Top-down semantic information influences early speech sound encoding

Laura M. Getz & Joseph C. Toscano, Villanova University

Short version

An unresolved issue in speech perception concerns how and when top-down linguistic information influences perceptual responses. We addressed this in two experiments using the event-related potential (ERP) technique to measure cross-modal semantic priming effects on the auditory N100. We provide evidence that top-down lexical information influences early perceptual encoding: listeners encode ambiguous speech sounds consistent with the phonological endpoint predicted by a semantic prime.

Long version

Introduction

One of the foundational questions in psycholinguistics is the extent to which top-down information from lexical representations can impact early speech perception. Early evidence for top-down processing of speech comes from the phoneme restoration effect (Warren, 1970; Samuel, 1981) and the Ganong effect (Ganong, 1980). In phoneme restoration, listeners use surrounding linguistic context to restore a degraded phoneme. Stronger evidence of restoration has been found for longer versus shorter words, words in a sentence versus words in isolation, and words versus non-words, all of which show the importance of top-down lexical context. In the Ganong effect, listeners' placement of the category boundary between phonemes shifts depending on the lexical status of the words at each end of the continuum, such that ambiguous stimuli are perceived consistent with the endpoint that is a word (compared to a non-word). These top-down effects can be instantiated in computational models, such as TRACE (McClelland & Elman, 1986), that show how feedback from the lexicon is used to augment information from the speech stream. However, despite this behavioral and modeling evidence for top-down influences, critics argue that phoneme perception can be explained entirely in a feedforward manner (e.g., MERGE model: Norris, et al., 2000).

We addressed this debate in two experiments using the event-related potential (ERP) technique to measure cross-modal semantic priming effects on initial perceptual encoding via the auditory N100. Previous research has found that this ERP component indexes both acoustic cue encoding (Toscano et al., 2010) and attention (Picton & Hillyard, 1974). If top-down information influences initial speech sound encoding, we predict that activation of a semantically-associated visual prime will affect N100 amplitude to an auditory target more than in a Neutral prime or visual Mask condition. In Experiment 1, we looked at attentional differences between the three conditions. In Experiment 2, we looked specifically at lexical effects on the N100 by adding targets of varying VOTs to investigate whether ambiguous stimuli would be encoded consistently with the prime in the Association condition.

Experiment 1

Method

Our task uses visual-auditory semantic priming, where a visual prime is displayed followed by an auditory target (at three SOAs: 300, 600, and 900ms). Stimuli were selected based on the results of a norming study in which participants (N=56) indicated via free response which word should follow the prime. Twenty-eight stop consonant target words (half voiced; balanced for log word frequency) and their *Association* primes (>85% agreement across subjects) were selected (e.g., marching BAND; amusement PARK). In addition, *Neutral* primes were selected (<15% agreement) to obtain a baseline N100 response in the absence of priming. Association and neutral primes were matched on proportion of proper nouns, alliteration between prime and target, and presence of the target phoneme elsewhere in the prime. A visual *Mask* control condition was also included. Auditory targets were recorded by a male talker in a sound-attenuated booth. Individual tokens were selected based on clarity, and the final 28 targets were equated for mean amplitude.

On each trial, participants (N=21) responded by indicating the sound that the auditory target started with (/b,d,g,k,p,t/) while reaction times and ERP data were recorded. Scalp EEG was recorded with a 64-channel Brain Products actiChamp active electrode system with electrodes at standard 10-20 sites and an average mastoid reference. Participants were seated in a sound-attenuated, electrically-shielded booth, and stimuli were delivered via E-A-RTONE 3A insert earphones.

Results

Behaviorally, participants were faster to respond to the target in the Association prime condition than the Neutral prime and visual Mask conditions. In subjects' brain responses, we find the expected effect of larger N100s for voiced than voiceless targets overall, replicating previous findings. We also found that Association primes produced smaller N100s overall, suggesting that semantically predictive contexts may cause listeners to attend less to subsequent targets.

Although we found an effect of attention on perceptual processing in Experiment 1, we have not specifically shown a top-down *lexical* effect on N100 amplitude. Therefore, in Experiment 2, we added semantically unexpected VOT (i.e., the opposite voicing endpoint) and ambiguous VOT targets so listeners had to equally attend to all prime conditions. We predicted that ambiguous targets may show more top-down influence than stimuli that are clearly voiced or voiceless.

Experiment 2

Method

Experiment 2 incorporated the following changes: (a) all trials used a 600ms SOA, as no significant differences by SOA were found in Experiment 1, (b) only /b,d,p,t/ phonemes were used, and (c) three VOTs were used for each target word. The auditory stimuli were created using the cut-back procedure, where an approximately 0ms, 25ms, and 50ms VOT were cut from the same sound and attached to an identical coda. For example, for the target word "band", listeners could hear a 0ms VOT (i.e., BAND), 50ms VOT (i.e., PAND), or 25ms VOT (ambiguous between BAND and PAND).

Participants (*N*=22) responded by indicating the sound that the auditory target started with while reaction times and ERP data were recorded as in Experiment 1.

Results

Behaviorally, we again found a priming main effect in participants' reaction times. There was also an interaction between prime type and stimulus VOT, with ambiguous stimuli interpreted consistent with the target in the Association prime condition. This interaction was evident in brain responses as well, with a three-way interaction between prime type, stimulus VOT, and original target voicing condition. The N100 amplitude for the ambiguous target VOT was similar to 0ms targets when a voiced target was expected and similar to 50ms targets when a when voiceless target was expected *only* in the Association prime condition, showing a top-down effect of lexical information.

Discussion

In two experiments, we found evidence that top-down semantic information influences initial cue encoding. In Experiment 1, smaller N100s in the Association prime condition suggest that semantic context modulates attention to upcoming targets. In Experiment 2, ambiguous VOTs were encoded similarly to the voicing endpoint elicited by the prime in the Association condition. Our results show that top-down lexical information influences very early perceptual responses, thereby addressing a long-standing issue in spoken language processing.

N=1000

References

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110-125.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1-86.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*(3), 299-325.

Picton, T. W., & Hillyard, S. A. (1974). Human auditory evoked potentials. II: Effects of attention. *Electroencephalography and Clinical Neurophysiology, 36*, 191-200.

Samuel, A. G. (1981). Phonemic restoration: insights from a new methodology. *Journal of Experimental Psychology: General*, *110*(4), 474-494.

Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous perception and graded categorization electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological Science*, *21*(10), 1532-1540.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167*, 392-393.