Research Report

KNOWING A WORD AFFECTS THE FUNDAMENTAL PERCEPTION OF THE SOUNDS WITHIN IT

Arthur G. Samuel

SUNY Stony Brook

Abstract—*Understanding spoken language is an exceptional computational achievement of the human cognitive apparatus. Theories of how humans recognize spoken words fall into two categories: Some theories assume a fully bottom-up flow of information, in which successively more abstract representations are computed. Other theories, in contrast, assert that activation of a more abstract representation (e.g., a word) can affect the activation of smaller units (e.g., phonemes or syllables). The two experimental conditions reported here demonstrate the top-down influence of word representations on the activation of smaller perceptual units. The results show that perceptual processes are not strictly bottom-up: Computations at logically lower levels of processing are affected by computations at logically more abstract levels. These results constrain and inform theories of the architecture of human perceptual processing of speech.*

Humans can recognize tens of thousands of words in their native language, even under listening conditions that pose severe problems for the best computer-based speech recognition systems. For the past half century, researchers have been trying to delineate the architecture of the perceptual process that accomplishes this remarkable feat. Modern theories of word recognition are generally in agreement that there is a mental lexicon that represents the words that we know, and that there are some smaller units (e.g., syllables, phonemes, phonemic features) that are activated in the course of recognizing a spoken word.

Superimposed on this theoretical consensus has been a theoretical question of intense and long-standing contention: Is the flow of information through this system strictly bottom-up, or is there also a top-down component? In a purely bottom-up model, the smallest units are computed first, and the results of these computations are sent on to the next more abstract level, with this increasing abstraction culminating in word recognition. For example, the system could first recognize certain phonemic features (e.g., voicing, place of articulation, and manner of articulation), then recognize the phonemes that comprise these features, and finally recognize the word that is "spelled out" by the sequence of extracted phonemes. In a system with top-down processing, the more abstract representations can actually influence the computation of the units that are logically less abstract. For example, if the representation of a word becomes active, this lexical activation might increase the activation of the phonemes that make up the word, allowing them to be recognized with less acoustic evidence than they would need without the top-down lexical influence. There are currently both strong proponents of models that are purely bottom-up (Cutler, Mehler, Norris, & Segui, 1987; Massaro, 1989; Norris, 1994; Norris, McQueen, & Cutler, 2000) and researchers who in contrast insist on the necessity of top-down lexical influences on phonemic identification (Connine & Clifton, 1987; Marslen-Wilson, 1984; McClelland & Elman, 1986; Samuel, 1996). The present study was designed to provide a critical test to discriminate between these two views. The results clearly support the position that information about a word penetrates downward in a top-down fashion. This penetration causes listeners to hear the sounds of speech in a manner that is based on their knowledge of the word.

Although there have been many demonstrations of lexical information influencing phonemic responses (e.g., Connine, Titone, Deelman, & Blasko, 1997; Marslen-Wilson & Warren, 1994; McQueen, Norris, & Cutler, 1999; Newman, Sawusch, & Luce, 1997; Samuel, 1981), proponents of purely bottom-up models have argued that these results do not truly reflect top-down perceptual effects. Instead, these demonstrations have been interpreted as merely showing that words can influence phoneme responses at a decision stage that occurs postperceptually. Consider, for example, a study by Ganong (1980). Ganong generated speech continua that varied in the voicing feature of the initial consonant, such as /dæ/–/tæ/. If /š/ is appended to each member of this continuum, the resulting stimuli sound like the word "dash" at one end and the nonword "tash" at the other. If, instead, /sk/ is appended to members of the same continuum, the nonword "dask" is obtained at the /d/ end and the word "task" is obtained at the other. For an item near the middle of the /dæ/–/tæ/ continuum, this creates an interesting situation: Such an item corresponds to a word if the initial consonant is heard as a /d/ when the ending is /š/, but the same sound must be heard as a /t/ when the ending is /sk/ if the item is to be heard as a word. If top-down models are correct, then listeners should hear this ambiguous initial consonant as a /d/ in the context of "_ash" and as a /t/ in the context of "_ask."

Exactly this result has been observed, in many studies (e.g., Connine & Clifton, 1987; Ganong, 1980; Pitt & Samuel, 1993). However, theorists who do not believe in top-down perceptual effects argue that the data cannot be unambiguously interpreted. In their view, the initial consonant is not perceived differently in the "_ash" and "_ask" contexts: In both cases, it is heard as ambiguous. But because the experiment requires one response or another, the lexical context biases the responses postperceptually. From this perspective, the perceptual flow of information remains purely bottom-up. Only the later decision stage is influenced by the lexical context. Nonperceptual interpretations of this sort have been offered for virtually all the demonstrations of lexical influences on phonemic judgments.

As long as a task requires listeners to make phonemic decisions, it is not possible to rule out a decision-level locus for any lexical influence on phoneme processing. With this constraint in mind, in our laboratory we have conducted a series of experiments that provide a test for top-down lexical influences on phoneme perception, using conditions that rule out a decision-level interpretation. Preliminary evidence for a true top-down lexical effect was reported in a previous study (Samuel, 1997). The current study extends this effort, and includes a crucial new procedure that conclusively demonstrates that lexical knowledge does indeed penetrate down into the perceptual encoding of speech sounds.

Address correspondence to Arthur G. Samuel, Department of Psychology, SUNY Stony Brook, Stony Brook, NY 11794-2500; e-mail: asamuel@ms.cc.sunysb.edu.

In these experiments, the key innovation is looking for evidence of top-down influences in the *consequences* of having heard a lexically driven phoneme, rather than asking listeners to identify the phoneme. By not requiring any phonemic decision, we avoid any opportunity for decision-level factors to drive the results. The particular consequence examined in this research was the occurrence of a selective-adaptation effect.

Selective adaptation is a widely used psychophysical procedure (e.g., Albrecht, Farrar, & Hamilton, 1984; Blakemore & Sutton, 1969; Green, 1980). Eimas and Corbit (1973) showed that this procedure works well with speech stimuli, using a test series that varied in voicing, /da/–/ta/. In a baseline condition, listeners identified members of this continuum many times. Then, in the adaptation phase, continuum items were interspersed with repeated presentations of a continuum endpoint (the adaptor). As in other psychophysical adaptation studies, the repeated presentation reduced the likelihood of perceiving stimuli that were similar to the adaptor. After hearing /da/ repeatedly, listeners identified fewer test stimuli as /da/; the /ta/ adaptor reduced reports of /ta/.

Many experiments have explored speech-adaptation effects (e.g., Cheesman & Greenwood, 1995; Samuel, 1986; Samuel & Kat, 1996; Sawusch, 1977). For the present purposes, two results are most germane. First, adaptation effects can be generated at three different levels of perceptual analysis: (a) a very early, monaurally driven level; (b) a binaurally driven level that combines acoustic properties, such as different formant paths; and (c) a binaurally driven level that appears to be phonemelike (Samuel & Kat, 1996; Sawusch, 1986; Sawusch & Nusbaum, 1983; Sawusch & Pisoni, 1976). Second, no adaptation has been found at the word level itself. A previous study using four different tests (Samuel, 1997) repeatedly demonstrated that there are no direct adaptation effects at the lexical level.

The localization of adaptation effects at three sublexical levels, together with the clear absence of adaptation at the lexical level itself, guarantees that the phenomenon is appropriate for our purposes. Our interest in adaptation in the current study lay in the potential for adaptation to reveal top-down lexical influences on phonemic processing. The first application of this potential (Samuel, 1997) tested whether lexical information could determine the identity of a phoneme, resulting in adaptation effects on the identification of test syllables. Because adaptation is limited to sublexical processing, such an effect would necessarily entail top-down lexical influences on lower-level perception. In this study, the lexical determination of a phoneme was accomplished through phonemic restoration (Samuel, 1981, 1996; Warren, 1970; Warren, Obusek, & Ackroff, 1972). In phonemic restoration, a portion of a word (typically, a phoneme plus its transitions to adjacent phonemes) is excised and replaced by a broadband sound such as white noise. When the resulting stimulus is played to listeners, they cannot indicate the location of the replacement (Warren, 1970), and they show very poor discrimination of this stimulus from one in which noise is merely superimposed on the corresponding portion of the waveform (Samuel, 1981, 1996). Listeners appear to have perceptually restored the missing speech.

Samuel (1997) selected five words with /b/ in the third syllable (e.g., "exhibition") and five with /d/ in the third syllable (e.g., "armadillo"). For all 10 words, two types of stimuli were made. Experimental items were made using the noise-replacement procedure just described, deleting and replacing each /b/ or /d/. Control items were made by deleting exactly the same portions of each waveform, without filling the resulting silent gap with noise. Previous research had shown that the experimental items would produce phonemic restoration (i.e., listeners would hear the /b/ or /d/, determined by the lexical context), but the control items would be heard veridically (i.e., as words with silent gaps). Listeners identified members of a /bɪ/–/dɪ/ test series, before and after adaptation with either the /b/-based or the /d/-based items, for both the experimental and the control stimuli. The experimental items produced reliable adaptation effects: Listeners identified fewer syllables as /bɪ/ after hearing the words with the /b/ replaced than after hearing the words with the /d/ replaced. As expected, the control items produced no such shifts. These results meet the criteria for a true top-down effect: The lexical context determined what phoneme was heard (/b/ or /d/), and the resulting phoneme (which listeners never had to report) produced the desired consequence—an adaptation shift at a sublexical level.

We have recently completed a new set of experimental conditions that confirm and extend this finding. These conditions used Ganong's (1980) procedure, rather than phonemic restoration, to impose lexical information on phonemic perception. The appeal of this procedure is that it requires that exactly the same acoustic stimulus be heard as different phonemes, depending on the lexical context (in the restoration work, the replacement noise differed from stimulus to stimulus, preserving the duration and amplitude envelope of the replaced phoneme). The current study also included a test that completely controlled for any acoustic artifacts due to coarticulation, a control that was not included in the previous study. As in that study, the listeners never reported the lexically determined phoneme. Instead, we assessed its identity by its ability to shift the identification of other speech sounds, via adaptation.

METHOD

Materials

Four words that end with the phoneme /s/ ("bronchitis," "embarrass," "malpractice," "tremendous") and four words that end in /š/ ("abolish," "demolish," "diminish," "replenish") were selected. These words were chosen because they were expected to produce strong lexical effects by virtue of their length and the word-final position of the critical phoneme, and because they have the same vowel preceding the final phoneme in a mid-Atlantic American English dialect, as verified with the Carnegie Mellon Pronunciation Dictionary (1995) and by directly measuring the frequencies of the first two formants. These eight words and the syllables /ɪs/ and /ɪš/ were digitized (16 kHz A/D, 7.8-kHz low-pass filter) and stored on computer disk. The /ɪs/ and /ɪš/ tokens were digitally mixed in various proportions, to produce an eight-step /ɪs/–/ɪš/ test series.

Experimental and control items were constructed to serve as adaptors. The control items were made by removing the final consonant (/s/ or /š/) from each word, along with enough of the preceding vowel to eliminate any perception of the final consonant. The experimental items were copies of the control items, with one change: An identical piece of sound was appended to each of the eight control items. This addition was an ambiguous consonant sound, taken from one of the middle items of the /ɪs/–/ɪš/ test series. This ambiguous consonant was expected to be identified as /s/ after contexts such as "bronchiti_", but as /š/ in contexts like "demoli_." The critical theoretical question is whether this identification reflects true lexically determined perception of this consonant or is merely a postperceptual decision effect. If the effect is perceptual, then the resulting phonemic percept should produce adaptation effects on the identification of other sounds, even

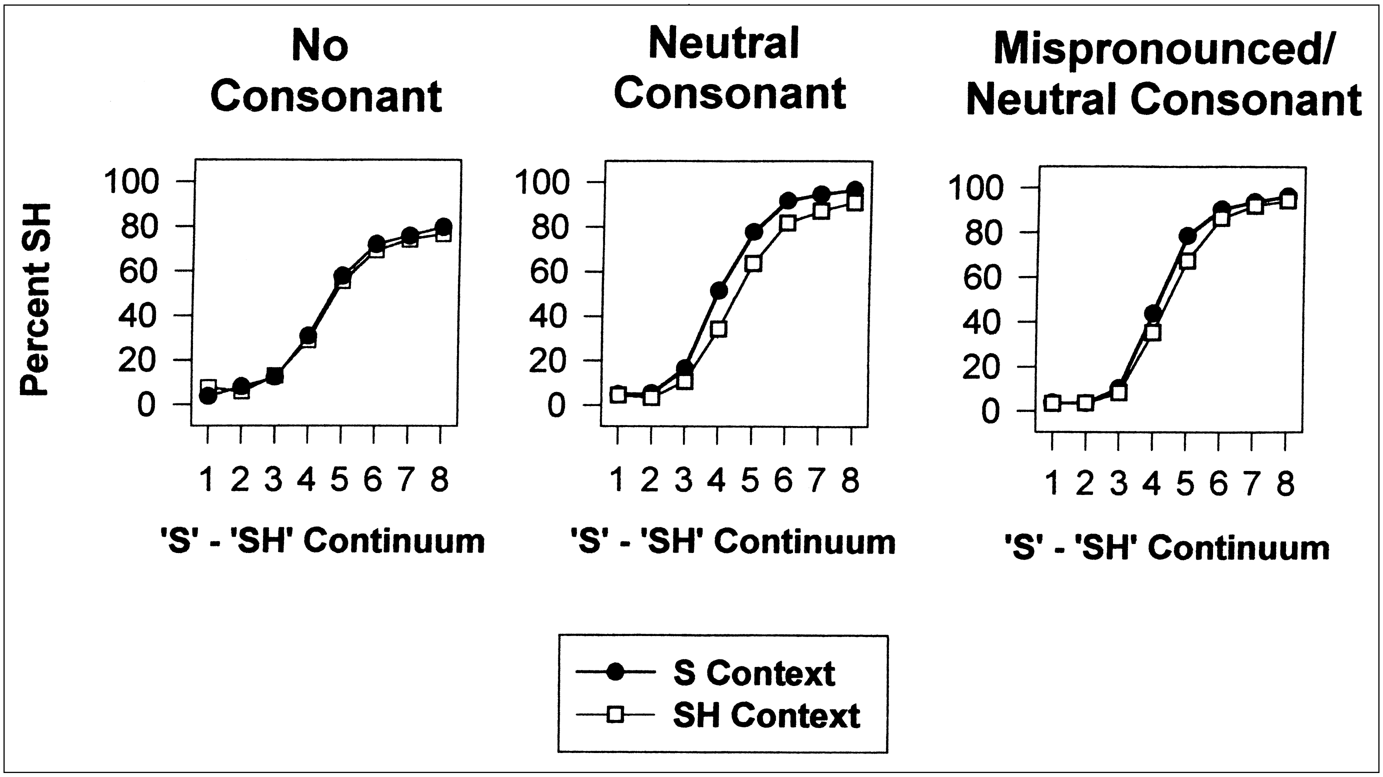

Fig. 1. Percentage of test syllables on the /ɪs/–/ɪš/ continuum identified as /ɪš/. Results are shown separately for test syllables presented after no-consonant control items as adaptors (left panel), experimental items with neutral consonants as adaptors (middle panel), and mispronounced words with neutral consonants as adaptors (right panel). In "s contexts," the adaptors were derived from words that end in /s/; in "sh contexts," the adaptors were derived from words ending in /š/.

though listeners never had to report the identity of the sounds in the adaptors.

Participants

Forty listeners participated in each of two adaptation sessions (one with adaptors based on /s/ words and one with adaptors based on /š/ words). One participant's data were eliminated because the participant was unable to identify the test syllables consistently. Half the listeners received the experimental items as adaptors; the other half received the control items.

Procedure

The task was to identify members of the /ɪs/–/ɪš/ test series as ending in either /s/ or /š/. The test syllables were played in blocks of eight randomly ordered items, alternating with 30-s adaptation blocks. During each adaptation block, eight randomizations of the four adaptors for that session were presented. Participants simply listened during these adaptation phases. They responded to the test syllables by pushing one of two labeled buttons ("s" or "sh"). Each session included 24 identification-adaptation passes. Order of sessions (/s/ words vs. /š/ words) was counterbalanced.

RESULTS AND DISCUSSION

Figure 1 shows the resulting identification functions for the test syllables, broken down by the lexical context (/s/ words or /š/ words). For the no-consonant control items (left panel), no differential adaptation occurred; there was no information in those truncated words to drive any adaptation shifts. When exactly the same stimuli were followed by a neutral sound (middle panel), a very different result obtained: Listeners reported fewer syllables as "ish" when the adaptors were made from words that normally end in /š/ than when the adaptors were derived from /s/-final words. A *t* test using the difference in average "sh" report as a function of adaptor showed that the shift was robust, $t(18) = 3.60$, $p = .002$. In contrast, there was no difference for the control items, $t(19) = 0.51$, n.s.

A third condition provides even more definitive evidence for top-down lexical effects. The eight words were rerecorded, but with the final consonant of each word intentionally mispronounced (/s/ and /š/ substituted for each other; e.g., "aboliss," "bronchitish"). Experimental items were made by deleting these final consonants, and replacing them with a neutral "s/sh" (individually selected for each listener to be neutral, based on a pretest). In addition, most of the vowel preceding the final consonant was deleted and replaced by white noise. Retaining the first couple of the vowel's pitch periods meant that a strong percept of the vowel could be generated by the perceptual restoration process, but at the same time the strongest coarticulatory cues to the following fricative were eliminated. These procedures produced adaptors in which any possible residual acoustic information would work *against* the lexically driven adaptation shifts. Adaptation was conducted as before, with 21 new listeners. As the right panel of Figure 1 shows, although the mispronunciation of the words stacked the deck against the top-down effect, a reliable shift was found, $t(20) = 2.61$, $p < .02$. This result rules out any possible acoustic artifact as an explanation of the effect.

The results of previous studies (Samuel, 1997; Samuel & Kat, 1996) have clearly shown that adaptation shifts occur at early acoustic-phonetic levels of perceptual analysis, and not at the lexical level itself.[1] The information that determined the identity (and therefore, the adapting property) of the critical phonemes in the current study was lexical: The same physical sound was an /s/ in the context of "bronchiti_," but was an /š/ in the context of "aboli_." Taken together, these findings and the results of our previous study (Samuel, 1997) demonstrate that the perceptual system includes top-down connections from the lexical level to the phonemic: Listeners *hear* speech sounds in accord with the representations of words stored in their mental lexicons.

Acknowledgments—This work was supported by National Institute of Mental Health Grant R01-51663. I thank Donna Kat and Mark Pitt for their help with this project.

REFERENCES

Albrecht, D.G., Farrar, S.B., & Hamilton, D.B. (1984). Spatial contrast adaptation characteristics of neurones recorded in the cat's visual cortex. *Journal of Physiology*, *347*, 713–739.

Blakemore, C., & Sutton, P. (1969). Size adaptation: A new aftereffect. *Science*, *166*, 245–247.

The Carnegie Mellon Pronouncing Dictionary (Version 4) [Electronic database]. (1995). Pittsburgh, PA: Carnegie Mellon University.

Cheesman, M.F., & Greenwood, K.G. (1995). Selective adaptation by context-conditioned fricatives. *Journal of the Acoustical Society of America*, *97*, 531–538.

Connine, C.M., & Clifton, C. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 291–299.

Connine, C.M., Titone, D., Deelman, T., & Blasko, D. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language*, *37*, 463–480.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, *19*, 141–177.

Eimas, P.D., & Corbit, J.D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, *4*, 99–109.

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

Green, M. (1980). Orientation-specific adaptation: Effects of checkerboards on the detectability of gratings. *Perception*, *9*, 369–377.

Marslen-Wilson, W.D. (1984). Function and process in spoken word recognition: A tutorial review. In H. Bouma & D.G. Bouwhuis (Eds.), *Attention and performance X* (pp. 125–150). Hillsdale, NJ: Erlbaum.

Marslen-Wilson, W.D., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, *101*, 653–675.

Massaro, D.W. (1989). Testing between the TRACE model and the Fuzzy Logical Model of Speech Perception. *Cognitive Psychology*, *21*, 398–421.

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

McQueen, J.M., Norris, D.G., & Cutler, A. (1999). Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology*, *25*, 1363–1389.

Newman, R.S., Sawusch, J.R., & Luce, P.A. (1997). Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 873–889.

Norris, D.G. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*, 189–234.

Norris, D.G., McQueen, J.M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*, 299–370.

Pitt, M.A., & Samuel, A.G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 699–725.

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

Samuel, A.G. (1986). Red herring detectors and speech perception: In defense of selective adaptation. *Cognitive Psychology*, *18*, 452–499.

Samuel, A.G. (1996). Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General*, *125*, 28–51.

Samuel, A.G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive Psychology*, *32*, 97–127.

Samuel, A.G., & Kat, D. (1996). Early levels of analysis of speech. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 676–694.

Sawusch, J.R. (1977). Peripheral and central processes in selective adaptation of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, *62*, 738–750.

Sawusch, J.R. (1986). Auditory and phonetic coding of speech. In E. Schwab & H. Nusbaum (Eds.), *Pattern recognition by humans and machines* (pp. 51–88). Orlando, FL: Academic Press.

Sawusch, J.R., & Nusbaum, H.C. (1983). Auditory and phonetic processes in place perception for stops. *Perception & Psychophysics*, *34*, 560–568.

Sawusch, J.R., & Pisoni, D.B. (1976). Response organization in selective adaptation to speech sounds. *Perception & Psychophysics*, *20*, 413–418.

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

Warren, R.M., Obusek, C.J., & Ackroff, J.M. (1972). Auditory induction: Perceptual synthesis of absent sounds. *Science*, *176*, 1149–1151.

(Received 7/1/00; Revision accepted 11/7/00)

1. The likely locus of the adaptation effect in the current study is the "categorical" (third) level (Samuel & Kat, 1996). Shifts at this level appear to be mediated by criterion shifts in phonemic identification. Critically, in the current study, these shifts were necessarily determined by lexical influences, with the shifts affecting test syllables presented *after* listeners had heard the adaptors.