Are two-year-olds sensitive to anticipatory coarticulation?

Caterina Minaudo, and Elizabeth K. Johnson

ARTICLES YOU MAY BE INTERESTED IN

Proceedings of Meetings on Acoustics

ICA 2013 Montreal

Montreal, Canada

2 - 7 June 2013

Speech Communication

Session 2aSC: Linking Perception and Production (Poster Session)

2aSC14.   Are two-year-olds sensitive to anticipatory coarticulation?

Caterina Minaudo* and Elizabeth K. Johnson

 *Corresponding author's address: Psychology, University of Toronto, 3359 Mississauga Road N., Mississauga, L5L 1C6, ON, Canada, c.minaudo@mail.utoronto.ca

  Eyetracking studies have shown that adults are highly sensitive to subphonemic detail in speech (e.g. Shatzman & McQueen, 2006). In some circumstances, adults use subphonemic coarticulatory information to anticipate which word will occur next in the speech stream (McDonough et al., 2009). In the current study, we ask whether two-year-old children use anticipatory coarticulation in a similar manner. Sixteen children were presented with pairs of images. In half of the trials, the names of the images presented on the screen had matching phonological onsets (e.g. doggy and ducky) that also matched in syllable length (e.g. monosyllabic or disyllabic). In the remaining trials, the names of the images had mismatching phonological onsets (e.g. cake and strawberry). In addition, a portion of each trial type was identity spliced (e.g. informative anticipatory coarticulation) and a portion was cross-spliced (e.g. misleading anticipatory coarticulation). We predicted that if two-year-olds are sensitive to anticipatory coarticulation, then they should be slowest to recognize named targets when the heard label was cross-spliced and the two objects on the screen had mismatching phonological onsets. However, all children looked to the named targets equally fast regardless of trial condition. Thus, no evidence of sensitivity to anticipatory coarticulation was observed.

Published by the Acoustical Society of America through the American Institute of Physics

INTRODUCTION

Language users are highly skilled listeners, rapidly integrating prosodic, rhythmic and subsegmental cues to interpret the unfolding speech signal as efficiently as possible (e.g. Marslen-Wilson & Warren, 1994; Dahan et al., 2001; Shatzman & McQueen, 2006). Coarticulation is one such type of cue, providing listeners with information regarding upcoming speech sounds in the speech stream (Ladefoged, 1993). For example, when a speaker produces the word *smack,* the /s/ is nasalized in anticipation of the upcoming /m/. Research suggests that listeners can precisely use this sort of information to speed online word recognition (Salverda et al., 2007). The usefulness of anticipatory coarticulation during online word recognition is likely affected by many factors, such as speech rate and syllable or segment type preceding the target word (Fowler, 2005).

Although there is evidence that adult listeners use coarticulation to anticipate upcoming words in the speech stream, little research exists on children's use of this information. There is evidence that infants as young as 5 months of age are sensitive to coarticulatory information (Fowler et al., 1990), and that 8-month-olds use this information to find word boundaries in an artificial language (Jusczyk & Johnson, 2001; Curtin et al., 2001). However, no study to date that we are aware of has directly asked whether young children use coarticulation to assist word recognition by anticipating upcoming words in the speech stream.

In the current study, we examined 32- to 34-month-olds' ability to use anticipatory coarticulation during an online word recognition task. Using a child-friendly eyetracking procedure, we presented children with appropriate or misleading coarticulatory information prior to the target word. Target objects visualized on screen were presented along side distractor objects that were matched (e.g. *bike & bee)* or mismatched (e.g. *bike & doggy)* target word onset. Our predictions were as follows. First, when presented with pictures of objects with mismatched onsets (e.g. *cat & strawberry*), word recognition should be more efficient than when presented with pictures of objects with matched onsets (e.g. *cat & cake*). We base this prediction on previous work showing that toddlers are slower to shift their gaze to the target if the competitor shares phonetically similar onsets, such as *doggie* and *doll* (Swingley et al., 1999). Our second prediction is that if children use coarticulatory cues to predict upcoming words, recognition should be most efficient when correct coarticulatory information cues were given (e.g. identity-spliced). Therefore, word recognition would be most efficient for toddlers presented with informative coarticulatory information and distractors that have mismatched onsets from the target word.

METHOD

Participants

Sixteen 32- to 34-month old children (Mean Age=1008 days; Range=974-1035 days; 7 female, 9 male) from the Greater Toronto Area were tested. Parental reports indicated participating toddlers were exposed to a minimum of 90% North American English. Participants were screened to ensure no history of ear infections in the two weeks preceding participation. The data from three additional participants were excluded from our data prior to coding because the children were too fussy to complete the study.

Visual Stimuli

Sixteen test stimuli were presented side by side on a Sony LDC television against a white background. Test stimuli were of similar sizes and roughly matched in interest. Stimuli included: *bee, bike, boat, book, cake, candy, castle, cat, doggy, ducky, goat, goose, strawberry, stroller, train* and *tree*. An initial 'label' phase was included at the beginning of each video to label all target items. Looking to correct target was reinforced during each test trial, where the target object would jiggle at 6000-ms after onset of trial with 1000-ms duration.

Auditory Stimuli

Auditory stimuli were recorded by a native Canadian English female speaker from Southern Ontario in an infant-directed register. Carrier phrases for each test trial began with *'do you see the [adjective]'* (e.g. *pretty, funny, silly, happy*) preceding the target word. All phrases were manipulated using Praat. Matched coarticulatory information (identity-spliced) was spliced to occur with a carrier phrase from another token of the target word. For

mismatched coarticulatory information (cross-spliced), the target word (e.g. *cat*) was spliced to occur with a carrier phrase from the target's mismatched onset distractor (see Figure 2 for an example). Therefore, cross-spliced phrases contained misleading anticipatory cues, suggesting the distractor would be labeled. Cross-spliced targets represented four of the sixteen targets per order, identity-spliced represented twelve of the sixteen targets to ensure that participants did not attribute the speaker as unreliable. An initial prompt (e.g. *"Wow", "Look"*) began 1000-ms after onset of each test trial; onset of each target word began at 4000-ms into the trial. Figure 1 shows a timeline of the auditory stimuli presented during each test trial.
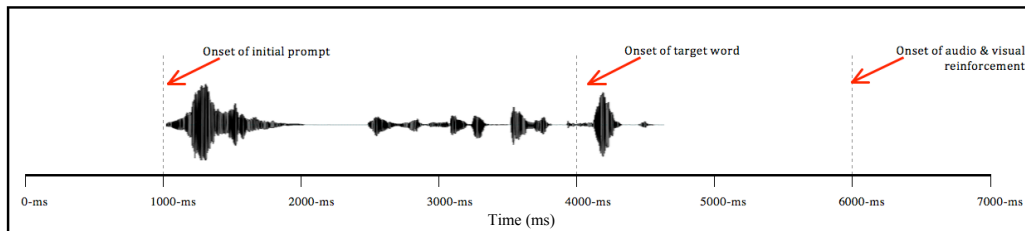

FIGURE 1. A sample timecourse of an experimental trial, with target word onset occurring exactly at 4000-ms into the 7000-ms trial. The target object jiggled at 6000-ms, creating an audio & visual reinforcement.

Design

This study included four trial types: matched onset, cross-spliced; matched onset, identity-spliced; mismatched onset, cross-spliced, mismatched onset, identity-spliced (see Figure 2). Four cross-spliced trial types and twelve identity-spliced trials were presented to each participant, with assignment of target word to trial type counterbalanced across participants. The sixteen target words were presented in a pseudo-randomized order with the constraint that no more than three targets in a row could occur on the same side. Targets were labeled once per participant, with targets appearing equally often on the left and right side of the screen counterbalanced across participants. Half of the targets occurred with a distractor that matched in onset and same syllable length (e.g. *cake* and *cat;* both monosyllabic). The other half were with paired with a distractor that mismatched in onset and syllable length (e.g. *cake* and *strawberry;* one monosyllabic and one polysyllabic). The items pictured on mismatched onset trials differed not only in onset but also number of syllables; thus mismatched cross-spliced trials provided children with two misleading cues to upcoming words: rhythmic structure and coarticulation, to maximize subsegmental cues to word boundaries. Across participants, all targets appeared equally on the left and right sides.

	Matched Phonetic Onset	**Mismatched Phonetic Onset**
Mismatched coarticulatory information (cross-spliced)	[Do you see the funny$_{cat1}$] [cake$_{cake2}$]?	[Do you see the funny$_{strawberry1}$] [cake$_{cake2}$]?
Matched coarticulatory information (identity-spliced)	[Do you see the funny$_{cake1}$] [cake$_{cake2}$]?	[Do you see the funny$_{cake1}$] [cake$_{cake2}$]?

FIGURE 2. Visual representation of each of the four trial types. All participants were presented with each trial type. Assignment of target words to trial type was counterbalanced across participants such that each participant was asked to find each of the sixteen target words only once.

Procedure

The experiment was conducted in a dimly lit double walled sound attenuating IAC booth. The booth contained a central Sony LDC Monitor to present visual stimuli. Auditory stimuli were presented through a concealed speaker behind the monitor. The parent sat on a chair with the child on his or her lap, one meter from the video camera. The child watched a video containing label, filler and test trials on the Sony monitor. Within each test trial, the target and distractor were presented on either side of the screen. The target object was labeled with a verbal phrase (e.g. "Can you find the [target]?). Participants' eye movements were recorded using a video camera situated under the Sony LDC television. In addition, parents wore headphones presented with masking music to reduce parental bias (Pinto et al., 1999).

Coding

Recordings of the child's eye movements were converted from DV tape to movie file before being coded using SuperCoder. All coding was done with the volume muted, and each frame of every 7-second test trials was coded as a look to the left, right or away.

RESULTS

Figure 3 shows a timecourse of the proportion of looking times to the target averaged across trial types.
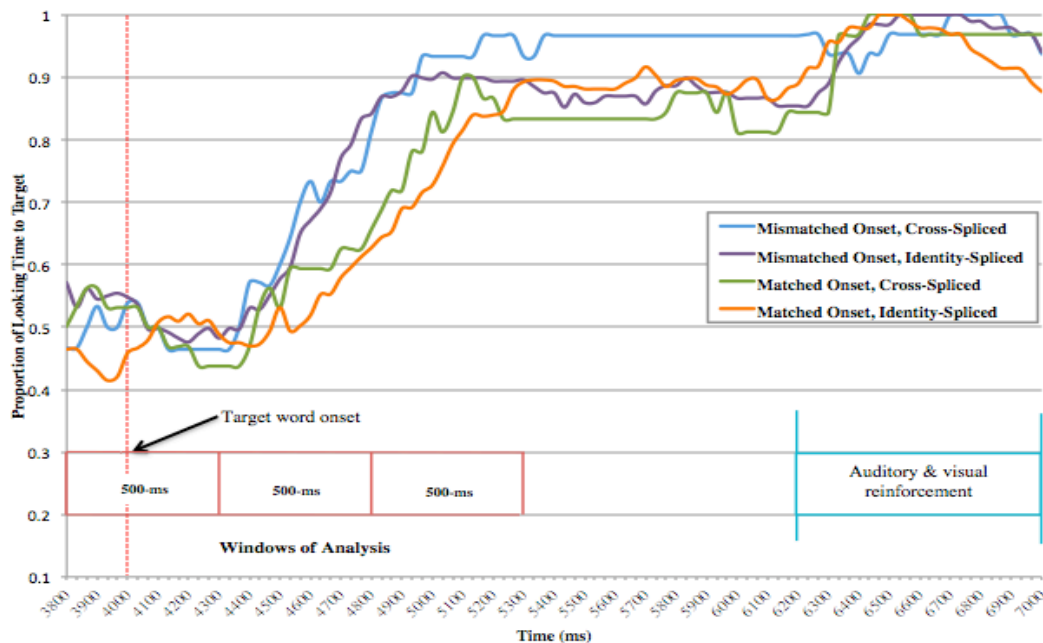

FIGURE 3. Timecourse data representing all four trial types. Note the three 500 ms analysis windows are marked. At 6000-ms into the trial, target words jiggled on-screen.

Proportion of looking time to target was analyzed across three time windows. Each window of analysis lasted 500-ms, with the first window beginning 200-ms prior to target word onset. The second and third window followed immediately after the first window of analysis. The first window started before target word onset and the last window coincided with average target word offset. A 2 (Context: cross spliced versus identity spliced) X 2 (Competitor: matched onset versus mismatched onset) repeated measures ANOVA revealed no significant main effects or interactions in the first two windows of analysis. In the third window of analysis, only the main effect of Competitor was significant, with the proportion of looks to target on different onset trials significantly higher than those to same onset trials, t (15) = 2.817, p=.0065.

In summary, in line with past studies, we found that children recognize words more efficiently when they are presented alongside a competitor that has a mismatching compared to matching onset. However, we found no evidence that children recognize words more efficiently on identity-spliced trials than cross-spliced trials.

DISCUSSION & CONCLUSION

The current study tested the possibility that young children, like adults, rely on coarticulatory information to help them anticipate upcoming words in the speech stream. Although our results replicate earlier work demonstrating that toddlers recognize target words more efficiently when they are presented alongside competitor items that are phonologically distinct from that of the target word (Swingley et al., 1999), we failed to observe any evidence that 32- to 34-month-olds use coarticulatory cues to predict upcoming words in speech. That is, the children we tested recognized target words equally well regardless of whether the targets were spliced into a context containing appropriate or inappropriate coarticulatory cues.

Do the findings of the current study indicate that young children are simply insensitive to coarticulatory information during online word recognition? We believe that this is not necessarily the case. As mentioned in the Introduction, coarticulatory information is strongly affected by factors such as syllable and segment type. In future studies, it might be interesting to manipulate characteristics of the carrier phrases preceding the target items, such as examining how "low resistant" phonemes transmit coarticulation across boundaries. Low resistant phonemes usually do not require fine motor control (e.g. [p]) and are more easily coarticulated to other segments. As certain phonemes tend to enhance or resist adjacent segments, selecting target words or preceding carrier phrases with low resistant phonemes may be better suited (Tobin et al., 2010).

In conclusion, although the children in the current study showed no evidence for using anticipatory coarticulation in a predictive manner, our results suggest toddlers more rapidly shift their gaze towards target words when phonetically similar competitors are absent. Future studies with both children and adults should examine listeners' use of coarticulatory information in a wider variety of phonological contexts.

REFERENCES

Curtin, S., Byrd, D., & Mintz, T. H. **(2001).** "Coarticulatory cues enhance infants' recognition of syllable sequences in speech," Proc. Mtgs. BUCLD. **25**, pp. 190-201.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. **(2001)**. "Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition," Lang. Cog. Processes. **16**, pp. 507–534.

Fowler, C. A., Best, C. T., & McRoberts, G. W. **(1990).** "Young infants' perception of liquid coarticulatory influences following stop consonants," Percept. Psychophys. **48**, pp. 559–570.

Fowler, C. A. **(2005)**. "Parsing coarticulated speech in perception: Effects of coarticulation resistance," J. Phon. **33**(2), 199-213.

Johnson, E.K. & Jusczyk, P.W. **(2001)**. "Word segmentation by 8-month-olds: When speech cues count more than statistics," J. Mem. Lang. **44**, pp. 548-567.

Ladefoged, P. **(1993)**. *A Course in Phonetics*, (Harcourt Brace, Fort Worth, TX), pp. 55-56.

Marslen-Wilson, W. D., & Warren, P. **(1994)**. "Levels of perceptual representations and processes in lexical access: Words, phonemes, and features," Psychol. Rev. **101**, pp. 653-675.

Pinto, J.P., Fernald, A., McRoberts, G.W., & Cole S. **(1999)**. *Reliability and validity in infant auditory preference procedures*. In C. Rovee-Collier, L.P. Lipsitt, & H. Hayne (Eds.), Advances in Infancy Research, (Ablex Publishing Co., Norwood, NJ), pp. 278 – 297.

Salverda, A.P., Dahan, D., Tanenhaus, M.K., Crosswhite, K., Masharov, M., & McDonough, J. **(2007).** "Effects of prosodically-modulated sub-phonetic variation on lexical competition," Cognition. **105**(2). pp. 466–476.

Shatzman, K. B., & McQueen, J. M. **(2006).** "Segment duration as a cue to word boundaries in spoken-word recognition," Percept. Psychophys. **68**(1). pp.1-16.

Swingley, D., Pinto, J.P., and Fernald, A. **(1999)**. "Continuous processing in word recognition at 24 months," Cognition. **71**, pp. 73-108.

Tobin, S. J., Whan Cho, P., Jennett, P. M., & Magnuson, J. S. **(2010)**. "Effects of anticipatory coarticulation on lexical access," Proc. Mtgs. Cognitive Sci. **2010**, 2200-2205.