Available online at www.sciencedirect.com

SCIENCE DIRECT®

ACADEMIC PRESS

Journal of Memory and Language 48 (2003) 416–434

Journal of Memory and Language

www.elsevier.com/locate/jml

Lexical activation (and other factors) can mediate compensation for coarticulation

Arthur G. Samuel[a,*] and Mark A. Pitt[b]

[a] *SUNY Stony Brook, Department of Psychology, Stony Brook, NY 11794-2500, USA*
[b] *Ohio State University, Columbus, OH 43210, USA*

Received 31 May 2002

Abstract

A key dispute in theories of spoken word recognition is whether activation of a lexical representation can affect the perception of sublexical components, such as phonemes. Elman and McClelland (1988) provided evidence for such top–down processing by showing that a prelexical process (compensation for coarticulation) could be affected by lexical activation. However, Pitt and McQueen (1998) reported that the observed compensation effects were in fact due to the transitional probability of certain phonemic sequences, rather than lexical activation. Part I of the current study shows that perceptual grouping must be considered in assessing compensation effects, reopening the question of whether lexical activation can drive compensation. Part II shows that when perceptual grouping is taken into account, lexically mediated compensation can indeed be observed, confirming the interactive nature of speech processing. Collectively, the results make it clear that compensation for coarticulation is affected by a number of factors, making it difficult to isolate lexical influences on the phenomenon.

Keywords: Lexical processing; Speech perception; Interactive models; Compensation for coarticulation

Speech recognition systems, whether human or machine, face a daunting mix of both too much and too little information. There is too much information in the sense that the initial representation must contend with a huge amount of frequency, amplitude, and timing data. There is too little information in the sense that a word in normal conversational speech often lacks many of the properties that seem to define it in its citation form; many studies have shown that the intelligibility of words taken out of conversation is very poor indeed (e.g., Bard, Schillcock, & Altmann, 1989; Pollack & Pickett, 1963, 1964).

One way to view the recognition process is as a data reduction mechanism. The input contains a wealth of acoustic–phonetic information, and the goal is to extract the sequence of words from this dense acoustic–phonetic array. Details of how a particular phoneme was realized are not needed for this output goal, and in fact such details must be smoothed over: "dog" must be recognized regardless of whether the /d/ was produced with a voice onset time of 5, 10, 15 ms, or any other particular value. Even if an impressive amount of the acoustic detail is ultimately retained in memory (as suggested by Goldinger, 1998; Nygaard, Sommers, & Pisoni, 1994), word recognition essentially involves mapping a large amount of acoustic information onto a particular representation—the intended word.

If one views word recognition as data reduction, then a model of word recognition is a description of the process by which the representation becomes progressively more abstract. This description should provide answers to questions such as, (1) What types of

* Corresponding author. Fax: 631-632-7876.
E-mail address: asamuel@ms.cc.sunysb.edu (A.G. Samuel).

0749-596X/02/$ - see front matter © 2002 Elsevier Science (USA). All rights reserved.
PII: S0749-596X(02)00514-4

abstractions (if any) occur between the richly detailed auditory starting point, and the ultimately activated lexical entry?, (2) When do less abstract representations pass information onward to more abstract ones?, and (3) Is the flow of information strictly from less abstract to more abstract, or can information from a more abstract representation affect processing of a less abstract form?

Research over the last two decades has supported at least preliminary answers to these questions. For example, many theorists have concluded that subsyllabic representations (e.g., phonemes, syllable onsets, and rimes) are computed by the perceptual system (e.g., McClelland & Elman, 1986; Nearey, 1990; Treiman, 1986; Treiman & Zukowski, 1996; but see Gaskell & Marslen-Wilson, 1997, for an alternative view). Similarly, there is relatively broad agreement that less abstract representations do not need to be fully computed before their contents become available to more abstract representations, a view that was first developed explicitly by McClelland (1979) in his Cascade model.

In contrast to these two points of relative consensus, there has been a longstanding sharp point of contention regarding the flow of information during word recognition. A number of theorists have argued that information flow is strictly unidirectional, from less abstract to more abstract (e.g., Cutler, Mehler, Norris, & Segui, 1987; Cutler & Norris, 1979; Eimas, Marcovitz Hornstein, & Payton, 1990; Massaro & Oden, 1995; Norris, McQueen, & Cutler, 2000). Others (e.g., Connine & Clifton, 1987; McClelland & Elman, 1986; Samuel, 1981, 1997) have suggested that word recognition also includes top–down information flow, in which activation of more abstract forms (e.g., words) can influence the activation of less abstract ones (e.g., phonemes).

At this point, there are many findings in the literature that demonstrate lexical influences on listeners' phonetic judgments. For example, monitoring for a target phoneme is faster in words than in matched pseudowords (e.g., Rubin, Turvey, & van Gelder, 1976); ambiguous phonemes are identified in accordance with their lexical context (e.g., Connine & Clifton, 1987; Ganong, 1980; Pitt, 1995; Pitt & Samuel, 1993); missing phonemes are perceptually restored more in words than in matched pseudowords (Samuel, 1981, 1996). Moreover, there are also studies showing that processing of phonemes within pseudowords depends on how similar the pseudowords are to real words (Connine, Titone, Deelman, & Blasko, 1997; Wurm & Samuel, 1997).

The accumulation of such evidence has made it clear that models of speech processing must include a mechanism that allows lexical activation to affect phonetic judgments. However, a remaining point of contention concerns exactly how lexical activation might influence phonetic judgments. In interactive models, as described above, this influence involves bidirectional connections between phonemic codes and lexical representations. Norris et al. (2000) have recently proposed an alternative conceptualization that they consider to be more consistent with the autonomous perspective. In their Merge model, lexical activation can influence phonetic decisions, but not phonetic perception. For example, listeners might report an ambiguous initial phoneme as /d/ if the remainder of the utterance is /æʃ/ ("ash"), because /dæʃ/ ("dash") is a word. But, this does not necessarily mean that the initial phoneme was actually perceived as more /d/-like than it would have been in a different context (e.g., in /dæsk/ ("dask")).

Although there are a number of grounds for questioning the Merge architecture (see the set of commentaries following the Norris et al., 2000, paper), choosing between a model of this sort and an interactive one must ultimately be determined empirically. The separation of the decision process from the perceptual process in Merge's architecture places severe constraints on any such empirical test. Tests in which listeners report on phonemes within words (or even within pseudowords resembling words) can potentially be handled by both Merge and interactive models. This is due to the fact that both approaches have mechanisms that allow activation of a word to affect the report of phonemes within that word. Any changes in responding, even when measured as changes in perceptual sensitivity (Pitt, 1995; Samuel, 1996) cannot be unambiguously attributed to perceptual processes when a separate decision stage is also postulated.

Distinguishing the models empirically is, therefore, a very subtle affair. To do so requires a task that measures lexical influences on phonetic processing indirectly. There are two lines of evidence in the literature that appear to have the necessary degree of indirection. One of these is a set of experiments using the selective adaptation paradigm in a lexical context (Samuel, 1997, 2001). In such studies, listeners identify simple syllables (e.g., /bI/ versus /dI/) before and after an adaptation phase in which an "adaptor" is presented repeatedly for a short time (e.g., 60 s). Many adaptation studies, beginning with Eimas and Corbit's (1973) seminal work, have shown that repeated presentation of the adaptor reduces report of sounds that are similar to it. Samuel (1997) used lexically based phonemic restoration to induce the perception of phonemes that could then potentially produce an adaptation effect. For example, the /b/ in words like "alphabet" and "inhibition" was replaced by white noise. Similarly, the /d/ in words like "armadillo" and "academic" was replaced. Previous work has shown that listeners use the lexical context to perceive the missing phonemes (Samuel, 1981, 1987, 1996). Samuel (1997) found that such lexically restored phonemes produced significant adaptation effects: Identification of the /bI/-/dI/ syllables was affected by hearing the restoration stimuli. Samuel (2001) has shown similar effects, using a different lexical manipulation to produce phonemic codes.

The critical feature of experiments of this sort is that the lexical influence on phonemic perception is not assessed directly: Listeners never report what they hear when the restoration items (e.g., "alphabet") are presented. Instead, lexical influences are inferred as a result of their consequential effects. Because the lexical context causes a /b/ or a /d/ to be heard, the /b/ or /d/ is able to affect the perception of syllables heard afterward. This result is predicted by interactive models, because in such models the lexical activation causes increased activation in the units involved in perceiving the word's constituent phonemes. Autonomous models, including Merge, cannot account for such consequential effects because they do not allow lexical activation to affect a constituent phoneme's perception.

The current study is concerned with a testing situation that shares the critical property of using a consequential measure of lexical activation. Rather than selective adaptation, this approach is based on the compensation for coarticulation phenomenon originally studied by Mann and Repp (1981) and Repp and Mann (1981, 1982). Mann and Repp found that the identification of syllable-initial stop consonants was influenced by the identity of a preceding fricative. In particular, listeners reported more velar stops (/g/ or /k/) following /s/ than following /ʃ/; they reported more alveolar stops (/d/ or /t/) following /ʃ/ than /s/. Mann and Repp explained this consequential effect as being a natural perceptual complement to the coarticulatory consequences of these sequences. The place of articulation of the preceding fricative alters the place of articulation of the following stop, and the observed perceptual effect is what should occur if the perceptual system compensates for these coarticulatory effects.

Because this compensation is clearly a relatively low-level prelexical phenomenon, Elman and McClelland (1988) realized that it presented an opportunity to test whether lexical activation can influence lower level representations. They constructed an ambiguous fricative that was perceptually between /s/ and /ʃ/, and used lexical contexts to influence how listeners identified this fricative. For example, when the neutral fricative was appended to /fulI/ ("fooli"), listeners should hear it as the /ʃ/ in "foolish," but after /krIsmI/ ("Christma") it should be /s/. The critical question for Elman and McClelland was whether these lexically determined fricatives would alter listeners' identification of following stop consonants. In a series of experiments, this is exactly what they found. Identification of words beginning with /t/ versus /k/ (e.g., /teIps/ ("tapes") versus /keIps/ ("capes")), or /d/ versus /g/ (e.g., /deIts/ ("dates") versus /geIts/ ("gates")) depended on the perception of the preceding fricative, even though the fricative was acoustically identical in the two situations. Given that compensation is clearly prelexical (a view strongly endorsed by Norris et al., 2000), these results appear to offer clear evidence for lexical activation having a top–down effect on phonetic perception.

However, as Elman and McClelland acknowledged, there was a possible confound in their stimuli, because the vowels preceding the final fricatives in their context words were not exactly the same (we will discuss Elman and McClelland's treatment of this issue below). Shillcock, Lindsey, Levy, and Chater (1992) pursued the potential artifact in Elman and McClelland's study, using a statistical analysis of English words to show that the fricatives could in theory be predicted by the preceding vowel, rather than by the lexical context as a whole. Such an effect of transitional probability would not constitute the kind of evidence needed to support top–down lexical influences on phonetic perception.

Pitt and McQueen (1998) conducted a set of experiments to determine whether the observed effects were really due to transitional probability, rather than lexical activation. These experiments dissociated the two possibilities very cleanly, by using either lexical contexts with matched vowel–fricative transitional probabilities (e.g., /dʒus/ ("juice") versus /bUʃ/ ("bush")), or nonlexical contexts with intentionally mismatched transitional probabilities (e.g., /dɚs/ ("ders") versus /neIʃ/ ("naish")). In several experiments, Pitt and McQueen found that the word contexts failed to produce compensation when the final fricative was ambiguous, whereas the pseudowords varying in transitional probability were successful under these conditions. These results seem to confirm Shillcock et al.'s concern that Elman and McClelland's results might be due to nonlexical factors.

Because of the critical role Elman and McClelland's findings have had in current theorizing, it is essential to know whether or not compensation for coarticulation really can be lexically mediated. We undertook an investigation of this issue, based in part on a concern that Pitt and McQueen considered and rejected: Perhaps the lexical conditions were not powerful enough to find what could be a rather small effect. As they noted, "the words were only three phonemes long, providing little context (or time) for biasing influences in fricative perception to emerge" (p. 357). Pitt and McQueen rejected this possibility on the basis of clear evidence showing lexical influences on labeling of the ambiguous fricative itself. However, it is possible that compensation (a between-word effect) may require a higher level of lexical activation than that required to bias the labeling of an ambiguous sound (a within-word effect).

Thus, a major motivation for the current study was our concern that simple monosyllabic words might not be sufficient to reliably produce lexically driven compensation. A second concern arose from a consideration of other factors that affect compensation. Mann and Repp (1981, Experiment 1) showed that the size of the compensation effect is influenced by the context preceding the fricative. When the fricative was presented alone with the stop + vowel syllable (e.g., /sta/), compensation was considerably larger (about 59% larger, based on the data in their Figs. 1 and 2) than when the preceding context was vowel + fricative (e.g., /asta/). A possible cause of this diminished effect is that the context preceding the fricative (in this case a single vowel) affected the degree to which the fricative perceptually cohered with the following stop. In effect, introduction of the vowel altered the perceptual grouping of the phonemes, loosening the bond between the fricative and stop and tightening it with the vowel, which resulted in reduced compensation.

If this interpretation is correct, then lexical status might also regulate how tightly the fricative binds with the stop, and thus the amount of compensation. When the context forms a word, the fricative might bind more tightly to the rest of the word, and, therefore, more loosely with the stop, than when the context forms a pseudoword. In fact, Shoaf and Pitt (2002) and Pitt and Shoaf (2001) have shown that the tendency for speech to perceptually transform when it is repeated (the verbal transformation effect; see below) depends on the lexical status of the original speech, with words being less likely to have their phonemes split off than nonwords are. With Pitt and McQueen's stimuli, this would imply that the final /s/ in /dʒus/ ("juice") and the final /ʃ/ in /bUʃ/ ("bush") were more likely to cohere with the rest of the syllable than the final /s/ in /dɚs/ or the final /ʃ/ in /neIʃ/. If compensation requires the fricative to group with the following stop (while still being influenced by its preceding context), then the pseudowords would be expected to be more effective because of the looser binding of their fricatives.

These two concerns could interact in a most pernicious way. The first concern is that longer, "better" lexical contexts may be needed to produce sufficient lexical activation to generate compensation (Elman and McClelland's context words were 2–4 syllables long). However, the second concern is that words may bind their final fricatives more than pseudowords, with binding strength being positively related to degree of lexical activation. If both of these concerns are valid, then there may only be very small regions of the parameter space in which lexically mediated compensation would be detectable. Elman and McClelland (1988) may have been fortunate to have tapped this region.

Part I of the current study examines whether the pattern of results reported by Pitt and McQueen (1998) could depend on differences in the perceptual grouping of the fricatives in their words versus their pseudowords. In Part II, we report the results of several compensation experiments involving context words of one, two, or three syllables, in tests that use manipulations of stimulus timing to try to control for differences in the perceptual grouping of final fricatives. The question addressed in these compensation experiments is whether the prelexical process of compensation for coarticulation can be lexically determined when the potential binding of the final fricative is controlled for.

Part I: Are final fricatives more tightly bound to words than to pseudowords?

Experiment 1

The idea that the results of Pitt and McQueen could be due to differences in the strength of fricative cohesion in word and pseudoword contexts follows logically from recent work examining what the Verbal Transformation Effect (VTE) can tell us about word recognition. The VTE (Warren, 1961, 1968) is an illusion in which listeners report illusory words and pseudowords while hearing an utterance repeat over and over. Analyses of the transformations that listeners report provide insight into the characteristics of the phenomenon. Three recent studies of the VTE are relevant to the present investigation. Shoaf and Pitt (2002) and Pitt and Shoaf (2001) replicated and extended a past result showing that lexical status of the repeating stimulus affects listeners' reports. Specifically, they found that repeating words were heard far more often as the veridical stimulus than repeated pseudowords, suggesting that the lexical representation of words makes them much more stable percepts.

Another set of experiments (Pitt & Shoaf, 2002) demonstrated that perceptual reorganization (Bregman, 1990) is one cause of the VTE. Repetitive presentation of an utterance (e.g., /pɛʃ/) causes phonemes at the beginning or end of the utterance (e.g., /p/ or /ʃ/) to split off and form a separate perceptual stream, causing listeners to report the remainder (e.g., /pɛ/) as a verbal transformation. Fricatives were particularly prone to splitting off, presumably because of their extreme difference in frequency and periodicity from the preceding vowel.

When these two results are combined, they lead to the prediction that words should yield relatively few transformations in which the boundary phonemes split off (i.e., "grouping" transformations), whereas pseudowords should produce such transformations more often. Such a difference is expected if words do in fact bind their constituent phonemes more tightly than pseudowords do. This prediction, in conjunction with the results of Mann and Repp (1981) that suggest grouping affects compensation, raises the question as to whether fricative binding may have influenced the findings of Pitt and McQueen (1998).

Experiment 1 is a VTE experiment that examines how well final fricatives are bound in lexical and nonlexical monosyllables, using the stimuli of Pitt and McQueen. The basic utterances were the eight syllables created by the crossing of three binary factors. One factor was whether a word could be formed from the CV base by appending /s/ or /ʃ/: The initial CVs in "juice" and in "bush" have this property, whereas those in /dɚs/ and /neIʃ/ do not. The second factor was whether the concatenated fricative was /s/ or was /ʃ/. Finally, for each pair of syllables, one could potentially exert a stronger hold on its fricative than the other, either through lexical factors ("juice" > /dʒuʃ/; "bush" > /bUs/), or through transitional probability (/dɚs/ > /dɚʃ/; /neIʃ/ > /neIs/).

If final fricatives bind more strongly to lexical than to nonlexical contexts, then there should be fewer grouping transformations with words than with pseudowords. Because such binding effects may well also be depend on transitional probability, strings with high transitional probabilities may yield fewer grouping transformations than strings with low transitional probabilities.

Method

Participants

Thirty two native English speakers participated in Experiment 1. None reported any hearing deficiencies. They received course credit or a small payment for participation.

Stimuli

Digital copies of the eight syllables used as context stimuli by Pitt and McQueen (1998, Experiment 2) were used. Each syllable consisted of a naturally produced initial CV, followed by one of two fricatives (/s/ or /ʃ/) that were created using a version of the Klatt (1980) synthesizer. Synthesis parameters are listed in the Appendix of Pitt and McQueen (1998). Each fricative was approximately 170 ms long, and the durations of the CV portions were 247 ms (/bUʃ/, /bUs/), 178 ms (/dɚʃ/, /dɚs/), 264 ms (/dʒus/, /dʒuʃ/), and 317 ms (/neIs/, /neIʃ/).

Eight VTE sequences were constructed, one for each syllable. Each sequence included 250 iterations of the syllable, with a 100 ms silent period between each iteration. Each sequence lasted approximately 2 min.

Apparatus and procedure

The eight syllables were stored on the disk of a Pentium III 450 computer. For presentation, they were output through a 12-bit D/A converter (10 kHz sample rate), low-pass filtered (4.8 kHz), amplified, and presented binaurally over high quality headphones. These conditions matched those used by Pitt and McQueen (1998).

Each participant listened to four experimental VTE sequences preceded by two practice VTE sequences. For half of the participants, the four sequences were based on /bUʃ/, /dɚs/, /dʒuʃ/, and /neIs/. For the other participants, the four sequences included /bUs/, /dɚʃ/, /dʒus/, and /neIʃ/. The practice sequences for all participants were based on the word "sleet" and the nonword /tʌls/, with half of the participants receiving them in each of the two possible orders. The order of the experimental sequences was counterbalanced across participants in a Latin Square design. Participants were tested in groups of 1–3 listeners in a sound attenuated chamber.

Participants were given a small stack of legal size paper, with each sheet providing two columns of 38 answer blocks. They were told that they would be hearing speech over their headphones, and that their task was to report what they heard. They were told to write down the first thing that they heard when a sequence began, and to then write down any subsequent changes. They were to start at the top of a column and to use as many lines on the answer sheet as they needed, using a new line for each change. The experimenter interacted with the participants after each of the practice sequences to make sure that the task was clear.

Results and discussion

Scoring

Each participant's responses to the four VTE sequences were scored in terms of the number of perceived changes, with subscores computed for the number of grouping-based transformations using some of the criteria of Pitt and Shoaf (2002). Reports were considered grouping transformations if they exhibited one of the following properties: (1) there was no fricative (e.g., /bUʃ/ being reported as /bU/, or as /bUk/); (2) the fricative moved from its syllable-initial position, the position required for compensation (e.g., /neIs/ being reported as /sneI/, or as /sneIk/); (3) the fricative was extended into an additional syllable, again taking it out of final position (e.g., /dʒus/ being reported as /tʃudʒu/). Note that some cases may involve more than one type of fricative change. The most common type of change by far was the first kind. There were some of the second type, and only a handful of the third.

For each participant, the total number of changes involving the fricative (grouping transformations) was tallied for each of the four VTE sequences. These scores were entered into a three-factor analysis of variance. The only between-subject factor was Group (one Group heard /bUʃ/, /dɚs/, /dʒuʃ/, and /neIs/; the other Group heard /bUs/, /dɚʃ/, /dʒus/, and /neIʃ/). One within-subject factor was whether a given item should have greater perceptual stability due to either lexicality (/bUʃ/, /dʒus/), or transitional probability (/neIʃ/, /dɚs/), or whether it should be perceptually less stable (/bUs/, /dʒuʃ/; /neIs/, /dɚʃ/). The third factor was also within-subject, and coded whether the syllables were based on words (/dʒus/, /dʒuʃ/, /bUʃ/, /bUs/), or pseudowords (/neIs/, /neIʃ/, /dɚs/, /dɚʃ/).

Fig. 1 presents the average number of verbal transformations involving a fricative change, broken down by whether a syllable was word-based or not, and whether the syllable was favored or disfavored by either lexicality or transitional probability. The results were very orderly. There was no effect of Group, or any interaction of Group with the other factors (largest $F(1,30) = 2.45$, n.s.). In contrast, both other main effects were quite robust. Pseudoword-based VTE sequences produced 66% more fricative-altering transformations (16.6 versus 10.0) than ones based on words, $F(1,30) = 12.19$, $p < .01$. Syllables with either lexical or transitional probability support for their fricatives produced significantly fewer fricative-altering transformations, $F(1,30) = 7.79$, $p < .01$. These two effects were additive, with no interaction whatsoever, $F(1,30) < 1$.

The results of Experiment 1 indicate that real words ("bush," "juice") exert a strong hold on their final fricatives, leading to very low verbal transformation rates (mean = 8.3 cases of final fricative loss). At the other extreme, pseudowords with low transitional probabilities from the vowel into the final fricative (/neIs/, /dɚʃ/) produce very weak binding of that fricative (mean = 18.1 cases of final fricative loss). If, as Mann and Repp's (1981) data suggest, compensation effects depend on the perceptual grouping of the fricative with the following stop consonant, then the differences revealed in Experiment 1 could certainly have affected the pattern of compensation effects in Pitt and McQueen's (1998) experiments. In particular, if the fricatives in the words were perceptually bound within them, then based on Mann and Repp's results, we would expect weak compensation effects. Conversely, if the fricatives in the pseudowords were poorly bound (especially those with low transitional probabilities), then the fricatives should have been available to produce compensation effects. Note that this analysis does not in any way negate the possible importance of transitional probability in producing compensation. However, it does suggest that the difference in compensation effects found by Pitt and McQueen might reflect a difference in perceptual grouping, rather than a lack of lexical influence on compensation.

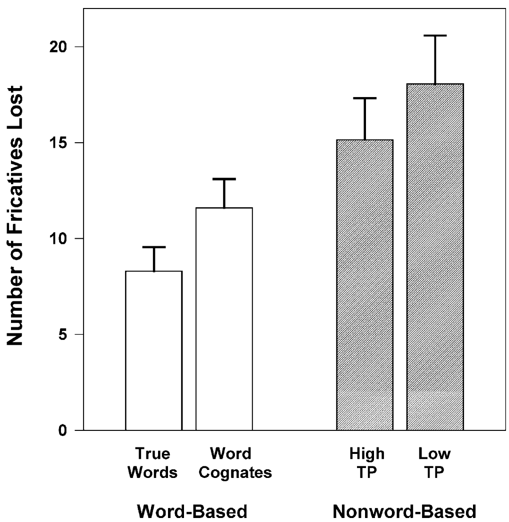

Fig. 1. The average number of verbal transformations involving a fricative change observed in Experiment 1 (TP = transitional probability). The True Words were /bUʃ/ and /dʒus/, the Word Cognates were /bUs/ and /dʒuʃ/, the High TP items were /neIʃ/ and /dɚs/, and the Low TP items were /neIs/ and /dɚʃ/.

In Part II, we revisit the question of lexically mediated compensation. In these experiments, we examine whether longer words (as in Elman & McClelland's (1988) study) may be needed to provide sufficient lexical activation. We couple this with time-based attempts to control for the perceptual grouping problem identified in Experiment 1.

Part II: Compensation for coarticulation, with temporal control of perceptual grouping

Preliminaries to testing compensation: Stimulus selection and construction

To test lexically mediated compensation for coarticulation, three types of stimuli are required. First, one or more pairs of lexical contexts are needed, with members of a pair contrasting in their final fricatives (e.g., "Christmas"–"foolish," or "juice"–"bush"). Second, a fricative continuum, ranging from /s/ to /ʃ/ is needed. Finally, a test series with initial stop consonants that vary in place of articulation must be constructed (e.g., "tapes"–"capes"). Each of these sets of stimuli must meet certain criteria in order to provide an appropriate test of compensation. In this section, we describe our choices of stimuli and indicate how they meet the necessary criteria. All stimuli were derived from recordings produced by the second author in a sound-attenuated room. Stimuli were recorded onto DAT tape and then digitally transferred to a PC, where they were stored as 16 bit, 16 kHz sound files.

Fricative continuum

An /s/-/ʃ/ continuum was created by blending, sample by sample, two equal-duration (249 ms) tokens of each fricative. Each was spoken in isolation, and then shortened to this duration by truncating the segment to 249 ms and then adjusting its amplitude contour. Pilot testing determined that an 8-step series beginning at a blending ratio of .8/.2 (/s/ versus /ʃ/) and ending at a ratio of .4/.6 (.05 step size) yielded satisfactory endpoints and a perceptual boundary near the middle of the continuum.

Stop-consonant test continuum

A /t/-/k/ test series was made following the same procedure as Pitt and McQueen (1998). The words "tame" and "came" were used instead of "tapes" and "capes" to avoid any possible effects of having the fricative /s/ present in the test words. Several tokens of each word were recorded, and two with similar stop durations (/t/ = 94 ms, /k/ = 106 ms) were used to create the continuum. The initial aspiration (which in the case of /t/ began with 12 ms of silence to equate its duration with /k/) plus the first 2–3 pitch periods of tokens of /teI/ and /keI/ were blended in varying ratios as described above (.05 step size), and /eIm/ from "tame" was then appended to each step. This continuum was narrowed down to six steps in which the /k/ endpoint had a blend ratio of .5/.5 (/t/ versus /k/) and the /t/ endpoint had a ratio of .75/.25. All steps were 503 ms in duration.

Before conducting any tests of lexically mediated compensation, it is important to demonstrate that the fricative and stop stimuli that we constructed are capable of producing a compensation effect under conditions similar to those of Mann and Repp (1981). We therefore conducted a pilot experiment to test the adequacy of these stimuli. The endpoint /s/, the endpoint /ʃ/, and one ambiguous fricative (step 5) were played before each of the steps of the /teIm/-/keIm/ series (0 ms ISI). Twelve listeners labeled every step 20 times in each of the three fricative contexts, with the order of the 18 cases (3 fricatives × 6 /t/-/k/ steps) randomized each time. The average proportion of /k/ responses for each context is shown in Fig. 2. The large shift in the /s/ and /ʃ/ functions is clear evidence of compensation: /k/ is reported more often when preceded by /s/ than when preceded by /ʃ/, $t(11) = 4.42$, $p < .001$ (see Experiment 2 for details of our statistical analysis procedure). The function for the ambiguous fricative falls in between the two endpoints, just as Mann and Repp (1981, Experiment 5) found, and is reliably different from the endpoints ($t(11) = 5.27$, $p < .001$ and $t(11) = 2.55$, $p < .02$, for the /s/ and /ʃ/ comparisons, respectively). These data clearly demonstrate that our fricative and stop continua can produce robust compensation for coarticulation.

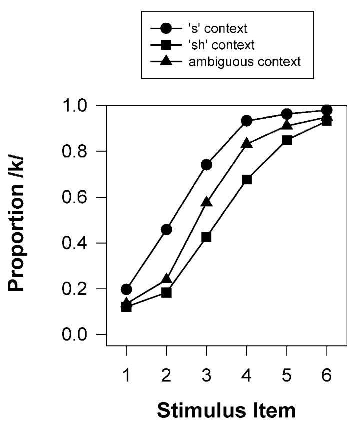

Fig. 2. Average report of /k/ for the six steps of the "tame–came" continuum. Data are shown separately for stop identification in the context of /s/, /ʃ/, and an ambiguous fricative.

The lexical context stimuli

As described in the Introduction, a central concern of the present study was whether the monosyllabic words used by Pitt and McQueen (1998) might have produced insufficient lexical activation to drive the compensation effect. Therefore, we selected four pairs of longer words that had the desired contrast in their final fricatives: "abolish"–"arthritis," "establish"–"malpractice," "distinguish"–"consensus," and "extinguish"–"contagious." The first two pairs differed from the second two in the size of their word-initial cohorts. The first four words had smaller cohorts and became lexically unique earlier in the word than the second four words. Because no differences emerged between these two types of words, this manipulation will not be discussed further. For comparison purposes, we selected two monosyllabic pairs ("wish"–"miss," and "fish"–"kiss"); we later added one bisyllabic pair as well ("punish"–"promise"). Each word was spoken in isolation with a final /θ/ (e.g., /kIθ/"kith"), which was then spliced off just prior to /θ/ onset, leaving word "bases" that lacked a final fricative. The word-initial cohorts of the one- and two-syllable words resembled those of the three-syllable words.

Our belief that longer words would provide stronger lexical activation was based on both theoretical and empirical grounds. If a word's activation level varies with the amount of bottom-up support for it in the input, then longer words have an inherent advantage over shorter words, because they (by definition) can receive support from more phonemes than shorter words. A related potential advantage is that longer words tend to have fewer near neighbors than shorter words, which can again provide an activation advantage for the longer words, due to their being subject to less competition. Studies of the phonemic restoration effect (Samuel, 1981, 1996) have demonstrated this difference, with three- and four-syllable words consistently providing stronger restoration of their constituent phonemes than two-syllable words. Moreover, recent experiments (Pitt & Samuel, 2002) with the longer words and the monosyllables used in the current study gave us good reason to believe that the longer words would generate strong lexical activation, significantly stronger than the monosyllables. These experiments explored lexically mediated shifts in the identification of the words' final fricatives (the "Ganong effect"; Ganong, 1980), and showed that the long words produced lexical shifts that were twice as large as those for the short words. In fact, even when the durations of the three-syllable and the monosyllabic words were equalized using speech compression/expansion, the lexical effect for the three-syllable words was still twice as large.

The word pairs were selected to prevent any possible influence of neighboring transitional probabilities on fricative perception by holding the preceding vowel constant: For all of the words, the final fricative was preceded by the vowel /I/ (the words were selected from a list generated by searching an online phonetic dictionary (cmudict4.0) for words ending in /Is/ or /Iʃ/). Therefore, if any compensation effects were to be observed as a function of lexical context, these effects could be unambiguously attributed to the lexical context, rather than to any confound of transitional probability that variation in the vowel might cause. As Shillcock et al. (1992) pointed out, the actual vowel realization for many words with /I/ in their final syllable can be quite variable. We produced our context words in citation form, to be sure that the vowel was in fact /I/. To check that the final vowels in all words were indeed matched, and were all instances of /I/, we measured the frequencies of F1 and F2 at the peak amplitude of each word's final vowel. We compared these values to the F1 and F2 values for a set of vowels chosen to represent our talker's vowel space. Fig. 3 presents these measurements, and shows clearly that all of the vowels were quite similar to each other, and were clustered around /I/. Importantly, there is no systematic displacement of the vowels as a function of the following fricative: The vowels in words ending in /s/ (shown with open symbols) did not differ from the vowels in words ending in /ʃ/ (shown in filled symbols). Thus, these stimuli satisfy the criteria necessary for testing whether compensation for coarticulation can be lexically mediated.

Experiment 2

As noted above, we undertook this project because there was reason to believe that longer context words were needed to produce enough lexical activation for lexically mediated compensation to be observed. In running our initial experiments for this project, we also began to appreciate the role that perceptual grouping might be playing in the observed results. This led us to the test of perceptual grouping using the VTE approach

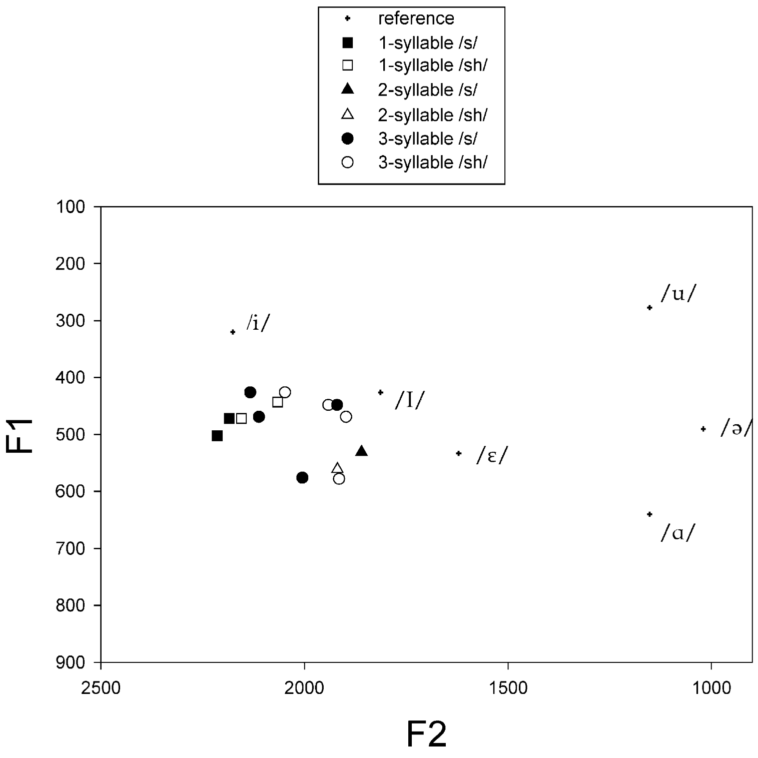

Fig. 3. F1–F2 vowel space plot for the final vowels in the 7 context pairs used in the current study, along with six reference vowels that were produced in the second syllable of the frame /pros_/. The first syllable received primary stress and the second was unstressed.

(Experiment 1) and to try various methods of controlling the extent to which the final fricative perceptually groups with its lexical context. For clarity of exposition, we will first present Experiment 2, which included several conditions that were run after settling on a time-based manipulation of perceptual grouping. Then, in the following section, we will summarize a number of the earlier experiments, to clarify the role of perceptual grouping in the compensation paradigm, and to provide converging evidence for the role of lexical activation in producing compensation.

Method

Stimuli

The stimuli used in these experiments were based on the three components described above: (1) the eight step /s/-/ʃ/ continuum; (2) the six-step /t/-/k/ (/teIm/-/keIm/) continuum; and (3) the seven different context pairs (see below).

Procedure

As a practical matter, we split the data collection between our two laboratories, with each lab testing both longer words (two word-pairs each), and monosyllables (one pair). When we added the bisyllabic pair, data were collected in both laboratories. The "arthritis–abolish" and "extinguish–contagious" pairs were tested at Stony Brook, along with the monosyllabic "wish–miss" pair. At Ohio State, "distinguish–consensus" and "establish–malpractice" were the focus, along with "fish–kiss."

Each participant received context stimuli based on a single word-pair, in an experimental session that took about 45 min. The experimental session consisted of two parts. The first was a short pretest designed to identify the step on the fricative continuum that was most ambiguous for the individual being tested. In the pretest, 16 stimuli were used, created by crossing the two relevant word pair bases (e.g., "aboli" and "arthriti") with the eight members of the /s/-/ʃ/ continuum. The participant categorized the final fricative of each resulting stimulus as either /s/ or /ʃ/, by pushing appropriately labeled buttons on a response board. Twelve randomizations of the 16 stimuli were presented. Each stimulus was presented one second after the participant's response to the previous stimulus. Mean proportion /s/ was calculated for each step on each continuum. The most ambiguous step was defined as the step at which the lexical context had its largest influence on fricative labeling, increasing /s/ responses the most following the /s/-biased context (e.g., "arthriti") and decreasing /s/ responses the most following the /ʃ/-biased context (e.g., "aboli"). In practice, this amounted to choosing the step with the largest difference score in fricative labeling between the two contexts.

The main compensation task followed the pretest. Each trial of the compensation task consisted of a context item followed by a target item. The *target* item was one of the six members of the "tame"–"came" continuum. For each participant, there were six different *context* items used. These six items were the result of crossing the two relevant bases (e.g, "aboli" and "arthriti") with three fricatives: the endpoint /s/, the endpoint /ʃ/, and the participant's individually determined neutral fricative. This procedure creates the three main conditions of Experiment 2: Compensation can be assessed with words in which the fricative is consistent with the context (e.g., "arthritis"), pseudowords in which the fricative is inconsistent with the context (e.g., "aboliss"), and with the critical items in which the fricative is ambiguous (e.g., "arthriti?").

Listeners were instructed to perform two tasks after hearing each context–target utterance. The first task was to identify the second item (the target) as either "tame" or "came" by pushing one of two labeled response buttons. Immediately after this response was made, the string "Word?" was printed in the center of a display in front of the participant. This served as a cue to initiate the second response, which was to decide whether the first word (the context) was a real English word or not (e.g., "abolish" is, but "aboliss" is not). Two additional buttons, one labeled "yes," the other labeled "no," were used to make this response. The purpose of the second task was to force the participants to listen to the context word. Pitt and McQueen (1998) used a similar procedure (see also Mann & Repp, 1981), requiring listeners to report the identity of both the final fricative and the initial stop. In pilot tests, we experimented with both of the tasks and found that listeners were more comfortable making the word judgment. For current purposes, the details of the second task are not critical, as neither such task can affect the phonetic processing of the following stop, in Merge (Norris et al., 2000) or any other autonomous model.

A maximum of 5000 ms was allowed for each response. After the second response, there was a 1500 ms pause, followed by the start of the next trial.

Experiment 1 demonstrated that perceptual grouping of a final fricative varies with both lexical status and transitional probability. Moreover, Mann and Repp (1981) showed that compensation effects depend on the perceptual grouping of the fricative; compensation was reduced when a vowel preceded the fricative, apparently making the fricative less available to affect the perception of the following stop consonant. We therefore tried to titrate the perceptual grouping of the fricative, trying to satisfy two somewhat contradictory goals: The ambiguous fricative must group sufficiently with the preceding lexical context to be lexically influenced, but this association must not be so strong as to prevent the fricative from interacting with the following stop to produce compensation.

There were two timing factors available to us to try to strike this balance. One factor was the interval between the fricative and the following stop, which we will call the interstimulus interval (ISI). In Pitt and McQueen's (1998) experiments, this ranged from 20 to 40 ms; Elman and McClelland's stimuli had an ISI of 0 ms. Presumably, longer ISIs should tend to reduce the influence of the fricative on the stop (Mann and Repp found some evidence for this). The second timing parameter available was the time between the end of the lexical base, and the onset of the final fricative; we call this the gap. By introducing a gap before the fricative, we can increase its likelihood of grouping with the following stop.

We experimented with various combinations of gap and ISI, as well as some additional factors. The results of this pilot work will be discussed in a later section. For Experiment 2, an ISI of 0 ms was selected for all of the three-syllable word pairs. This was coupled with a relatively long gap size of 40 ms for these items. This combination was intended to push the fricative's grouping more toward the following stop consonant, given the very strong lexical influences these long words produce. The situation for the shorter words is less clear, and may vary considerably with idiosyncratic properties of the individual word pairs. We used a long gap and short ISI (0 ms) for the "wish–miss" (gap = 40 ms) and "punish–promise" (gap = 30 ms) cases run at Stony Brook. We used the opposite grouping strategy for the "fish–kiss" and "punish–promise" cases at Ohio State (0 ms gap, 30 ms ISI).[1] The latter two cases also differed in one other way: Based on pilot testing, the ambiguous fricative was always step 5 of the /s/-/ʃ/ continuum, rather than being set individually.

Apparatus

All stimuli were originally stored in digital form on the disk of a Pentium PC (16-bit, 16 kHz sample rate). Stimuli used at Stony Brook were scaled to 12-bit form. For presentation to participants, the stimuli were output via a digital to analog converter (16 kHz rate), low-pass filtered (7.8 kHz), amplified, and presented binaurally over high quality stereo headphones. Each participant was tested individually in a sound-shielded booth. Responses were entered on a 4-button response board with appropriate labels on each button. The experiment was controlled by a PC, which stored the responses for later analysis.

Participants

A total of 88 listeners participated. All were native English speakers with no known hearing problems. They received course credit or a small payment for their participation. At Stony Brook, 24 participants were tested on the three-syllable stimuli (12 on "abolish–arthritis," and 12 on "extinguish–contagious"), 12 were tested on the monosyllables, and 10 were tested on the bisyllabic stimuli. At Ohio State, 17 participants were tested on the three-syllable stimuli (9 on "distinguish–consensus," and 8 on "establish–malpractice"), 15 on the monosyllables, and 10 on the bisyllabic stimuli.

Results and discussion

The central question of the current study is whether lexically mediated compensation for coarticulation can be observed under the appropriate conditions. In Experiment 2, these conditions involved longer words than Pitt and McQueen (1998) used, in order to increase the putative lexical influences on the final fricative. To balance the anticipated corresponding increase in perceptual grouping of the fricative with the word containing it, we manipulated the timing parameters of stimulus presentation.

We demonstrated in the pretest (Fig. 2) that our fricatives were capable of producing appropriate compensation shifts on the labeling of the "tame–came" stimuli, when the fricatives were presented in isolation with an ISI of 0 ms. Before examining compensation with lexically determined ambiguous fricatives, it is also necessary to ascertain that compensation occurs using the fricatives presented in word contexts, with the gaps and ISIs used here. For example, if a true (endpoint) /s/ occurring in its appropriate lexical context (e.g., "contagious") does not produce compensation, then it makes no sense to look for such effects with an ambiguous fricative. Using the very conservative criterion that a majority of the subjects show appropriate effects in the intact, consistent word condition, 7 of our 8 cases allow us to look for the effect of interest. One pair, however, does not: For the "establish–malpractice" pair, a majority of the subjects produced reversals in the consistent word condition (and in the ambiguous fricative condition as well). The mix of 5 reversals with 3 in the predicted direction led to a completely null effect for the real word condition, $t(7) = 0.52$, n.s. The reversals in the ambiguous case were actually sufficient to reach significance, $t(7) = -2.43$, $p < .05$. Only the pseudoword case behaved as one would expect, $t(7) = 2.49$, $p < .04$. We can think of no obvious reason for this odd outcome, as other experiments using the same continuum did not yield such anomalous results (see below).

Fig. 4 presents the results for the long context words, averaged across the "arthritis–abolish," "contagious–extinguish," and "consensus–distinguish" pairs. Each

[1] As noted, other combinations were explored in pilot work that will be described in a later section. The values that we used in Experiment 2 were the ones that seemed to provide the best balance between the need for a strong lexical influence on the fricative and the need to prevent excessive perceptual grouping of that fricative with its carrier word.

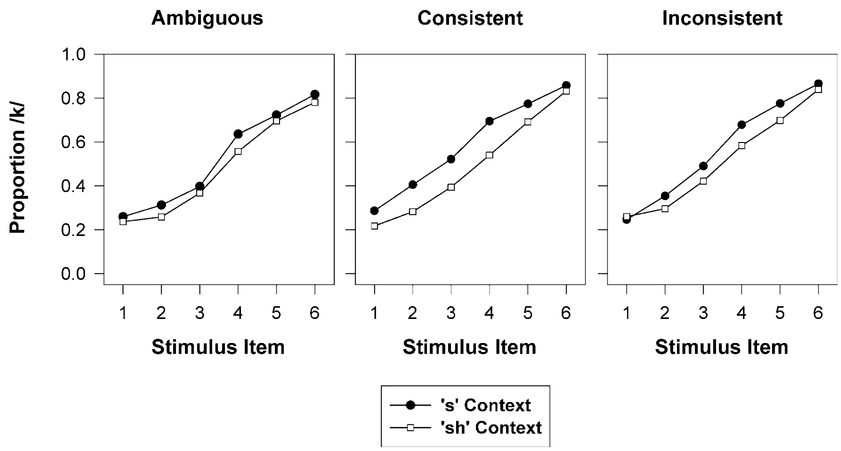

Fig. 4. Identification of the "tame–came" test words (proportion /k/), as a function of whether the preceding fricative was /s/ or /ʃ/ (determined either acoustically, and/or via lexical context). The /ʃ/ context words were "abolish," "distinguish," and "extinguish," and the /s/ context words were "arthritis," "consensus" and "contagious." The left panel shows the results for an acoustically ambiguous fricative, the middle panel shows the results for fricatives occurring in correctly pronounced words, and the right panel shows the results when the fricatives were placed in inconsistent lexical contexts.

panel shows the average identification data for the "tame–came" test words, in the context of either /s/ or /ʃ/. For the center panel, the fricative occurred in its normal lexical context (e.g., the /s/ in "arthritis"). In the right panel, these fricatives were defined entirely acoustically, because they were presented in inappropriate lexical contexts (e.g., /s/ in "aboliss"). The left panel presents the critical data: The fricatives were acoustically identical for the two contexts (i.e., curves), and only acquired their phonetic identity by virtue of their lexical context.

The data were analyzed using the same procedures as in previous studies of this sort (e.g., Pitt & McQueen, 1998; Pitt & Samuel, 1993; Samuel, 1986). For each subject, the average percentage of stimuli identified as /k/ was computed for trials in which the context was (acoustically and/or lexically) /s/, and when it was /ʃ/. These percentages were based on the four middle items of the /t/-/k/ test series. Each subject's two scores in each condition were used to compute t tests for matched samples.[2]

In the word-context condition (i.e., /s/ and /ʃ/ presented in their natural lexical contexts), shown in the center panel of Fig. 4, compensation is relatively robust, with fewer reports of /k/ following /ʃ/-biased words (e.g., "abolish") than following /s/-biased words (e.g., "arthritis"). Similarly, when the context was essentially pseudowords (right panel), we find the expected shifting of the labeling functions. The critical condition is of course the one in which the fricative is acoustically identical in the two cases (left panel). As the panel shows, lexical activation was sufficient to cause the perception of the appropriate fricative, which in turn produced the compensation effect on "tame–came" identification.

Looking first at the data for each word pair individually, the results are a bit variable due to the limited number of subjects tested on each pair, but they generally provide clear evidence of lexically induced compensation. For "extinguish–contagious," compensation was reliable for the true word condition, $t(11) = 3.98$, $p < .01$, but not for the pseudoword contexts, $t(11) = 0.06$, n.s. Critically, for the ambiguous fricative condition, the compensation effect was significant, $t(11) = 2.49$, $p < .04$, with 8 of the 10 subjects (plus two ties) shifting in the predicted direction. For "distinguish–consensus" compensation was reliable for the true word condition, $t(8) = 2.63$, $p < .03$, and marginally so for the pseudoword case, $t(8) = 2.17$, $p < .06$. For the critical ambiguous fricatives, the effect was reliable, $t(8) = 2.37$, $p < .05$, with 7 of the 9 subjects showing the effect.

The effects for "abolish–arthritis" were generally somewhat noisier: The shift for the true word context did not reach significance, $t(11) = 0.97$, n.s., though the

[2] We have used the procedure of calculating effects using the middle four items on the /t/-/k/ continuum, for continuity with previous work (e.g., Pitt & McQueen, 1998). This procedure is based on the general rule that effects are strongest in the most ambiguous portion of the continuum. With the /t/-/k/ tokens used in the current study, effects were actually spread across all six (intentionally ambiguous) items. The extra observations in some cases provide a little more stability for each observation. In general, the results of the two approaches are quite similar, so we have opted throughout the paper to maintain consistency with Pitt and McQueen, and to therefore report statistics based on the middle four items.

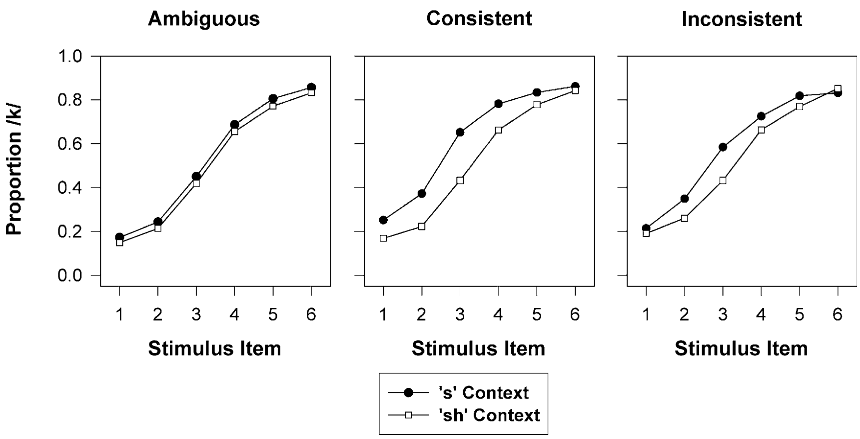

Fig. 5. Identification of the "tame–came" test words (proportion /k/), as a function of whether the preceding fricative was /s/ or /ʃ/ (determined either acoustically, and/or via lexical context). The /ʃ/ context words were "fish," "wish," and "punish," and the /s/ context words were "kiss," "miss," and "promise." The left panel shows the results for an acoustically ambiguous fricative, the middle panel shows the results for fricatives occurring in correctly pronounced words, and the right panel shows the results when the fricatives were placed in inconsistent lexical contexts.

one for the pseudoword context did, $t(11) = 2.67$, $p < .04$. The critical ambiguous fricative context behaved very much like the true word case, with the shift failing to reach significance ($t(11) = 0.74$, n.s.), despite 9 of the 12 subjects shifting in the predicted direction.

Taken together, the "arthritis–abolish," "contagious–extinguish," and "consensus–distinguish" stimuli produced impressively consistent shifts in the critical ambiguous fricative condition: 24 subjects produced compensation shifts in the predicted direction, versus only 7 reversals (by a sign test, $p < .01$).[3]

As expected, lexically mediated compensation was more difficult to find with the shorter words. Recall that there were three word pairs tested: monosyllabic "miss–wish" and "kiss–fish," and bisyllabic "promise–punish"; the latter pair was tested under two different sets of timing conditions. The data, combined across these four tests, are shown in Fig. 5.

Of these four tests, only one yielded reliable lexically mediated compensation, "kiss–fish" (word: $t(14) = 4.93$, $p < .001$; pseudoword: $t(14) = 4.73$, $p < .001$, and ambiguous: $t(14) = 2.14$, $p = .05$). The small shift shown in the left panel of Fig. 5 is entirely due to the "kiss–fish" results, as "miss–wish" produced a marginal reversal, and the two versions of "promise–punish" each produced no change at all. Of the 46 subjects who were tested across the four cases, 23 showed effects in the lexically mediated direction (with one tie).

For "miss–wish," both the true word case ($t(11) = 4.90$, $p < .001$) and the pseudoword context ($t(11) = 4.11$, $p < .002$) produced very robust compensation effects. The results for the two "promise–punish" tests were quite similar to each other for the word and the ambiguous contexts, but differed in the pseudoword case. For the true word contexts, both tests produced reliable shifts, $t(9) = 3.52$, $p < .006$ for one, and $t(9) = 3.01$, $p < .02$ for the second. For the pseudoword contexts, one test produced the predicted shift, $t(9) = 3.56$, $p < .006$, but the other yielded a reversal, $t(9) = -3.34$, $p < .01$. The results for the ambiguous condition were virtually identical in the two tests, with neither showing any shift at all, $t(9) = 0.64$, n.s.; $t(9) = 0.00$, n.s.

The results of Experiment 2 demonstrate that it is indeed possible to find lexically mediated compensation for coarticulation under conditions in which the transitional probability of the vowel–fricative pair is perfectly controlled. These results are consistent with interactive models of word recognition, in which lexical activation affects the perception of sublexical representations. They are inconsistent with autonomous models, even models such as Merge in which lexical information can affect phonetic judgments, but not phonetic perception. The key to teasing apart these cases is the use of a task in which lexical influences are inferred through their indirect impact on phonetic

[3] We would have liked to report a t test of the data combined over the three word pairs, in addition to this sign test, but we cannot because the data files for "distinguish–consensus" were lost due to experimenter error, after the initial analyses had been performed. In its place, we performed a meta-analysis using the method of adding weighted Z's (Rosenthal, 1991). The resulting p-value was .006, reinforcing the results of the individual analyses.

processing. Thus, the current results, together with other results using such indirect procedures (Samuel, 1997, 2001), indicate that an accurate model of auditory word recognition must incorporate top–down lexical-phonetic connections.

The rocky path to lexically mediated compensation

We noted above that the successful testing conditions used in Experiment 2 were developed as a result of extensive pilot work. In addition, one of the four longer word pairs ("establish–malpractice"), and most of the shorter word pairs failed to produce lexically mediated compensation, despite the use of several different choices of timing and stimulus parameters. In this section, we will briefly review the path that brought us to the generally successful conditions used in Experiment 2. This review serves two purposes. First, these efforts provide additional evidence of the important role that perceptual grouping plays in this paradigm. Second, despite the absence of clear lexically mediated compensation in these pilot experiments, there are patterns in the data that only make sense if such compensation was actually occurring for some of the participants, some of the time.

Finding lexically mediated compensation requires a rather daunting constellation of events to co-occur within an experiment. First, the fricative context continuum must be good enough to produce reasonable percepts of the two endpoints, while also having enough of an ambiguous range to find a perceptually malleable token. The same constraints apply to the initial stop consonant (/t/-/k/) series as well. Furthermore, lexical influences need to be strong enough to produce a robust shift in the perception of an ambiguous fricative. Even if lexical influences are found in the fricative identification task, they may not show up in the compensation task, simply because listeners may focus their attention on the required /t/-/k/ judgment, rather than on the preceding context. As we discovered in our pilot work, in addition to all of these constraints, the interaction of perceptual grouping effects with lexical activation can be particularly problematic: As lexical activation is increased (e.g., by using longer words), the final fricative becomes more tightly attached to its carrier word, thereby reducing any compensation effect.

The pilot work that we conducted involved a series of experiments that varied in a number of ways, as we discovered the constraints described above and tried to adapt to them. Our approach was to select a number of stimulus and task parameters, and to run about a dozen listeners through the compensation task. Over the course of this pilot work, we varied the range of mixtures making up each continuum, and in one case, whether the fricative continuum was made with mixtures of naturally produced sounds or synthetic ones.[4] We varied the nature of the second task done during compensation: In addition to the lexical decision task that we settled on, we tried a fricative judgment task (/s/ versus /ʃ/), or no second task at all. And, as described above, we tried a number of different combinations of gap duration and ISI.

Although our primary focus in the pilot work was to establish the conditions that would allow compensation effects to be observed for the ambiguous fricative contexts, we also examined the results for the real word (endpoint fricatives consistent with their word contexts) and pseudoword (endpoint fricatives inconsistent with their word contexts) conditions. As we collected a substantial amount of data, we noticed two interesting patterns. First, in most cases, we were finding larger compensation effects for the pseudoword condition than for the real word condition. Second, the size of the compensation effect in the ambiguous context condition appeared to correlate with the size of this effect for the real word case. Table 1 summarizes the observed shifts and the correlations of the size of these shifts for all of the data that were collected (including the data reported in Experiment 2).

Consider first the data shown in the last two columns. These are the average compensation shifts for the word and for the pseudoword conditions. For the longer words, the average compensation effect for the pseudoword case (12%) was three times larger than the corresponding effect for the word case (4%), $t(140) = 4.7$, $p < .001$. At first, this seemed quite odd—why should pseudowords produce a larger effect than words (which should have lexical activation helping their fricatives)? However, once we began to think in terms of perceptual grouping, this pattern made sense. For both the words and the pseudowords, the final fricatives were very good phonetic tokens, which meant that lexical activation could not change their perception substantially. However, such activation *could* bind the fricative to its *word* context, making the fricative less able to participate in compensation with the following stop consonant. Even the pseudoword contexts could produce some of this binding: The 12%

[4] One pilot experiment used Pitt and McQueen's "juice" and "bush" context words, but with the natural speech fricative continuum that we used in our experiments, rather than the synthetic fricatives they used. A group of 14 listeners did the compensation task with both the ambiguous context stimuli, and ones in which the fricatives were in real words (comparable to the left and center panels in Figs. 4 and 5). These listeners showed no compensation in either condition. Recall that Pitt and McQueen found no compensation for the ambiguous word case, but obtained robust shifts for real fricatives in real words. The disappearance of the effect with natural speech fricatives, like those used in the current study, underscores the fragility of compensation effects.

Table 1
Correlation between the size of the compensation effect in the ambiguous condition and the word and pseudoword conditions across word contexts

Context word pair	Ambiguous versus				N	Mean size of shift in the labeling of the fricative (%)		
	Word		Pseudoword					
	r	p	r	p		Ambiguous	Word	Pseudoword
Three syllable								
Abolish–arthritis	.24	.06	−.15	.02	32	.00	.01	.08
Extinguish–contagious	.30	.09	−.31	.10	12	.06	.17	.00
Establish–malpractice	.30	.02	−.34	.02	62	−.05	.01	.14
Distinguish–consensus	.30	.08	−.31	.08	35	.02	.09	.17
Two syllable								
Punish–promise	.48	.01	.09	.66	27	.02	.12	.12
One syllable								
Fish–kiss/wish–miss	.46	.001	−.08	.55	55	.02	.23	.22

effect they produced was only half as large as the 24% effect that the bare fricatives caused (shown in Fig. 2). This is similar to, but larger than, the grouping effect that Mann and Repp (1981) reported when they preceded their fricative with a single vowel. For the long words in our experiments, this effect was pulled all the way down to 4%, due to perceptual grouping.

The shorter words in our study provide a very interesting comparison, in two ways. First, the size of the compensation effect overall is clearly dependent on word length, with average shifts of 23%, 12%, and 4% for words of one, two, or three syllables, respectively, $F(2,219) = 27.90$, $p < .001$. Second, the difference between the word and pseudoword case that was so prominent with longer words is nonexistent for the shorter words. Both of these results are in accord with the view that longer words provide more lexical activation, which in turn increases the binding of the final fricative to the word.

The second striking result in Table 1 is the consistently good positive correlation across subjects in the size of the compensation effect in the word context with its size in the ambiguous context condition. In phonemic categorization experiments, listeners differ greatly in the extent to which lexical memory affects labeling (e.g., see Massaro & Oden, 1995). The present study was no exception, with large compensation effects (44%) and substantial reversals (−17%) being found across listeners for all word pairs. Such wide variation in lexical influences suggests that compensation effect sizes would covary in the two conditions in which lexical effects should be observed: the word condition and the ambiguous condition. Those subjects who (for whatever reason) exhibit strong lexical influences would show larger effects of compensation in both conditions. This correlation was actually our first clue that lexically driven compensation might in fact be occurring, without being visible in the overall labeling data.

Moreover, such individual variation should cause a negative correlation in compensation effect size between the ambiguous condition and the pseudoword condition. This is because the lexical information is *inconsistent* with the acoustic information for the fricatives in the pseudowords. The stronger the lexical effect for a given listener, the less of an effect the conflicting fricative should have on labeling the following stop consonant. In the extreme, the lexically consistent fricative would be restored and listeners would show a reverse compensation effect. As Table 1 shows, such negative correlations were consistently observed for the longer words. The absence of these correlations for the shorter words suggests that lexical influences are reduced in such short words. The only curiosity from this perspective is the pair of extremely robust positive correlations for the short words for the ambiguous and word conditions.

With the large number of listeners tested in the compensation experiments, the correlational data provide strong converging evidence with the main results reported in Experiment 2. In both analyses, the most parsimonious explanation of the data is that lexical activation affects the perception of the word's constituent phonemes, which in turn causes compensation for coarticulation.

General discussion

The central theoretical question that we set out to answer is whether perception of phonemic information is influenced by lexical activation. This question has been at the heart of the debate over the architecture of the system that processes spoken language. Autonomous theories (e.g., Cutler & Norris, 1979; Cutler et al., 1987;

Eimas et al., 1990; Massaro & Oden, 1995; Norris et al., 2000) assert that lexical activation cannot affect the perception of sublexical codes. Interactive theories (e.g., Connine & Clifton, 1987; McClelland & Elman, 1986; Samuel, 1981, 1997), in contrast, predict such top–down influences.

As we noted in the Introduction, there are many demonstrations in the literature of lexical influences on phonetic decisions. As a result of these demonstrations, autonomous theorists have focused their claims on a distinction between phonetic *percepts* and phonetic *decisions*. The clearest statement of this distinction is Norris et al.'s (2000) Merge model. In Merge, lexical activation is explicitly allowed to influence the *report* of a phoneme, but it is not allowed to affect its *perception*. By separating these two processes, almost all of the many demonstrations of lexical influences on phonemic processing are cast aside, because almost all of these tests were designed to show that changes in lexical activation would affect the phonemes that listeners report.

If one accepts how Norris et al. (2000) frame the debate (and as we noted, a number of the commentaries on the Norris et al. paper do not), then the only tests capable of distinguishing between the two model types are ones that rely on measuring lexical influences on phoneme categorization indirectly. The compensation for coarticulation phenomenon is one of the very few domains that have this property, which is why we undertook this project. The results of our experiments, and those in the literature, suggest that clarity in this domain will require a good understanding of three interrelated topics: (a) lexical effects in the compensation phenomenon, (b) the role of transitional probability, and (c) the effects of perceptual grouping.

In trying to integrate these three topics, it is useful to frame the discussion in terms of two statements from Pitt and McQueen's (1998) study. After their first demonstration of significant effects of transitional probability, coupled with a null effect of lexicality on compensation, Pitt and McQueen concluded that "lexical influences are not necessary to observe compensation following ambiguous fricatives" (p. 357). This claim is well supported, as they repeatedly found compensation when transitional probability was manipulated in the absence of lexicality. After their second experiment, which replicated both the significant compensation with the transitional probability manipulation, and the null lexicality effect, Pitt and McQueen stated: "it appears that the lexicon, even when it is involved in fricative decisions, does not influence compensation for coarticulation in perception of the following stop" (p. 359). This claim is quite different than the first one. The initial claim essentially asserted that compensation could be driven by something other than lexicality, e.g., transitional probability. The second claim, in contrast, is that lexicality does not drive compensation. Although the data within Pitt and McQueen's study are consistent with this stronger claim, we believe that the results of our study, and those in the literature, do require a role for lexically driven compensation. We think that the full complement of results can only be accounted for by positing a role for *both* transitional probability and for lexicality, with perceptual grouping having an important influence as well. Therefore, in order to develop a coherent account of the results, we will summarize the evidence for each of these three factors:

Transitional probability

There is now a rather substantial body of evidence showing that listeners are sensitive to the long term probability structure of phoneme sequences. For example, Pitt and Samuel (1995) found that phoneme monitoring was faster in high frequency syllables (i.e., high transitional probability) than in less common ones, where the words in which they occurred were matched for frequency. Vitevitch and Luce (1998, 1999) have found similar sensitivity to the frequency of sublexical units using naming and matching tasks. A growing number of studies have found that initial language learning by infants shows that they are sensitive to the probability structure of their speech input (Jusczyk, Luce, & Charles-Luce, 1994; Mattys, Jusczyk, Luce, & Morgan, 1999; Saffran, Newport, & Aslin, 1996). Thus, in a wide set of circumstances, transitional probability clearly influences speech processing.

In the specific circumstances that currently concern us—compensation and perceptual grouping—there is also strong evidence that transitional probabilities affect processing. In our first experiment, we demonstrated that these statistical properties can influence the perceptual stability of speech input, with low transitional probabilities associated with higher rates of verbal transformations. In the compensation paradigm, Pitt and McQueen (1998) provided unambiguous evidence for an effect of transitional probability on the strength of compensation. In fact, as Pitt and McQueen pointed out, even within Elman and McClelland's (1988) original study there was some evidence that compensation was being influenced by transitional probability. In Elman and McClelland's Experiment 2, the context stimuli were vowel + (ambiguous) fricative syllables based on the endings of their "ridiculous" and "Spanish" context words. Elman and McClelland found a clear but nonsignificant trend toward compensation in the direction predicted by the transitional probabilities of the vowel–consonant sequences. Taken together, all of the evidence suggests that statistical regularities are encoded by listeners, and that the resulting representations influence both compensation and perceptual grouping.[5]

Perceptual grouping

One of the most critical methodological findings of the current study is that the compensation paradigm is extremely sensitive to effects of perceptual grouping. In fact, this sensitivity seriously undermines the utility of the task, because the perceptual grouping effect will generally tend to oppose the lexical effect of interest. In Mann and Repp's (1981) seminal compensation study there was some early evidence for an influence of perceptual grouping. A 32% compensation effect for an isolated fricative preceding a stop was reduced to 19% simply by prepending a vowel, which reduced the fricative's tendency to group with the following stop.

In the current study, three results illustrate how strongly compensation depends on perceptual grouping. First, context stimuli in which the final fricative completed a real word (e.g., "abolish") yielded much *weaker* compensation shifts (4%) than the shifts for ones that created pseudowords (12%; e.g., "aboliss"); words bind their final fricatives more strongly than pseudowords do (as shown in Experiment 1), thereby reducing the compensation shifts. Second, even the 12% shifts found for these three-syllable pseudowords are dwarfed by the 24% effect we found with simple VC contexts. Again, the stronger context binds the fricative more strongly, reducing compensation. Finally, for real word context stimuli, the compensation shifts were very large for monosyllables (23%), moderate for bisyllables (12%) and quite small for three-syllable stimuli (4%). In all three cases, these effects involve unambiguous (i.e., endpoint) fricatives, for which stronger context cannot significantly alter fricative perception. As such, the perceptual grouping effect is not offset by a lexical effect on the fricative, which is why such large grouping effects were evident.

Lexical effects on compensation

Although we believe that the insights gained into the effects of transitional probability and perceptual grouping are quite important, the most critical issue is still whether or not lexical activation can drive compensation for coarticulation. Several results in the current study provide a positive answer to this question, and additional evidence in the literature supports this conclusion. Recall that one of our initial hypotheses was that the different results found by Elman and McClelland (reliable lexical effects) and by Pitt and McQueen (null lexical effects) might be due to differences in the nature of the context words that each used. Elman and McClelland used polysyllabic contexts, whereas Pitt and McQueen used monosyllables. Even though Pitt and McQueen's monosyllables were clearly sufficient to influence fricative identification, they might not have provided sufficiently long-lasting lexical activation to drive compensation for coarticulation.

Thus, our central manipulation involved three-syllable context pairs, with shorter context pairs included for comparison purposes. As predicted, the longer words generally supported compensation, whereas the shorter ones generally did not. Three of the four three-syllable pairs produced lexically mediated compensation, with 24 subjects shifting in the predicted direction, and only 7 producing reversals. We did find one monosyllabic pair that produced reliable shifts as well. This pattern of results is consistent with fricative identification data that we have collected for the shorter and longer words (Pitt & Samuel, 2002), with much larger Ganong effects for the trisyllables than for the monosyllables.

In addition to the four cases of lexically mediated compensation, the set of correlations shown in Table 1 provides further evidence for lexical influences on compensation. In particular, the lexical influences can be seen in the very consistent positive correlation between compensation magnitude for the critical ambiguous condition and that in the real word condition. Similarly, the very reliable negative correlation between the ambiguous and pseudoword conditions is exactly what would be expected if the strength of lexical activation varied across listeners. We believe that the correlational data and the four successful demonstrations of lexically mediated compensation can be accounted for only if lexical activation can affect sublexical perception.

This conclusion is bolstered by several other findings in the literature. Most notably, it is worth revisiting the original Elman and McClelland (1988) paper in the context of the current findings. We have already reviewed the three successful compensation cases in their Experiment 1, and we have noted the preliminary evidence for transitional probability effects in their Experiment 2. For the current purposes, their less-cited Experiment 3 is most interesting. The test series in this experiment was the same "dates–gates" continuum that was one of the three used in their first experiment. Experiment 3 tested two pairs of context items. One pair was based on "ridiculous" versus "Spanish." For this pair, Elman and McClelland created a final VC that was designed to be

[5] One might wonder whether longer-range transitional probabilities (e.g., triphone probability, p(/s/|/kI/)) could account for why some contexts produced compensation and others did not. When the triphone probabilities for all context word pairs were compared with the size of their labeling shifts, this explanation proved unsatisfactory. In particular, there were too many instances in which a triphone account made the wrong prediction. For example, the "fish–kiss" pair yielded one of the most robust effects of lexically mediated compensation, yet based on triphone probabilities a large reversal should have been obtained because /ʃ/ given /kI/ is more than twice as frequent as /s/ given /kI/ (0.214 versus 0.108).

midway between the final VC in "ridiculous" and the final VC in "Spanish." This neutral VC replaced the final VC in each context word. Note that by using this design, Elman and McClelland effectively addressed the vowel–consonant transitional probability issue, because exactly the same vowel (and ambiguous consonant) was used in both context words. The second pair of context words took this control one step further. This pair was based on "ridiculous" versus "foolish," and the final *three* phonemes in both were replaced by a common neutral /l + vowel + fricative/. For both the VC-matched case, and the CVC-matched case, Elman and McClelland found reliable lexically mediated compensation effects.

We are aware of one other case of lexically mediated compensation. Magnuson, McMurray, Tanenhaus, and Aslin (2002) used a "tapes–capes" test series, and critical context stimuli based on "bliss" versus "brush." As Magnuson et al. note, a nice feature of their design is that the final VC in their context items reverses the vowel–consonant co-occurrence of Elman and McClelland. For example, the transitional probability concern over Elman and McClelland's stimuli was based on the possible pronunciation of the final vowel in "Christmas" as more like /^/ or /ə/ than like the /I/ in "foolish." In Magnuson et al.'s contexts, the /^/ occurs in "brush," with the final /ʃ/, while /I/ is coupled with the /s/ in "bliss." By reversing these co-occurrences, Magnuson et al. must reverse any transitional probability effects. Nonetheless, they found a reliable lexically mediated compensation effect.

When the findings of the current study are combined with those of Elman and McClelland and Magnuson et al., there is a very substantial body of evidence that shows lexical influences on compensation for coarticulation. Elman and McClelland report five independent cases, we report four, and Magnuson et al. report one. In addition, the correlational results in Table 1 strongly support this interpretation. Collectively, these data provide a solid basis for asserting that lexical activation can influence phonemic perception.

This conclusion is strengthened by results from a modified selective adaptation paradigm that, like the compensation paradigm, meets the standard of using a consequential measure of lexical influences on phonemic perception. As discussed in the introduction, Samuel (1997) found that listeners' categorization of syllables along a /bI/-/dI/ continuum was affected by prior exposure to an adapting stimulus (e.g., "alphabet"), whose medial stop (e.g., /b/ in this case) was replaced with white noise. This indirect influence on categorization could have occurred only if listeners restored the missing phoneme.

Samuel (2001) recently extended this approach, using a Ganong-type lexical manipulation, rather than phonemic restoration, to drive the adaptation effects. Again, in two experiments, reliable lexically mediated adaptation shifts were found. Thus, in addition to the ten lexically mediated compensation results summarized above (plus the correlational data), there are four additional experiments in the literature that meet the consequential criterion.

Given the sizable number of findings that cannot be attributed to phonemic decisions, we conclude that lexical activation does in fact affect the encoding of lower level information. This result is predicted by interactive models, and is at odds with autonomous ones. Our results also add to the growing literature that shows the perceptual importance of the statistical regularities of the speech input. And, for both the lexical influences on phonemic perception, and the effects of transitional probability, we have seen that more general perceptual grouping effects (Bregman, 1990) can have a critical moderating influence.

The effects of transitional probability and perceptual grouping must be carefully considered in any attempt to study lexically mediated compensation for coarticulation. These possible confounding factors make it extremely difficult to use the compensation paradigm to examine lexical influences on phonemic perception. Until investigators can devise alternative paradigms that meet the standard of using consequential procedures to assess lexical influences on sublexical processing, we believe that the lessons of the current study will be useful to those who wish to pursue this important question.

Acknowledgments

This research was supported by NIMH grant R01-51663 to Arthur Samuel, and by NIDCD grant R29 DC01774 to Mark Pitt. We thank Anna Volkova, Naomi Fetterman, Erik Tracy and Lisa Shoaf for their help in data collection, the anonymous reviewers of a previous draft of this paper, and Donna Kat for her assistance and expertise in many aspects of this project.

References

Bard, E. G., Schillcock, R. C., & Altmann, G. T. M. (1989). The recognition of words after their acoustic offset in spontaneous speech: Effects of subsequent context. *Perception & Psychophysics, 44*, 395–408.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT press.

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.

Cutler, A., & Norris, D. (1979). Monitoring sentence comprehension. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 113–134). Hillsdale, NJ: Erlbaum.

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

Eimas, P. D., Marcovitz Hornstein, S., & Payton, P. (1990). Attention and the role of dual codes in phoneme monitoring. *Journal of Memory and Language, 29*, 160–180.

Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language, 27*, 143–165.

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

Gaskell, G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12*, 613–656.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251–279.

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language, 33*, 630–645.

Klatt, D. H. (1980). Software for a cascade/parallel format synthesizer. *Journal of the Acoustical Society of America, 67*, 971–975.

Magnuson, J., McMurray, B., Tanenhaus, M., & Aslin, R. (2002). Lexical effects on compensation for coarticulation: The ghost of Christmash past. Manuscript in preparation.

Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America, 69*, 548–558.

Massaro, D. W., & Oden, G. C. (1995). Independence of lexical context and phonological information in speech perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1053–1064.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38*, 465–494.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86*, 287–330.

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

Nearey, T. (1990). The segment as a unit of speech perception. *Journal of Phonetics, 18*, 347–373.

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

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5*, 42–45.

Pitt, M. A. (1995). The locus of the lexical shift in phoneme identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1037–1052.

Pitt, M. A., & McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language, 39*, 347–370.

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.

Pitt, M. A., & Samuel, A. G. (1995). Lexical and sublexical feedback in auditory word recognition. *Cognitive Psychology, 29*, 149–188.

Pitt, M. A. & Samuel, A. G. (2002). Lexical activation strength is determined by word length. Manuscript in preparation.

Pitt, M. A., & Shoaf, L. (2001). Beyond traditional measures of lexical influences on perception. *Language and Cognitive Processes, 16*, 715–721.

Pitt, M. A., & Shoaf, L. S. (2002). Linking verbal transformations to their causes. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 150–162.

Pollack, I., & Pickett, J. M. (1963). The intelligibility of excerpts from conversation. *Language and Speech, 6*, 165–171.

Pollack, I., & Pickett, J. M. (1964). Intelligibility of excerpts from fluent speech: Auditory vs. structural context. *Journal of Verbal Learning and Verbal Behavior, 3*, 79–84.

Repp, B. H., & Mann, V. A. (1981). Perceptual assessment of fricative-stop coarticulation. *Journal of the Acoustical Society of America, 69*, 1154–1163.

Repp, B. H., & Mann, V. A. (1982). Fricative-stop coarticulation: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America, 71*, 1562–1567.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newburry Park, CA: Sage Publications.

Rubin, P., Turvey, M. T., & van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in spoken nonwords. *Perception & Psychophysics, 19*, 384–398.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606–621.

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. (1987). Lexical uniqueness effects on phonemic restoration. *Journal of Memory and Language, 26*, 36–56.

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. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science, 12*, 348–351.

Shillcock, R., Lindsey, G., Levy, J., & Chater, N. (1992). A phonologically motivated input representation for the modeling of auditory word perception in continuous speech. In *Proceedings of the 14th Annual Cognitive Science Society Conference, Bloomington* (pp. 408–413).

Shoaf, L. S., & Pitt, M. A. (2002). Does node stability underlie the verbal transformation effect? A Test of Node Structure Theory. *Perception & Psychophysics*.

Treiman, R. (1986). The division between onsets and rimes in English syllables. *Journal of Memory and Language, 25*, 476–491.

Treiman, R., & Zukowski, A. (1996). Children's sensitivity to syllables, onsets, rimes, and phonemes. *Journal of Experimental Child Psychology, 61*, 193–215.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in spoken word perception. *Psychological Science, 9*, 325–329.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40*, 374–408.

Warren, R. M. (1961). Illusory changes of distinct speech upon repetition—the verbal transformation effect. *British Journal of Psychology, 52*, 249–258.

Warren, R. M. (1968). Verbal transformation effect and auditory perceptual mechanisms. *Psychological Bulletin, 70*(4), 261–270.

Wurm, L. H., & Samuel, A. G. (1997). Lexical inhibition and attentional allocation during speech perception: Evidence from phoneme monitoring. *Journal of Memory and Language, 36*, 165–187.