How many ducks did Heidi see swimming in the pond: Altering context speech rate creates real-time expectations that can cause words to appear and disappear

Meredith Brown (University of Rochester), Laura C. Dilley (Michigan State University), & Michael K. Tanenhaus (University of Rochester)

mbrown@bcs.rochester.edu

Speech rate; Prosody; Expectations; Speech processing; Visual world paradigm; English

Expectation-based approaches in which perceptual input is evaluated with respect to internally generated forward models provide compelling and increasingly influential explanations of phenomena in the perception and motor control literatures [1,2]. Similar forward-modeling approaches may also provide a promising explanatory framework for spoken language processing. Perceptual expectations can readily account for the speed of real-time language processing, listeners' sensitivity to fine-grained contextually-conditioned subphonetic variation, and rapid adaptation to speaker-specific characteristics.

From this perspective, recent studies demonstrating effects of acoustic manipulations early in a spoken sentence on the interpretation of sentence material several syllables downstream are particularly striking [3,4]. For example, manipulations of speech rate distal to the potential location of a function word affect whether listeners report perceiving it, regardless of whether it is present or absent in an utterance [3]. The distal locus of these effects suggests that they are rooted in listeners' expectations about the acoustic-phonetic realization of upcoming segments.

To provide a stronger test of this forward-modeling account, we investigated the time-course of speech rate effects on the interpretation of indefinite articles in a visual-world experiment. Participants (*n*=32) listened to utterances containing a singular or plural expression immediately followed by a sibilant-initial word [6], and selected the picture mentioned in each utterance from a display containing singular and plural alternatives. For singular expressions, two manipulations were conducted to discourage the perception of the acoustically present determiner. In the *proximal-manipulation condition,* the determiner and surrounding segments (the *determiner region*, e.g. the underlined segments in [6]) were sped up; in the *distal-manipulation condition*, the preceding and following utterance context were slowed down. Both manipulations resulted in the determiner region having a faster speech rate than the surrounding context. For plural expressions, corresponding manipulations were conducted to slow the rate of the determiner region relative to the surrounding context and thereby encourage the perception of an acoustically absent determiner.

Consistent with predictions of the forward-modeling account, effects of speech rate emerged shortly after the processing of the determiner region. Proximally- and distally-manipulated singular expressions elicited more fixations to plural pictures than unmanipulated singular expressions, whereas proximally- and distally-manipulated plural expressions elicited more fixations to singular pictures than unmanipulated plural expressions. These findings demonstrate that listeners' expectations about the acoustic realization of spoken words in context are strong enough to cause words to effectively appear or disappear during online comprehension. However, suggestive effects of the duration of the ambiguous sibilant following the target word imply that determiner perception is probabilistic, rather than all-or-none. Longer sibilants resulted in more fixations to plural pictures relative to singular pictures, demonstrating that determiner perception, whether hallucinatory or veridical, can be modulated by phonetic information encountered several syllables later. This finding suggests that listeners maintain uncertainty about previously encountered input [5].

These results set the stage for explicit quantitative tests of forward models of spoken language processing based on cue-reliability, e.g., manipulating the relative reliability of asynchronous determiner- and sibilant-based cues to examine how the strength of listeners' determiner-based expectations affects the integration of these cues.

[1] Jordan MI, Rumelhart DE (1992). *Cognitive Science, 16,* 307–354.
[2] Guenther FH, Micci Barreca D (1997). In P.G. Morasso and V. Sanguineti (eds.), *Self-organization, Computational Maps and Motor Control*, pp. 383–421. Amsterdam: Elsevier-North Holland.
[3] Dilley LC, Pitt MA (2010). *Psychological Science, 21,* 1664–1670.
[4] Brown M, Salverda AP, Dilley LC, Tanenhaus MK (2011). *Psychonomic Bulletin and Review, 18,* 1189–1196.
[5] Levy R, Bicknell K, Slattery T, Rayner K (2009). *PNAS, 106,* 21086–21090.
[6] Heidi sometimes saw (a) white duck(s) swimming in the pond.