

2-1-2006

Putting phonetic context effects into context: A commentary on Fowler (2006)

AJ. Lotto

L L. Holt
lholt@andrew.cmu.edu

Follow this and additional works at: <http://repository.cmu.edu/psychology>

Published In

PERCEPTION & PSYCHOPHYSICS, 68, 2, 178-183.

This Article is brought to you for free and open access by the Dietrich College of Humanities and Social Sciences at Research Showcase @ CMU. It has been accepted for inclusion in Department of Psychology by an authorized administrator of Research Showcase @ CMU. For more information, please contact research-showcase@andrew.cmu.edu.

Putting phonetic context effects into context: A commentary on Fowler (2006)

ANDREW J. LOTTO
University of Texas, Austin, Texas

and

LORI L. HOLT
Carnegie Mellon University, Pittsburgh, Pennsylvania

On the basis of a review of the literature and three new experiments, Fowler (2006) concludes that a *contrast account* for phonetic context effects is not tenable and is inferior to a gestural account. We believe that this conclusion is premature and that it is based on a restricted set of assumptions about a general perceptual account. Here, we briefly address the criticisms of Fowler (2006), with the intent of clarifying what a general auditory and learning approach to speech perception entails.

Perception is not an isolated event. Context stimuli neighboring in time and space affect the encoding, representation, and interpretation of target stimuli. Phonetic context effects are a set of particularly interesting examples of context-sensitive perception, whereby identification of target speech sounds is shifted by the preceding or following phonetic context. For example, identification of an ambiguous target syllable as /da/ or /ga/ is shifted by preceding /ar/ or /al/ contexts (Mann, 1980).

A variety of studies have demonstrated that the perceived identity of target speech sounds also can be shifted by nonspeech sounds that mimic the spectral characteristics of effective phonetic contexts. For example, Lotto and Kluender (1998) presented /da/-/ga/ syllables preceded by tone glides matched in frequency to the third formant (F_3) transition of /al/ or /ar/. Listeners' identification of target syllables was shifted in the same direction as when targets were preceded by the speech contexts. Even steady-state tones at the offset frequency of /al/ or /ar/ F_3 resulted in a significant shift in speech target identification. In fact, the shift was the same size as that reported for speech contexts (Mann, 1980). Multiple replications of this effect have been found across classes of targets and contexts (see Table 1).

In all of these cases, the shift in target identification can be described as contrastive. For example, an acoustically ambiguous syllable between /da/ and /ga/ is labeled as /ga/ (lower F_3 onset frequency) following a higher frequency tone (modeling /al/ F_3 offset frequency) and is labeled as /da/ (higher F_3 onset) following a lower frequency tone (modeling /ar/ F_3 offset). We have referred to this pattern of perceptual behavior as *spectral contrast* and have proposed that it is a result of general auditory

processes that affect speech and nonspeech perception alike (Holt, in press; Holt & Lotto, 2002; Lotto & Kluender, 1998; Lotto, Sullivan, & Holt, 2003). In agreement with this proposal, contrastive effects of speech contexts on the perception of nonspeech targets also are observed (Stephens & Holt, 2003).

On the basis of our previous work, Fowler (2006) ascribes to us a *spectral contrast account* of context effects and compares it with gestural accounts, such as motor theory (Liberman & Mattingly, 1985) and direct realism (Fowler, 1986). Our theoretical stance is that speech perception can be understood in terms of the operating characteristics of perceptual and cognitive systems, without requiring a specialized module for speech (motor theory) or access to invariant information specifying speakers' articulations (direct realism). (For a recent review of these theories, see Diehl, Lotto, & Holt, 2004.) Given the results summarized in Table 1, we have proposed that perceptual contrast plays a substantial role in explaining many of the phonetic context effects that have been studied. However, it may be too generous to refer to spectral contrast as an *account*. In no way does contrast account for all of speech perception. It is clear that there are many other potential sources for context effects, including knowledge and expectations (e.g., syntactic, lexical, and transitional probability information in speech), as well as information from other senses (e.g., vision). The point is that these various sources can be integrated into a cognitive model of perception in context. An important component of the model is contrast, which describes the patterns of responses seen with speech and nonspeech contexts and speech and nonspeech targets alike.

On the basis of a review of the literature and three new experiments, Fowler (2006) concludes that a "contrast account" of phonetic context effects is not tenable and is inferior to a gestural account. We believe that this conclusion is premature and that it is based on a restricted set of assumptions about a general perceptual account. Below,

Correspondence concerning this article should be addressed to A. J. Lotto, Center for Perceptual Systems, University of Texas, Austin, TX 78712 (e-mail: andrew@mail.cps.utexas.edu).

Table 1
Partial Summary of Effects of Context

Speech → speech	
Human native English adult listeners	e.g., Mann, 1980
Nonnative listeners who do not discriminate the context speech sounds	Mann, 1986
4-month-old human infants	Fowler, Best, & McRoberts, 1990
Nonhuman subjects, Japanese quail	Lotto, Kluender, & Holt, 1997
Female speech contexts affect perception of male speech targets.	Lotto & Kluender, 1998
Voiceless contexts with acoustic characteristics less likely to promote spectral contrast do not influence speech perception, whereas their voiced counterparts with similar articulations do.	Holt, Lotto, & Kluender, 2000
Cochlear implant users who can phonetically label speech contexts do not show evidence of context effects on speech targets due to poor spectral resolution; they do exhibit temporal effects.	Aravamudhan & Lotto, 2004, 2005
Nonspeech → speech	
Nonspeech contexts affect consonant identification.	Coady, Kluender, & Rhode, 2003; Fowler, Brown, & Mann, 2000; Holt, 1999; Holt & Lotto, 2002; Lotto & Kluender, 1998
Nonspeech contexts affect vowel identification.	Holt, Lotto, & Kluender, 2000
Nonspeech context effects persist with dichotic presentation of context and target, indicating that effects are not purely cochlear.	Lotto, Sullivan, & Holt, 2003
Nonspeech effects decay with silence between context and target along same time course as speech effects.	Holt & Lotto, 2002; Lotto, Sullivan, & Holt, 2003
Nonspeech contexts that are nonadjacent to the speech target affect speech identification.	Holt, 2005
Nonspeech effects persist across 1.3 sec of silence between context and speech target, indicating cortical processes.	Holt, 2005
Nonspeech effects persist when up to 13 sounds intervene between context and speech target, indicating cortical processes.	Holt, 2005
Nonspeech affects preceding speech targets (a right-to-left effect).	Wade & Holt, 2005
Speech → nonspeech	
Preceding speech affects discrimination of nonspeech targets as a function of the spectral make-up of stimuli.	Stephens & Holt, 2003
Speech affects identification of preceding nonspeech targets (a right-to-left effect).	Fowler, 2006
Nonspeech → nonspeech	
Nonspeech context affects nonspeech identification.	Aravamudhan & Hawks, 2004

Note—All effects shown here are contrastive in their influence.

we will briefly address the criticisms of Fowler (2006), with the intent of clarifying what a general auditory and learning approach to speech perception entails.

Criticism 1: Contrast Effects Are Restricted to Left-to-Right Effects

Most of the demonstrations of phonetic context effects that have been studied are designed so that context precedes the target (e.g., /a/ before /da-/ga/). Likewise, demonstrations of spectral contrast with nonspeech contexts have been *left to right* in time. However, there are some phonetic context effects that are *right to left*, or backward, with the target *preceding* the context.

Fowler (2006) argues that these effects invalidate the “spectral contrast account” because contrast cannot work backward in time. However, there have been previous demonstrations of seemingly right-to-left nonspeech contrast effects. Diehl and Walsh (1989) have reported that the perception of the rate of a frequency-modulated tone can be influenced (contrastively) by the duration of a following steady-state tone.

Backward contrastive effects have also been observed for nonspeech interactions with speech. Experiment 3 in Fowler (2006) demonstrates a right-to-left contrast effect

with nonspeech target and speech context. Similarly, Wade and Holt (2005) have reported a backward contrast effect whereby nonspeech contexts influence preceding speech targets. Interestingly, these nonspeech/speech backward context effects can be contrastive or assimilative, depending on small changes in the timing of the target and context stimuli. The response patterns in Fowler’s (2006) Experiment 2, which appear to be assimilative, may have analogues in nonspeech effects.

Fowler (2006) provides two reasons for discounting these examples of right-to-left contrast effects. The first is that the logic of comparing listeners’ responses to speech and nonspeech is flawed, because the perception of nonspeech is likely to be different from speech perception, even if the response patterns are qualitatively similar (see Fowler, 1990). However, this criticism does not apply to Wade and Holt’s (2005) results or to many of the previous contrast examples. In these cases, listeners’ *speech perception* was affected by the nonspeech contexts.

Fowler (2006) also maintains that contrast effects are by definition left-to-right. We do not accept that there is such a restrictive definition of contrast. Spectral contrast is a description of a pattern of results mapped onto relations between context and target stimuli. It does not refer

to a particular mechanism. Initially, we hypothesized that peripheral neural adaptation may play a substantial role in spectral contrast effect, but our own subsequent studies have demonstrated that the time course of the effect and its robustness to dichotic presentation require central mechanisms (Holt, 2005; Holt & Lotto, 2002; Lotto et al., 2003). Higher level perceptual interactions provide a wider time window of integration and the possibility of later information's affecting the encoding of preceding information.

Criticism 2: Some Speech Effects Are Simultaneous or Are Not Contrastive

Fowler (2006) presents a few examples of effects for which spectral contrast is not appropriate. For example, the perception of a fundamental frequency (f_0) contour can change as a function of vowel height (Silverman, 1987) or consonant voicing (Pardo & Fowler, 1997). In both cases, listeners judge pitch relative to what is typical for the phonetic segment; an /i/ is perceived as lower in pitch, relative to an /a/ with the same f_0 , because /i/ typically has a higher f_0 . This is not a case in which spectral contrast applies. We believe that this result is predictable from a general learning account whereby categories are formed from correlations among stimulus parameters. Moreover, we have demonstrated that even birds can learn such correlations from speech sounds (Holt, Lotto, & Kluender, 2001).

As we argued above, it is unreasonable to expect that spectral contrast, a single perceptual pattern, can account for all speech perception phenomena. Fowler (2006) argues that this demonstrates the superiority of gestural accounts, because they provide one parsimonious explanation for all of speech perception (parsing of the signal along gestural lines), whereas a nongestural account requires multiple mechanisms. However, the application of Occam's razor depends on how one chooses to divide the phenomena of interest. The nongestural approach provides one explanation for the context effects summarized in Table 1, whereas gestural theories require multiple explanations to cover these observations (Fowler, Brown, & Mann, 2000).

Criticism 3: Spectral Contrast Results in a Loss of Sensitivity to Coarticulatory Information

Fowler's (2006) critique of spectral contrast in speech perception repeatedly states that contrast results in a "transient loss of sensitivity to spectral information" or that "it eliminates perceptual sensitivity to assimilative effects." This description presumes a particular type of mechanism underlying spectral contrast. We have ruled out the possibility that spectral contrast is due solely to peripheral adaptation or masking (Holt, 2005; Lotto et al., 2003). Neural mechanisms exist (e.g., active recalibration [Anstis, 1975] or adaptation of suppression [Viemeister & Bacon, 1982]) that can produce contrast without loss of sensitivity (Holt & Lotto, 2002).

According to Fowler (2006), a loss of sensitivity is incompatible with results in which coarticulatory effects are used by listeners as information for context phonemes. For example, Whalen (1984) observed that when frication from a natural production of /sa/ is spliced to precede /u/, listeners' labeling of the cross-spliced /su/ is slower than labeling of natural /su/ tokens. Thus, it appears that information in the coarticulated vowel aids fricative identification.

It is almost certainly the case that there are regularities in acoustics that result from coarticulation (see, e.g., Sussman, McCaffrey, & Matthews, 1991) and that listeners are sensitive to these regularities in speech perception (Holt, Lotto, & Kluender, 1998; Nearey, 1997; Smits, 2001). Nevertheless, spectral contrast may play a supporting role even in these cases. The original samples of speech to which Whalen (1984) compared the cross-spliced examples were naturally coarticulated. Due to coarticulation, adjacent frequencies of sounds are assimilated. Cross-splicing destroys this naturally assimilative relationship, moving frequencies of adjacent sounds further apart. This difference matters because spectral contrast is graded in its influence. Holt (1999) demonstrated that there is a range across which contrast has its greatest influence and that contrastive effects decrease with spectral distance. Cross-splicing serves to distance spectral energy of adjacent sounds, moving energy outside the most effective range of spectral contrast (that of naturally coarticulated speech) and, thus, lessening the effects of contrast. As such, the influence of contrast can be expected to be stronger for the original coarticulated speech productions than for the cross-spliced stimuli. These effects do not require gestural mediation for speech perception.

Criticism 4: Context Effects Can Occur Without Changes in the Acoustic Makeup of the Context

Fowler et al. (2000) have presented what appears to be a visually moderated phonetic context effect whereby a video corresponding to /al da/ or /ar da/ accompanies an acoustically ambiguous context syllable preceding /da/–/ga/ targets. The acoustic context was thought to be disambiguated by the visual information, thus shifting speech identification to more /ga/ responses with the /al da/ video. Fowler et al. (2000) correctly pointed out that this effect cannot be explained by spectral contrast, because there was no change in the acoustic context.

In fact, there have been previous demonstrations of context effects in which acoustically ambiguous contexts were disambiguated by other information, such as lexical status (Elman & McClelland, 1988; Magnuson, McMurray, Tanenhaus, & Aslin, 2003; Samuel & Pitt, 2003) and transitional probabilities (Pitt & McQueen, 1998). These designs are not tests of a spectral contrast "account," because we have never stated that context effects cannot occur in the absence of spectral contrast; our claim is that when contexts differ in spectral characteristics, spectral contrast affects the encoding of the target stimulus. There is nothing about the use of visual information in speech

perception that is inconsistent with a general nongestural approach, as has been demonstrated by the work of Masaro (1987, 1998).

Given these caveats, it is actually surprising how little evidence there is for a robust visually moderated context effect. Using new stimulus materials, Holt, Stephens, and Lotto (2005) recently attempted replication of Fowler et al. (2000) but failed to find any shift in target identification related to participants' identification of the context (see also Vroomen & de Gelder, 2001). A major difference in the stimuli used in these two studies was that Fowler et al. (2000) included video during the target syllable, whereas Holt et al. (2005) displayed video only during the context. When the Fowler et al. (2000) stimuli were edited to include video only during the context, they produced no effect. However, when only the target audio and video of the Fowler et al. (2000) stimuli were presented (i.e., when there was no context), a shift in responses equivalent to the original Fowler et al. (2000) context effect was observed. Holt et al. (2005) concluded that there were subtle differences in the video presented simultaneously with the target that had nothing to do with disambiguating the context and that these differences influenced the listeners' speech identification. That is, the original results were not a context effect.

Fowler (2006) suggests that differences in the video simultaneous with the target are informative of the context and that the participants are, in fact, parsing from the acoustics the presumed coarticulatory effects of the context gestures that are apparent in video simultaneous with the target. The problem with this account is that it does not explain why there was no effect of context when the context was actually present. That is, in the original replication attempted by Holt et al. (2005), context was physically present, but there was no video during the target and no context effect; the listeners did not parse the effects of context. On the other hand, Fowler (2006) argues that when there was no context, the listeners parsed the context from the subtle video clues present during the target. This would appear to violate the expectations of a gestural theory. As is noted by Fowler (2006), "A left context provides even better information than does the coarticulatory information with /da/ or /ga/ as to the nature of the coarticulatory influence" (see note 6). However, that is clearly not the case for the audiovisual effects observed by Fowler et al. (2000).

A similar problem is evident in Experiment 1 in Fowler (2006). In that experiment, /da/ and /ga/ tokens were spliced from natural productions of /al da/, /al ga/, /ar da/, and /ar ga/. These /da-/ga/ syllables were used to create series of stimuli that were presented in isolation for identification. Again, this was not a test of spectral contrast in any way, because there was *no* context. The results indicated more /ga/ responses for syllables originating from productions following /al/. Note that this demonstration is different from previous examples of phonetic context effects. It is typical to demonstrate that stimuli with identical acoustics are perceived differently or that stimuli with dif-

ferent acoustics are perceived the same. This experiment shows that different acoustics are perceived differently.

Fowler (2006) claims that listeners are compensating for the coarticulatory effects of the missing contexts. There are several problems with this account. The first is that if this had really been an example of perceptual compensation, all of the series of stimuli should have been labeled similarly. If coarticulation perverts the acoustics of the syllable, compensating for its effects will result in percepts like those that would result if the acoustics were similar. But this is not what happens. Fowler (2006) therefore suggests that listeners are *overcompensating*. It is hard to imagine why this would happen if the perceptual system is really fine-tuned to the invariants that specify articulations, as would be expected according to a direct realist theory. Also, although there was actually a context stimulus present in the original demonstration of this phonetic context effect, listeners *undercompensated* (as can be judged from Figures 2 and 4 in Mann, 1980). A second problem with this explanation is that the silence preceding these stimuli is not just an absence of context; it is also *information* that no context is present (as opposed to a preceding noise that may be masking the context). It is odd to posit that listeners overcompensate for a context that is clearly not present. Finally, the participants in this task could not identify above chance the contexts from which these syllables originated. Fowler (2006) suggests that this is just a case of unconscious perception, but this explanation is hard to motivate. According to the direct realist approach, listeners are parsing the information for the context and the target from the acoustics, and this information overlaps in time. Why, then, would the target, but not the context, information be available to consciousness, given that both of them are specified by invariant information in the signal? Experiment 1 in Fowler (2006) does not seem to offer support for any current theory.

The Adequacy of Gestural Theories

Fowler's (2006) review of context effects gives the impression that gestural theories are capable of accounting for all of the context effects involving speech and, if one disregards effects involving nonspeech or animals, gestural theories are adequate. However, there are effects involving speech contexts and speech targets that raise problems for gestural theories. Lotto and Kluender (1998) found that listeners' identification of male /da-/ga/ targets shifted as a function of female /al/ and /ar/ contexts. This is troubling for accounts of context effects that consider them to be compensation for coarticulation, because there are no coarticulatory influences across two distinct speakers. If listeners have a special system that has evolved a knowledge of vocal tract dynamics (motor theory) or can pick up these dynamics directly from the signal (direct realism), they should not compensate across independent vocal tracts. However, if the spectral makeup of the context drives the effect, these effects are predictable from the spectra.

A second example comes from Holt, Lotto, and Kluender (2000). In their Experiment 1, the identity of a medial

vowel was shifted, depending on whether it was abutted on each end by /d/s or /b/s. This is a classic contrast effect involving the frequency of the second formant (F_2) and also can be induced by tones that follow the F_2 trajectories of the consonants. In their Experiment 4, /b/ and /d/ were replaced by /p/ and /t/. Whereas there was a large context effect elicited by the /b/ and /d/ contexts, there was no context effect for /p/ and /t/. This is a problem for gestural theories because the supraglottal gestures for /p/ and /t/ are similar to those for /b/ and /d/. If listeners are parsing the coarticulatory consequences of the context, the voicing status of the context should not matter. The result was predicted from spectral contrast, because the formant trajectories of the contexts were excited only by aspiration energy and, therefore, had lower amplitudes and broader bandwidths than did their voiced counterparts. These acoustic characteristics made spectral contrast with the target less likely for /p/ and /t/ contexts.

As a final example, postlingually deafened adults with cochlear implants do not show the same phonetic context effects that normal hearing listeners do, despite being able to accurately label the phonetic contexts (Aravamudhan & Lotto, 2004, 2005). The degraded spectral resolution of the implants appears to be sufficient for labeling context sounds, but not specific enough to produce spectral contrast. In support of this explanation, cochlear-implanted listeners exhibit strong temporal contrast effects (identification of /ba/ or /wa/ as a function of duration of following vowel; Miller & Liberman, 1979) because the temporal resolution of implants is rather good. Gestural theories cannot account for these results.

The Viability of Contrast in a General Perceptual and Cognitive Account of Speech Perception

Our conclusions from the experiments summarized in Table 1 are that spectral contrast is real and that the critique and experiments of Fowler (2006) do not invalidate its explanatory potential. It is hard to imagine that such a robust phenomenon would have no effect on the encoding of speech sounds. But is it possible that its role is minimal and that other aspects of speech processing easily override it?

Holt (in press) placed nonspeech contexts in competition with speech contexts for influence on a target speech syllable. Targets from a /da/-/ga/ series were preceded by a speech /al/ or /ar/ context, which itself was preceded by a series of tones shown by Holt (2005) to affect /da/-/ga/ identification. Despite the fact that the speech context was temporally adjacent to the target, target identification was best predicted by the nonspeech contexts (e.g., more /ga/ responses with high-frequency tones, even with a preceding /ar/, which predicts more /da/ responses). This influence indicates that spectrally contrastive effects of nonspeech are not just minor illusions of perception but are indicative of important interactions that occur in the perception of sounds in context.

We take the large set of findings in Table 1 to be *prima facie* evidence for the existence of a general auditory op-

erating characteristic that we call *spectral contrast*. This contrastive perceptual behavior is similar in many respects to what occurs in the visual perception of brightness, color, visual curvature, and so forth, in that perception is relative. That is, instead of providing veridical values of sensory parameters at each moment in time, perceptual systems encode stimulus values relative to surrounding context. This principle can result in contrastive perception: for example, the F_3 onset of a syllable will be perceived as lower in frequency, relative to neighboring high-frequency energy. Similar demonstrations of how spectral manipulations to sentence length speech contexts influence target speech identification have been made by a number of researchers (Kluender, Coady, & Kiefte, 2003; Ladefoged & Broadbent, 1957; Watkins & Makin, 1996). In each of these cases, the long-term spectral characteristics of the context phrase shifts perception of target speech stimuli in such a way that listeners are most likely to report hearing the speech sound with the spectrum most different from the precursor.

Fowler (2006) rightly points out that spectral contrast cannot account for all of speech perception. We agree. The spectral contrast "account" in no way presumes to be a full theory of speech perception. However, the significant body of empirical findings that indicate a role for spectral contrast in speech perception is wholly consistent with our theoretical stance that speech perception relies on general perceptual and cognitive mechanisms. To account for the full constellation of findings presented in Table 1, it is necessary to move past gestural theories and examine speech communication within the framework of general perceptual and cognitive sciences.

REFERENCES

- ANSTIS, S. M. (1975). What does visual perception tell us about visual coding? In M. S. Gazzaniga & C. Blakemore (Eds.), *Handbook of psychobiology* (pp. 269-323). New York: Academic Press.
- ARAVAMUDHAN, R., & HAWKS, J. W. (2004). Role of experience in eliciting perceptual overshoot with sine waves [Abstract]. *Journal of the Acoustical Society of America*, **115**, 2629.
- ARAVAMUDHAN, R., & LOTTO, A. J. (2004). Perceptual overshoot in listeners with cochlear implants [Abstract]. *Journal of the Acoustical Society of America*, **116**, 2523.
- ARAVAMUDHAN, R., & LOTTO, A. J. (2005, March). *Phonetic context effects in adult cochlear implant listeners*. Paper presented at the 10th Symposium on Cochlear Implants in Children, Dallas.
- COADY, J. A., KLUENDER, K. R., & RHODE, W. S. (2003). Effects of contrast between onsets of speech and other complex spectra. *Journal of the Acoustical Society of America*, **114**, 2225-2235.
- DIEHL, R. L., LOTTO, A. J., & HOLT, L. L. (2004). Speech perception. *Annual Review of Psychology*, **55**, 149-179.
- DIEHL, R. L., & WALSH, M. A. (1989). An auditory basis for the stimulus-length effect in the perception of stops and glides. *Journal of the Acoustical Society of America*, **85**, 2154-2164.
- ELMAN, J. L., & MCCLELLAND, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory & Language*, **27**, 143-165.
- FOWLER, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, **14**, 3-28.
- FOWLER, C. A. (1990). Sound-producing sources as objects of perception: Rate normalization and nonspeech perception. *Journal of the Acoustical Society of America*, **88**, 1236-1249.

- FOWLER, C. A. (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception & Psychophysics*, **68**, 161-177.
- FOWLER, C. A., BEST, C. T., & McROBERTS, G. W. (1990). Young infants' perception of liquid coarticulatory influences on following stop consonants. *Perception & Psychophysics*, **48**, 559-570.
- FOWLER, C. A., BROWN, J. M., & MANN, V. A. (2000). Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 877-888.
- HOLT, L. L. (1999). Auditory constraints on speech perception: An examination of spectral contrast. *Dissertation Abstracts International*, **61**, 556B.
- HOLT, L. L. (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychological Science*, **16**, 305-312.
- HOLT, L. L. (in press). Speech categorization in context: Joint effects of nonspeech and speech precursors. *Journal of the Acoustical Society of America*.
- HOLT, L. L., & LOTTO, A. J. (2002). Behavioral examinations of the level of auditory processing of speech context effects. *Hearing Research*, **167**, 156-169.
- HOLT, L. L., LOTTO, A. J., & KLUENDER, K. R. (1998). Incorporating principles of general learning in theories of language acquisition. In M. Gruber, C. D. Higgins, K. S. Olson, & T. Wysocki (Eds.), *Chicago Linguistic Society: The panels* (Vol. 34, pp. 253-268). Chicago: Chicago Linguistic Society.
- HOLT, L. L., LOTTO, A. J., & KLUENDER, K. R. (2000). Neighboring spectral content influences vowel identification. *Journal of the Acoustical Society of America*, **108**, 710-722.
- HOLT, L. L., LOTTO, A. J., & KLUENDER, K. R. (2001). Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement? *Journal of the Acoustical Society of America*, **109**, 764-774.
- HOLT, L. L., STEPHENS, J. D. W., & LOTTO, A. J. (2005). A critical evaluation of visually-moderated phonetic context effects. *Perception & Psychophysics*, **67**, 1102-1112.
- KLUENDER, K. R., COADY, J. A., & KIEFTE, M. J. (2003). Sensitivity to change in perception of speech. *Speech Communication*, **41**, 59-69.
- LADEFOGED, P., & BROADBENT, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, **29**, 98-104.
- LIBERMAN, A. M., & MATTINGLY, I. G. (1985). The motor theory of speech perception revised. *Cognition*, **21**, 1-36.
- LOTTO, A. J., & KLUENDER, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*, **60**, 602-619.
- LOTTO, A. J., KLUENDER, K. R., & HOLT, L. L. (1997). Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *Journal of the Acoustical Society of America*, **102**, 1134-1140.
- LOTTO, A. J., SULLIVAN, S. C., & HOLT, L. L. (2003). Central locus for nonspeech context effects on phonetic identification. *Journal of the Acoustical Society of America*, **113**, 53-56.
- MAGNUSON, J. S., McMURRAY, B., TANENHAUS, M. K., & ASLIN, R. N. (2003). Lexical effects on compensation for coarticulation: The ghost of Christmas past. *Cognitive Science*, **27**, 285-298.
- MANN, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, **28**, 407-412.
- MANN, V. A. (1986). Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of English "l" and "r." *Cognition*, **24**, 169-196.
- MASSARO, D. W. (1987). *Speech perception by ear and eye: a paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum.
- MASSARO, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.
- MILLER, J. L., & LIBERMAN, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semi-vowel. *Perception & Psychophysics*, **25**, 457-465.
- NEAREY, T. M. (1997). Speech perception as pattern recognition. *Journal of the Acoustical Society of America*, **101**, 3241-3254.
- PARDO, J. S., & FOWLER, C. A. (1997). Perceiving the causes of coarticulatory acoustic variation: Consonant voicing and vowel pitch. *Perception & Psychophysics*, **59**, 1141-1152.
- PITT, M. A., & McQUEEN, J. M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory & Language*, **39**, 347-370.
- SAMUEL, A. G., & PITT, M. A. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory & Language*, **48**, 416-434.
- SILVERMAN, K. (1987). *The structure and processing of fundamental frequency contours*. Unpublished doctoral dissertation, Cambridge University.
- SMITS, R. (2001). Evidence for hierarchical categorization of coarticulated phonemes. *Journal of Experimental Psychology: Human Perception & Performance*, **27**, 1145-1162.
- STEPHENS, J. D. W., & HOLT, L. L. (2003). Preceding phonetic context affects perception of nonspeech. *Journal of the Acoustical Society of America*, **114**, 3036-3039.
- SUSSMAN, H. M., McCAFFREY, H. A., & MATTHEWS, S. A. (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America*, **90**, 1309-1325.
- VIEMEISTER, N. F., & BACON, S. P. (1982). Forward masking by enhanced components in harmonic complexes. *Journal of the Acoustical Society of America*, **71**, 1502-1507.
- VROOMEN, J., & DE GELDER, B. (2001). Lipreading and the compensation for coarticulation mechanism. *Language & Cognitive Processes*, **16**, 661-672.
- WADE, T., & HOLT, L. L. (2005). Effects of later-occurring nonlinguistic sounds on speech categorization. *Journal of the Acoustical Society of America*, **118**, 1701-1710.
- WATKINS, A. J., & MAKIN, S. J. (1996). Effects of spectral contrast on perceptual compensation for spectral-envelope distortions. *Journal of the Acoustical Society of America*, **99**, 3749-3757.
- WHALEN, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception & Psychophysics*, **35**, 49-64.

(Manuscript received March 14, 2005;
revision accepted for publication April 11, 2005.)