation based on the assumption that the *t*-distribution converges to the *z*-distribution.

General discussion

In this research, participants were tested on how they perceived coarticulation cues of vowel nasalization during spoken word recognition. In Experiments 1 and 2, adults and children looked faster to targets that had matching coarticulation cues (W1W1). In the mismatch cross-splice condition (W2W1), both adults and children initially fixated more on the competitor image. Thus, the results demonstrate that, like adults, 2- and 3-year-old children are able to perceive and make use of within-word nasal coarticulation cues during on-line spoken word recognition. Despite the similarities across the age span, differences between adults and children were seen in the timing of the effects as well as in the ability of participants to recover after miscues.

In the same-splice (W1W1) condition with adults, there was a significant increase in looking to the target starting 500 ms after word onset. This is similar to Dahan et al. (2001), where a difference between conditions emerged at 600 ms after the target word onset. Both studies have an important similarity in that participants needed to differentiate between cohort competitors; however, in the current study participants were presented with a two-image display, whereas Dahan and colleagues' study used four images. For the children, increased looking to the target was not observed until 900 ms (400 ms later than with adults). This latency is not surprising because it takes children more time than adults to initiate an eye movement in response to an auditory stimulus. Fernald et al. (2006) found that 25-month-olds had a mean reaction time of 769 ms to look to a target after an auditory prompt. Because our task includes two cohort competitors, a slightly longer latency can be expected as children interpret the phonetic coarticulation cues that distinguish the two images.

Our results support spreading activation type models of spoken word recognition such as the TRACE model (Mayor & Plunkett, 2014; McClelland & Elman, 1986), where coarticulation cues can be used to further activate targets and inhibit competitors. The TRACE model explains how adults in Experiment 1 were able to recover from the misleading coarticulation nasal cues to eventually arrive at the target in the cross-splice condition; once the final consonant was presented, target activation was greater than that of the competitor. With both the target and competitor receiving activation and inhibition due to mismatching cues, it is not surprising that whereas adults resolved the ambiguity in the cross-splice condition, adults' looking to the target did not reach the same levels as in the same-splice condition. In contrast, although the results from Experiment 2 are still in line with the TRACE model because children were sensitive to coarticulation cues during on-line processing, children did not fully resolve the mismatching nasal vowel cues. Children's gaze began to shift away from the competitor on the W2W1 trials once the final consonant was presented, but looking to the target did not differ from chance. This could be due to fewer exemplars in their memories and/or less robust representations for the target words, which may result in less activation for the target even after the final consonant was realized. This may be combined with a less mature processing system, as argued in other research showing that 5-year-old children show continued interference from competitors (Huang & Snedeker, 2011), which suggests that children's processing system does

not inhibit competing representations in the same way as adults. (See also Mayor & Plunkett, 2014; Rigler et al., 2015, and Swingley & Aslin, 2007, for further discussion about lexical competition and the development of children's processing system.)

The current findings add support to the notion that young children show a maturing but incomplete processing system. This is reflected in their sensitivity to coarticulation and, at the same time, their inability to fully resolve the ambiguity introduced by subphonemic miscues. We predict that if one reduces the amount of nasalization coarticulation cued, 2- and 3-year-olds would be able to resolve the competition (on average 84% of the vowels duration was nasalized; see Table 1). Another possibility is that the apparent inability of children to fully resolve the ambiguity could be the result of children giving subphonemic cues more weight than adult listeners. A child with a newly developing phonological system may "over"-represent details (Houston & Jusczyk, 2000; Singh, Morgan, & White, 2004), and the level of detail in exemplars is pared down as the child's speech recognition system matures. (We thank a reviewer for pointing out this interpretation.) Future research could examine representational detail by looking at whether children's attention to subphonemic detail depends on a word's frequency and the word's status as a minimal pair. One would predict earlier sensitivity to vowel nasalization with higher frequency words and with words that do and do not have minimal pairs because the distinction between oral and nasalized vowels is expected to be more enhanced with words with minimal pairs (Goldrick, Vaughn, & Murphy, 2013).

This study has presented new data with regard to the way in which listeners treat nasal coarticulation in English across development. However, many questions remain unanswered such as how these cues are perceived across languages. If we limit our research to English, we may mis-estimate children's skills because coarticulation cues differ across languages and vary in their predictability and saliency. Vowel nasality in English is a consequence of the presence of a nasal consonant, and it is motivated by articulation. In contrast, French uses vowel nasality to make phonemic and lexical distinctions between nasal and oral vowels. Thus, although English has nasalized vowels created by an upcoming nasal consonant, in French the difference between a nasal vowel and an oral vowel is phonemic; that is, they define a lexical distinction as in *paon* /pɑ̃/ 'peacock' and *pas* /pa/ 'step'. A good way in which to have a broader idea of how children process coarticulation cues would be to test English–French bilinguals. One could investigate whether English–French bilingual children perceive and process nasalization miscues differently from monolinguals given the phonetic and phonemic status across the two languages. For example, English–French bilingual children may use nasalization cues even when processing English words because the learner would be exposed to both phonetic and phonemic vowel nasality. Studies with bilinguals provide an opportunity to see how two differing levels of phonetic/phonemic awareness interact in the same mind (Byers-Heinlein & Fennell, 2014).

The results demonstrate continuity across development, indicating that learners incorporate early sensitivity to subphonemic cues in speech perception (e.g., Curtin et al., 2001; Fowler et al., 1990; Johnson, 2008; Johnson & Jusczyk, 2001; McMurray & Aslin, 2005) into a spoken word recognition system that uses this fine-grained detail. Children perceive coarticulation cues in a way that resembles adult patterns. Both adults and children in the W2W1 trials looked to the competitor, which matched the coarticulation nasal cues. This parallel points to the idea that coarticulation cues are a fundamental component of the word recognition mechanism and that the ability to make use of these cues arises in tandem with the rest of the system (Mahr et al., 2015). (See also Paquette-Smith, Fecher, & Johnson, *in press*, who found that toddlers are more sensitive to phonemic mismatches compared with subphonemic mismatches.) Thus, young learners are sensitive to subphonemic cues created by anticipatory coarticulation in a way that resembles adult behavior, as seen both in previous research looking across word boundaries (Mahr et al., 2015) and in the current study, within words. However, at the same time, children did not show the same resolution as adults of the mismatching competitor inhibition. This indicates that children's processing systems are still developing.

Acknowledgments

This research was supported by a Social Sciences and Humanities Research Council (SSHRC) – Canada research grant awarded to T.S.Z., a SSHRC M.A. grant awarded to C.M., and a Fonds québécois de recherche–Société et culture (FQRSC) and SSHRC doctoral fellowship awarded to F.D.-T. We thank

Sarah Colby for her contribution, Yaroslav Konar for contributing to the statistical analyses, and anonymous reviewers for providing helpful comments.

References

- Archibald, L., & Joannis, M. (2011). Electrophysiological responses to coarticulatory and word level miscues. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 1275–1291.
- Barr, D. J. (2008). Analyzing “visual world” eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*, 457–474.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*. R package (Version 1.1-7). <http://CRAN.R-project.org/package=lme4>.
- Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, *85*, 785–821.
- Beddor, P. S., McGowan, K., Boland, J., Coetzee, A., & Brasher, A. (2013). The timecourse of perception of coarticulation. *Journal of the Acoustical Society of America*, *133*, 2350–2366.
- Boersma, P., & Weenink, D. (2010). *(P)raat: Doing phonetics by computer* (Version 5.3.79) [computer program]. <http://www.praat.org/>.
- Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, *6*, 201–251.
- Buckler, H., & Fikkert, P. (2016). Dutch and German 3-year-olds’ representations of voicing alternations. *Language and Speech*, *59*, 236–265.
- Byers-Heinlein, K., & Fennell, C. T. (2014). Perceptual narrowing in the context of increased variation: Insights from bilingual infants. *Developmental Psychobiology*, *56*, 274–291.
- Cohn, A. C. (1993). Nasalisation in English: Phonology or phonetics. *Phonology*, *10*, 43–81.
- Curtin, S., Byers-Heinlein, K., & Werker, J. F. (2011). Bilingual beginnings as a lens for theory development: PRIMIR in focus. *Journal of Phonetics*, *39*, 492–504.
- Curtin, S., Mintz, T. H., & Byrd, D. (2001). Coarticulatory cues enhance infants’ recognition of syllable sequences in speech. In A. H. J. Do, L. Dominguez, & A. Johansen (Eds.), *Proceedings of the 25th annual Boston University Conference on Language Development* (vol. 1, pp. 190–201). Somerville, MA: Cascadia Press.
- Curtin, S., & Zamuner, T. S. (2014). Understanding the developing sound system: Interactions between sounds and words. *WIREs Cognitive Science*, *5*, 589–602.
- Dahan, D., Magnuson, J., Tanenhaus, M., & Hogan, E. (2001). Subcategorical mismatches and the timecourse of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, *16*, 507–534.
- Desmeules-Trudel, F. (2015). Propriétés aérodynamiques et comparaison des voyelles nasales et potentiellement nasalisées en français québécois. In S. Vinerte (Ed.), *Proceedings of the 2015 annual conference of the Canadian Linguistic Association*. Ottawa, Ontario: Canadian Linguistic Association.
- Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*, 98–116.
- Flege, J. (1988). Anticipatory and carry-over nasal coarticulation in the speech of children and adults. *Journal of Speech Language and Hearing Research*, *31*, 525–536.
- Fowler, C. A., Best, C. T., & McRoberts, G. W. (1990). Young infants’ perception of liquid coarticulatory influences on following stop consonants. *Perception & Psychophysics*, *48*, 559–570.
- Goldrick, M., Vaughn, C., & Murphy, A. (2013). The effects of lexical neighbors on stop consonant articulation. *Journal of the Acoustical Society of America*, *134*, EL172–EL177.
- Gow, D. W., & McMurray, B. (2007). Word recognition and phonology: The case of English coronal place assimilation. In J. Cole & J. I. Hualde (Eds.), *Laboratory phonology 9* (pp. 173–2000). Berlin: Mouton de Gruyter.
- Hillenbrand, J. (1984). Speech perception by infants: Categorization based on nasal consonant place of articulation. *Journal of the Acoustical Society of America*, *75*, 1613–1622.
- Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1570–1582.
- Huang, Y. T., & Snedeker, J. (2011). Cascading activation across levels of representation in children’s lexical processing. *Journal of Child Language*, *38*, 644–661.
- Johnson, E. K. (2008). Infants use prosodically conditioned acoustic–phonetic cues to extract words from speech. *Journal of the Acoustical Society of America*, *123*, EL144–EL148.
- Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, *44*, 548–567.
- Johnson, E. K., & Zamuner, T. S. (2010). Using infant and toddler testing methods in language acquisition research. In E. Blom & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 73–93). Amsterdam: John Benjamins.
- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). Dordrecht, Netherlands: Kluwer Academic.
- Mahr, R., McMillan, B. T. M., Saffran, J. R., Weismer, S. E., & Edwards, J. (2015). Anticipatory coarticulation facilitates word recognition in toddlers. *Cognition*, *142*, 345–350.
- Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, *53*, 372–380.
- Mayor, J., & Plunkett, K. (2014). Infant word recognition: Insights from TRACE simulations. *Journal of Memory and Language*, *71*, 89–123.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.
- McMurray, B., & Aslin, R. N. (2005). Infants are sensitive to within-category variation in speech perception. *Cognition*, *95*, B15–B26.
- McQueen, J. M., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1363–1389.

- Miller, J. L., & Dexter, E. R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 369–378.
- Minaudo, C., & Johnson, E. K. (2013). Are two-year-olds sensitive to anticipatory coarticulation? *Journal of the Acoustical Society of America*, 133, 3335.
- Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, 59, 475–494.
- Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science*, 13, 407–420.
- Paquette-Smith, M., Fecher, N., & Johnson, E. K. (in press). Two-year-olds' sensitivity to subphonemic mismatch during online spoken word recognition. *Attention, Perception, & Psychophysics*.
- R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org/>.
- Rigler, H., Farris-Trimble, A., Greiner, L., Walker, J., Tomblin, J. B., & McMurray, B. (2015). The slow developmental timecourse of real-time spoken word recognition. *Developmental Psychology*, 51, 1690–1703.
- Salverda, A., Kleinschmidt, D., & Tanenhaus, M. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, 71, 145–163.
- Seidl, A., Cristià, A., Bernard, A., & Onishi, K. H. (2009). Allophonic and phonemic contrasts in infants' learning of sound patterns. *Language Learning and Development*, 5, 191–202.
- Singh, L., Morgan, J. L., & White, K. S. (2004). Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory and Language*, 51, 173–189.
- Swingle, D., & Aslin, R. N. (2007). Lexical competition in young children's word learning. *Cognitive Psychology*, 54, 99–132.
- Swingle, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, 71, 73–108.
- Van Berkum, J., Brown, C., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 443–467.
- Versteegh, M., Seidl, A., & Cristià, A. (2014). Acoustic correlates of phonological status. In H. Li, H. M. Meng, B. Ma, E. Chng, & L. Xie (Eds.), *Proceedings of Interspeech 2014* (pp. 91–95). Singapore: International Speech Communication Association.
- Whalen, D. H. (1991). Subcategorical phonetic mismatches and lexical access. *Perception & Psychophysics*, 50, 351–360.