ELSEVIER

Hearing Research 167 (2002) 156–169

HEARING RESEARCH

www.elsevier.com/locate/heares

Behavioral examinations of the level of auditory processing of speech context effects

Lori L. Holt [a,*], Andrew J. Lotto [b]

[a] *Department of Psychology, The Center for the Neural Basis of Cognition, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA*

[b] *Department of Psychology, Washington State University, Pullman, WA, USA*

Received 20 July 2001; accepted 19 March 2002

Abstract

One of the central findings of speech perception is that identical acoustic signals can be perceived as different speech sounds depending on adjacent speech context. Although these *phonetic context effects* are ubiquitous in speech perception, their neural mechanisms remain largely unknown. The present work presents a review of recent data suggesting that spectral content of speech mediates phonetic context effects and argues that these effects are likely to be governed by general auditory processes. A descriptive framework known as *spectral contrast* is presented as a means of interpreting these findings. Finally, and most centrally, four behavioral experiments that begin to delineate the level of the auditory system at which interactions among stimulus components occur are described. Two of these experiments investigate the influence of diotic versus dichotic presentation upon two phonetic context effects. Results indicate that context effects remain even when context is presented to the ear contralateral to that of the target syllable. The other two experiments examine the time course of phonetic context effects by manipulating the silent interval between context and target syllables. These studies reveal that phonetic context effects persist for hundreds of milliseconds. Results are interpreted in terms of auditory mechanism with particular attention to the putative link between auditory enhancement and phonetic context effects.

Key words: Speech perception; Phonetic perception; Context effect; Spectral contrast; Auditory enhancement

1. Introduction

In perception, context matters. Both behaviorally and physiologically, there is substantive evidence across perceptual systems that adjacent stimuli have an influence on how their neighbors are encoded and perceived. In the domain of audition, context is extraordinarily important to the perception of speech. Whereas the neural mechanisms underlying effects of context are quite well understood for some perceptual context effects (e.g. Mach bands and lateral inhibition in vision, Hartline and Ratliff, 1954), the neural underpinnings of effects of context on speech perception remain to be determined. Recent findings in behavioral speech perception indicate that context effects in speech may be well predicted from spectral information. As such, these effects appear to be quite general in nature and amenable to neurophysiological investigation in non-human species. In hopes of guiding future neurophysiological investigation, we describe these recent results and report new data from four experiments that explore the level of auditory processing that gives rise to effects of context in speech perception.

Despite the fact that there is little known about the physiology underlying context effects in speech, there is a great deal of behavioral evidence to demonstrate that perception of speech sounds is strongly influenced by adjacent speech context. Examples among the long list of context effects observed in phonetic perception occur at a number of linguistic levels. For example, phoneme identification can be shifted by semantic information (Connine, 1987; Borsky et al., 1998). An intermediate member of a phonetic series that varies from *goat* to *coat*, for instance, is more likely to be identified as

---

* Corresponding author. Tel.: +1 (412) 268 4964; Fax: +1 (412) 268 2798.
E-mail address: lholt@andrew.cmu.edu (L.L. Holt).

0378-5955/02/$ – see front matter
PII: S0378-5955(02)00383-0

'goat' when it is presented in the context of a sentence like 'The laughing dairyman hurried to milk the [*goat/coat*] in the drafty barn' than when it is accompanied by a sentence like 'The expert tailor tried to shorten the [*goat/coat*] in the cluttered attic' (Borsky et al., 1998).

Lexical information also produces effects of context in phonetic categorization. Ganong (1980) had listeners label stimuli from a phonetic series ranging perceptually from [d] to [t] that formed either words or non-words (e.g. *dash–tash* or *dask–task*). Under these conditions, there was a lexical context effect such that listeners were more likely to label perceptually ambiguous mid-series [d] to [t] stimuli as words than non-words. For example, listeners more often labeled members of the [d] to [t] series as [d] for the series *dash–tash* and [t] for the series *dask–task*.

At a somewhat lower level of analysis, context effects also are observed under conditions in which there are no lexical or semantic sources of information. The fundamental characteristic of these *phonetic context effects* is that a target sound's perceived identity may change as a function of neighboring speech sounds even though the target's acoustics remain constant. Mann (1980), for example, has reported that a syllable synthesized to be perceptually ambiguous in isolation – sometimes perceived as /ga/, sometimes perceived as /da/ – is strikingly more often identified as /ga/ when the syllable /al/ precedes it. The same syllable is perceived as /da/ when /ar/ precedes it. The preceding context has an influence on perception even when a 50-ms silent interval separates context and target and even though the acoustics of the target remain the same.

Behavioral evidence for phonetic context effects is extremely rich. Such effects have been documented across many phonetic contexts (e.g. Lindblom and Studdert-Kennedy, 1967 [vowels]; Mann and Repp, 1981 [fricatives]; Mann and Repp, 1980 [stop-consonants]; see Repp, 1982 for a review), as well as among pre-linguistic infant listeners (Fowler et al., 1990) and among speakers of foreign languages (Mann, 1986).

To date, most accounts of phonetic context effects have suggested that they are mediated by speech-specific mechanisms or mechanisms closely linked to representations of the speech production system (e.g. Mann and Repp, 1980; Repp, 1982; Fowler et al., 1990, 2001). An influential example, the Motor Theoretic account, suggests that speech signals have properties uniquely from other sounds in that they carry information about the gestures of speech articulations. By this account, special perceptual mechanisms have evolved to accommodate this distinctive information. Phonetic context effects are explained as a consequence of perceptual compensation for coarticulation, the natural articulatory overlap among adjacent sounds in fluent speech. The Motor Theory suggests that the mechanisms of compensation are intimately linked with the representations responsible for speech production. A second view suggests that listeners attune to specific structure in the acoustic speech signal that serves as information about its source, i.e. the articulatory gestures that produced it (e.g. Fowler, 1996). Fowler et al. (2001), for example, have argued that listeners' use of the structure in the acoustic speech signal as information for its articulatory cause underlies the tendency of listeners to exhibit phonetic context effects.

Several recent empirical investigations have provided two lines of evidence to suggest that speech context effects are less tightly coupled to speech production, per se, and more dependent on the spectral content of the speech (Lotto et al., 1997; Lotto and Kluender, 1998; Holt, 1999; Holt et al., 2000a,b). As such, phonetic context effects appear to be much more amenable to psychoacoustic and physiological investigation than once thought.

Here, we describe the recent results in the descriptive framework of 'spectral contrast' (Lotto and Kluender, 1998) and present data from experiments that use psychophysical techniques to begin to tease apart how audition may arrive at distinct percepts from identical signals. We wish to provide behavioral data that begin to characterize how context effects may arise in audition with the hopes of precipitating new physiological investigation of the mechanisms involved in phonetic context effects.

1.1. Evidence for the role of spectral content

The first line of evidence suggesting that phonetic context effects are a consequence of interactions of the spectral content of adjacent sounds involves a correspondence between speech and non-speech sounds in their effectiveness in modulating perception of neighboring speech. Illustrative of this correspondence, Holt (1999) examined the influence that a preceding vowel (/i/ as in 'heat' or /u/ as in 'hoot') has upon listeners' identification of consonants varying perceptually across a syllable series ranging from /ba/ to /da/. Members of the /ba/ to /da/ series were distinguished by the onset frequency of the second formant (F2), with lower-frequency and higher-frequency endpoint stimuli perceived as good examples of /ba/ and /da/, respectively. The vowel pair /i/ and /u/ likewise were distinguished by F2 frequency, with /i/ having a higher F2 frequency than /u/.

Holt (1999) had listeners identify the /ba/–/da/ series members in the context of the two vowels. Listeners heard /i/ followed by a 50-ms silent interval preceding /ba/–/da/ series stimuli. The task was to identify the second syllable as 'ba' or 'da'. The same listeners identified identical /ba/–/da/ series members preceded by /u/

and a 50-ms silent interval. Listeners exhibited a context effect in that they labeled the /ba/–/da/ syllables more often as /ba/ when /i/ preceded them and more often as /da/ when /u/ was the precursor.

To examine the generality of the context effect exerted by the vowels, Holt created non-speech caricatures of the vowels by synthesizing tone glides that mimicked the vowels' F2 frequencies. These glides roughly modeled the essential acoustic characteristic distinguishing /i/ from /u/ (their F2 center frequencies). Although the non-speech glides shared spectral qualities with the F2 of the vowels, they fell far short of perceptual or acoustic equivalence with their vowel complements.

Vowels possess rich harmonic structure with energy at each multiple of the fundamental frequency (f_0). The glides, in contrast, had energy only at the nominal F2 center frequency, with no fine harmonic structure and no energy mimicking F1 or F3. The formant transitions of speech stimuli are not much like frequency-modulated tones comprising the glides because component frequencies of speech do not vary (when f_0 is constant, as it is for these stimuli). Instead, relative amplitudes of harmonics change with resulting changes in the shape of the spectral envelope. Despite these considerable acoustic differences, however, the very simple tone glides capture some of the putatively important spectral energy in the region of /i/ and /u/ F2.

Using these glides as context stimuli, Holt (1999) created non-speech/speech hybrid stimuli in which a glide preceded members of the /ba/–/da/ speech series described above. Again, 50 ms of silence separated context and target. When listeners identified the consonants in the context of these glides, their pattern of consonant identification was very similar to that observed for vowel context stimuli. Listeners more often labeled consonants as /ba/ when a glide modeling /i/ served as preceding context and more often as /da/ when the glide mimicking /u/ preceded the consonants. Neighboring non-speech sounds shifted listeners' identification of speech in the same direction as the vowels they mimicked.

In the same vein, Lotto and Kluender (1998) examined context effects within the /al/, /ar/, /ga/, /da/ stimulus paradigm of Mann (1980) described above. Acoustically, these syllables are distinguished primarily by third formant (F3) characteristics. Members of a /ga/–/da/ series can be distinguished by their F3 onset frequency, with /ga/ having a lower F3 onset than /da/ and intermediate series members varying stepwise between these two extremes. Likewise, the higher-frequency F3 offset of /al/ differentiates it from /ar/. As noted, the perception of a series of syllables as /da/ or /ga/ can be modulated by the presence of preceding context; /ga/–/da/ series syllables are more often perceived as /ga/ when preceded by /al/ and as /da/ when preceded by /ar/.

Lotto and Kluender (1998) created sine-wave tone glides modeling the F3 transitions of the /al/ and /ar/ stimuli. Thus, like the glide precursors of Holt (1999), these stimuli shared some of the same spectral energy as the speech sounds that they modeled. However, they were not perceived as speech. When these sounds preceded members of the phonetic series varying from /ga/ to /da/, they produced a pattern of /ga/–/da/ identification responses very similar to that elicited by synthetic and naturally produced speech tokens of /al/ and /ar/. In an additional experiment, Lotto and Kluender found that simple steady-state sine-wave tones situated at the final offset frequency of F3 for /al/ and /ar/ produced the same pattern of results.

Thus, this first line of evidence suggests that spectral content predicts the context-dependent shifts observed in phonetic context effects. A second line of evidence indicates that general auditory interactions of spectral characteristics may underlie the phonetic context effect. Lotto et al. (1997) trained Japanese quails (*Coturnix coturnix japonica*) to peck a lighted key in response to endpoints of the same /ga/–/da/ series as used by Lotto and Kluender (1998). Once quails reached a training criterion, novel intermediate stimuli drawn from the /ga/–/da/ series were interspersed in quails' daily training sessions. On these novel trials, /al/ or /ar/ preceded the /ga/–/da/ syllables. Birds trained to peck to /ga/ pecked most vigorously to novel intermediate members of the /ga/–/da/ series that were preceded by /al/. Correspondingly, /da/-positive quails pecked most robustly when novel stimuli were preceded by /ar/. Thus, Japanese quails exhibited shifts in pecking behavior contingent upon preceding context and their shifts were analogous to human shifts in consonant identification. The quails had no previous experience with speech so their behavior cannot be explained on the basis of learned covariance of acoustic attributes of speech in context or upon existing phonetic categories. Quails' behavior appears to be guided by the spectral information in the speech signal.

1.2. Spectral contrast

What remained constant between conditions in the experiments outlined above was some of the spectral content of the context stimuli. Tones or tone glides modeling the spectral characteristics of speech signals influenced neighboring speech in a manner comparable to the speech they mimicked. These data suggest that spectral content may be the critical variable in explaining phonetic context effects.

For all the examples above, context effects in speech perception can be described as contrastive. Fig. 1 depicts pseudospectrograms of the speech stimuli utilized by Holt (1999). The left panel illustrates the vowel /i/ preceding consonant–vowel syllables ranging perceptually from /ba/ (lowest F2 frequency) to /da/ (highest F2 frequency). Referring to Fig. 1, one can see that /i/ is distinguished from /u/ by F2 frequency. Holt (1999) found that listeners are more likely to report hearing /da/ following /u/. Thus, a precursor with *low*-frequency energy /u/ begets more responses corresponding to the consonant with *higher*-frequency spectral composition /da/. Likewise, listeners report more /ba/ responses (lower-frequency F2) following /i/ (higher-frequency F2).

Referring to Fig. 2, it can be seen that this descriptive pattern holds also for effects of /al/ and /ar/ on perception of following /ga/ or /da/. When the frequency composition of the offset is higher (F3 for /al/) listeners are more likely to report hearing /ga/ (lower F3 onset frequency) and vice versa (Lotto and Kluender, 1998). Overall, these results can be described as ‘spectrally contrastive’. A general neural mechanism or set of neural mechanisms that promotes spectral contrast may be responsible for producing the ubiquitous phonetic context effects observed in speech perception.[1]

It is important to note that although spectral contrast is a useful description of these data, the effects might be referred to more accurately as a form of *amplitude* contrast because speech formants are composed of multiple harmonics of the fundamental frequency. As such, shifts in ‘formant frequency’ are truly more a result of a shift in the relative amplitude of adjacent harmonics than an actual shift in frequency. This understanding may have important implications for psychophysical experiments aimed at outlining the bounds of contrast in speech perception. However, because of its usefulness as a descriptive heuristic, the use of spectral contrast is maintained here.

Examining perception across perceptual modalities, contrast emerges as an important mechanism for exaggerating differences between neighboring objects and events. Many well-known examples of perceptual contrast are visual. Enhancement of edges produced by lateral inhibition (Hartline and Ratliff, 1954), lightness judgments (Koffka, 1935) and judgment of line orientation (Gibson, 1933) each exhibits contrastive characteristics. Contrastive context effects in behavior are as varied as tempo of behavior (Cathcart and Dawson, 1928), weight lifting (Guilford and Park, 1931) and haptic perception (von Bekesy, 1967). In fact, mechanisms of contrast exist for every perceptual modality (von Bekesy, 1967; Warren, 1985). Across domains, contrast is a familiar outcome of mechanisms that serve to exaggerate change in the physical stimulus and to maintain an optimal dynamic range. Perceptual contrast, in this case spectral contrast, appears also to play an important role in perception of speech.

Contrast in itself, however, is a designation that does not implicate any specific perceptual mechanism(s). However, alluring the broad concept of contrast may be, its ubiquity betrays the fact that it falls short as a rigorous explanation. Ultimately, careful physiological investigation will be necessary to fully understand how phonetic context effects may originate in the auditory system. Psychophysical studies can help to determine the boundaries of the effect such that physiological investigation may proceed informed by basic data. The experiments described here begin to delineate these characteristics.

1.3. Psychoacoustic similarities

As described above, the spectral composition of a preceding auditory stimulus shifts perception of a following stimulus such that frequencies with less energy in the precursor are enhanced relative to frequencies with strong amplitudes in the spectral makeup of the precursor. Preceding speech contexts containing predominantly high-frequency spectral composition, for example, generate more identification responses for following speech sounds with low-frequency spectral patterns.

At least one class of psychoacoustic findings, known collectively as auditory enhancement (Green et al., 1959; Viemeister, 1980; Viemeister and Bacon, 1982) bears note for its similarity to the context effects described here. Auditory enhancement refers, generally, to the observation that if one member of a set of equal-amplitude harmonics is omitted from a harmonic series in an initial stimulus presentation and then introduced in a subsequent presentation, it stands out perceptually from the rest of the set members. Viemeister and Bacon (1982) used forward masking to demonstrate that auditory enhancement can change the perceived amplitude of a particular harmonic. In their study, they measured the forward masking of a 30-ms, 2-kHz probe produced by a masker comprised of a complete harmonic complex (fundamental frequency 200 Hz, highest component 4 kHz) with equal amplitude for each component. In one condition of the experiment, an adaptor stimulus consisting of an identical harmonic complex minus the 2-kHz component preceded the masker. Despite the fact that the harmonics of the masker were equal in amplitude, listeners reported hearing a distinct 2-kHz tone when the adaptor preceded it. That is, the 2-kHz component perceptually ‘popped out’ of the complete harmonic masker

---

[1] It should be noted that, in addition to the spectral differences in F3, there are also distinctions in F2. These differences also support a contrastive influence of context.

after presentation of the adaptor stimulus that lacked the 2-kHz component.

Viemeister and Bacon (1982) quantified this exhibit of auditory enhancement by demonstrating that such a masker 'enhanced' by a preceding adaptor produces more forward masking of a subsequent 2-kHz probe tone. In fact, listeners' thresholds for detecting a 2-kHz probe were an average of 7.9 dB higher when the masker was 'enhanced' by a preceding adaptor lacking the 2-kHz component. Viemeister and Bacon also measured the recovery functions for enhanced forward masking, finding functions that mirror the change produced by increasing the intensity of a single tone masker. Rather than a physical increase in intensity, however, Viemeister and Bacon were able to boost the *perceived* amplitude of the component in the masker and, in turn, produce greater forward masking of the probe.

Summerfield and colleagues (Summerfield et al., 1984, 1987) have related auditory enhancement to speech perception. Specifically, when a uniform harmonic spectrum composed of equal-amplitude harmonics is preceded by a spectrum complementary to a particular vowel (with troughs of low-amplitude components replacing formants), listeners report hearing a vowel during presentation of the uniform spectrum (Summerfield et al., 1984). This observation is in line with the more general case of auditory enhancement offered above in that harmonic frequencies absent in the precursor stimulus (the 'troughs') are enhanced perceptually in the subsequent stimulus. Listeners report hearing an illusory vowel that corresponds to a signal with formants positioned at the troughs of the precursor despite the fact that the actual signal is a uniform harmonic spectrum. Spectral composition of a preceding auditory stimulus shifts perception of a following stimulus such that frequencies absent in the precursor are enhanced relative to frequencies represented in the spectral makeup of the precursor.

It has been suggested that spectral contrast in speech may be a special case of auditory enhancement or at least that the mechanisms responsible for the two processes are the same (Holt et al., 2000a,b). In particular, there has been suggestion that neural adaptation plays an important role for each class of effects. Neural adaptation or adaptation of suppression may serve to enhance changes in spectral regions where previously there had been relatively little energy. Although these alternatives, so far, have been explored only in service of explaining auditory enhancement effects (Palmer et al., 1995), they may also play a role in resolving context dependencies arising from fluent speech. Delgutte and his colleagues (Delgutte, 1996; Delgutte et al., 1996) have established a case for a broad role of adaptation in perception of speech, noting that adaptation may enhance spectral contrast between sequential speech segments. This enhancement is proposed to arise because neurons adapted by stimulus components spectrally close to their preferred (characteristic) frequency are relatively less responsive to subsequent energy at that frequency whereas components not present in a prior stimulus are encoded by more responsive unadapted neurons.

Although the similarities of auditory enhancement and spectral contrast are intriguing, whether the two arise from the same mechanisms remains to be determined. The data collected thus far on spectral contrast do not let us choose among these proposals (or any other possibilities). In order to move these proposals beyond mere conjecture, basic data that characterize spectral contrast in speech perception are necessary. Although there have been some attempts to examine these mechanisms directly through neural recordings at the auditory-nerve and cochlear-nucleus levels (Holt and Rhode, 2000; Holt et al., 2000b) results have proven inconclusive. One of the aims of the present experiments is to provide more perceptual data to generate hypotheses about the mechanisms underlying spectral contrast in speech and their place in the auditory pathway.

In an effort to define some of the characteristics of phonetic context effects, the first experiments reported here examine whether the influence of a preceding stimulus on speech perception is a phenomenon that requires context and target stimulus to be presented to the same ear or whether it maintains for dichotic presentation. The next pair of experiments explores the time course over which a precursor influences speech perception. From these studies, clues as to the level of the auditory system at which these effects are occurring can be garnered. Furthermore, these findings may be compared to what is known of auditory enhancement in service of discovering whether the two originate from common mechanisms.

2. Experiments 1a and 1b

Auditory enhancement appears to be a strictly monaural effect. Summerfield and Assmann (1989), for example, failed to find effects of a precursor stimulus in their vowel experiments when the precursor was presented to the contralateral ear. Given that auditory enhancement appears to require that context and target be presented to the same ear, one means to examine the relationship between auditory enhancement and spectral contrast is to assess whether spectral contrast is likewise constrained. Experiments 1a and 1b undertake this manipulation using the phonetic context effects examined by Holt (1999) and Lotto and Kluender (1998).

Further, in comparing diotic (context and target to each ear) versus dichotic (context to contralateral ear of target), it is also possible to examine coarsely the level of auditory processing at which phonetic context effects occur. In traditional psychophysics, a distinction is often drawn between low-level peripheral processes (presumed to operate in the cochlea or at the level of the auditory nerve) and more central mechanisms (beginning at the brainstem in the cochlear nucleus and extending to cortical levels) of auditory processing. One method that can be used to tease apart these levels of processing is manipulation of stimulus presentation such that a context stimulus and a target stimulus are presented to the same or to different ears (cf., Lotto and Kluender, 1996).

There is no conduit for direct information processing between ears at the level of the cochlea.[2] Inter-aural processing begins at the level of the cochlear nuclei in the brainstem. Extrinsic sources of input to cochlear neurons lie in the contralateral cochlear nucleus and more central processing stations such as the superior olivary complex, the nuclei of the lateral lemniscus, and the inferior colliculus (for a review, see Cant, 1992). Thus, if information presented to one ear has an influence of context upon that presented to the opposite ear, it is suggestive that more central, rather than peripheral, levels of auditory processing are involved. If, however, very low-level mechanisms such as cochlear masking or auditory-nerve adaptation account for phonetic context effects then the effects should be obtained only when both the context and the target syllables are presented diotically.[3] For dichotic presentation, an effect of context upon phonetic categorization should not be evident.

3. Method

3.1. Listeners

Thirty-four and sixteen undergraduates participated in Experiments 1a and 1b, respectively. All listeners reported normal hearing and were native English speakers. No listeners from Experiment 1a participated in Experiment 1b. Experiments were approved by an Institutional Review Board and were conducted in accord with the mandates of the Declaration of Helsinki.

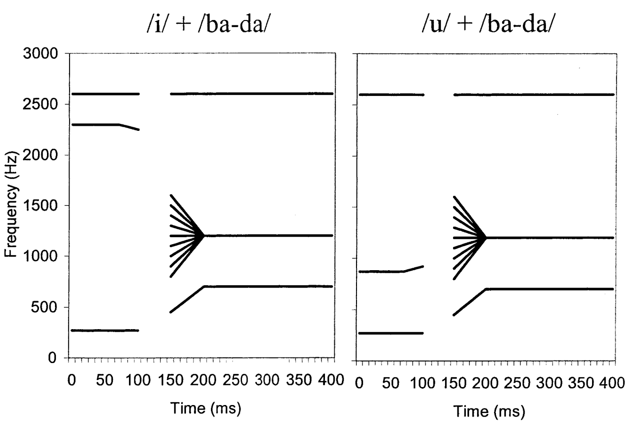

Fig. 1. Pseudospectrograms of Experiments 1a and 2a stimuli. Each of the nine possible second formant (F2) transitions is depicted on a single figure, illustrating the entire /ba/–/da/ series. The panel on the left represents /ba/–/da/ series members preceded by /i/. In the right panel, /u/ precedes the same stimuli. Note that the spectral characteristic distinguishing both target series members and context vowels is F2.

3.2. Stimuli

3.2.1. Experiment 1a

The stimulus set for Experiment 1a consisted of a nine-member series of CV syllables varying acoustically in F2 onset frequency and varying perceptually from /ba/ to /da/. The endpoints of this synthetic speech series were modeled after isolated productions of a male speaker. Across series members, F2 onset frequency varied from 800 to 1600 Hz in 100-Hz steps. Following onset, F2 frequency changed linearly over the next 50 ms to 1200 Hz, where it remained steady-state for the remainder of the 250 ms overall duration. All other synthesis characteristics were held constant across series members. First formant (F1) frequency increased linearly from 450 to 700 Hz over the first 50 ms. Thereafter, F1 frequency remained 700 Hz. The frequency of F3 was 2600 Hz and the fundamental frequency (f_0) was 120 Hz throughout the entire stimulus duration.

One of two synthesized speech stimuli preceded /ba/–/da/ series members. These vowels, /i/ and /u/, had acoustic characteristics based on productions of the same male after whom /ba/–/da/ syllables had been modeled. For both /i/ and /u/, F1 frequency was 270 Hz over the entire 100 ms duration. F2 frequency for /i/ was 2300 Hz over the first 70 ms after which it decreased linearly over 30 ms to 2250 Hz. For /u/, F2 frequency was 870 Hz for 70 ms and increased linearly over the last 30 ms to 920 Hz. F3 frequencies were

---

[2] This is not entirely correct because efferent neurons of the olivocochlear bundle do synapse on hair cells in the cochlea (Ryugo, 1992). However, theories of diotic (between ear) processing typically suggest that should such efferent synapses play a role in between-ear interactions, the origin of influence is likely to be more central stations of the auditory system.

[3] Of course, one also would predict the effect to be present for monaural (single ear) presentation. This redundant condition is not tested here.

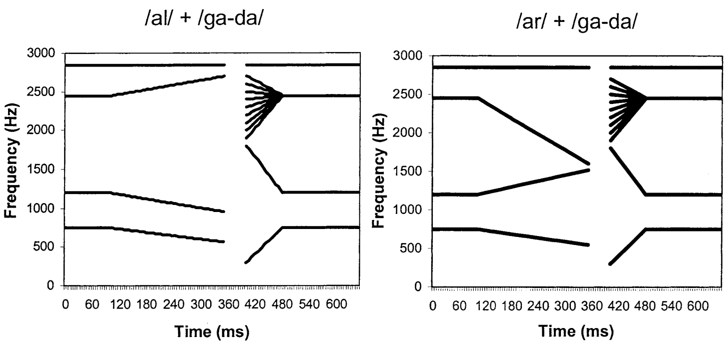

Fig. 2. Pseudospectrograms of Experiments 1b and 2b stimuli. Each of the 10 possible third formant (F3) transitions is depicted on a single figure, illustrating the entire /ga/–/da/ series. The panel on the left represents /ga/–/da/ series members preceded by /al/. In the right panel, /ar/ precedes the same stimuli. Note that the spectral characteristic distinguishing both target series members and contexts is F3.

steady-state at 2600 Hz for both vowels.[4] These stimuli are identical to those utilized by Holt (1999). Pseudospectrograms depicting caricatures of the regions of maximum energy as a function of frequency and time (i.e. the formants) for Experiment 1a stimuli are shown in Fig. 1. The left panel depicts /i/ preceding the nine /ba/–/da/ series members. The right panel illustrates /u/ before the same syllables.

Each stimulus was synthesized with 12-bit resolution using the cascade branch of the Klatt (1980) speech synthesizer implemented on a Pentium microcomputer at a sampling rate of 10 kHz. With a 50-ms silent interval separating vowel and consonant–vowel syllable, stimuli were converted from digital to analog (Ariel DSP-16) and low-pass filtered (4.8-kHz cutoff frequency, Frequency Devices #677). The RMS energy of each syllable was matched in amplitude prior to being amplified (Stewart HDA4) and presented over headphones (Beyer DT-100) at 70 dB SPL(A).

3.2.2. Experiment 1b

A 10-member series of synthetic speech varying acoustically in F3 onset frequency and varying perceptually from /ga/ to /da/ was created using the cascade branch of the Klatt (1980) synthesizer at a sampling rate of 20 kHz. For these stimuli, F3 onset frequency varied from 1800 to 2700 Hz in 100-Hz steps. From onset, F3 frequency changed linearly to a steady-state value of 2450 Hz across 80 ms. All other synthesis parameters were constant across series members. F1 frequency increased linearly from 300 to 750 Hz and F2 frequency declined from 1650 to 1200 Hz across 80 ms. The fourth formant (F4) had a steady-state value of 2850 Hz. f_0 was 110 Hz over the first 200 ms and decreased to 95 Hz over the last 50 ms. Total stimulus duration was 250 ms.

Two additional stimuli were created to serve as precursor context to the /ga/–/da/ series members. These stimuli were synthesized using the Klatt (1980) speech synthesizer and were identical to the /al/ and /ar/ precursor stimuli used by Lotto and Kluender (1998). The first 100 ms of all syllables were identical, with steady-state formant frequencies of 750, 1200, 2450, and 2850 Hz for the first four formants, respectively. After this 100-ms vowel, each stimulus had 150-ms linear formant transition. Offset frequencies for the first four formants of the /al/ syllable were 564, 956, 2700, and 2850 Hz, respectively. For the /ar/ syllable, formant frequencies changed linearly to arrive at 549, 1517, 1600, and 2850 Hz at stimulus offset. Each syllable was 250 ms long and had a constant f_0 of 110 Hz. Pseudospectrograms of /al/ and /ar/ preceding /ga/– /da/ series members are shown in Fig. 2. A 50-ms silent interval separated the precursor and /ga/–/da/ series members.

All stimuli were synthesized with 16-bit resolution at a 20-kHz sampling rate and stored on a computer disk following synthesis. Stimulus presentation was under the control of a microcomputer and Tucker Davis Technologies (TDT) hardware. Following D/A conversion (TDT, DD1), stimuli were low-pass filtered at a 9.8-kHz cutoff frequency (TDT, FTG2), amplified (TDT, PA4), and presented over headphones (Sennheiser HD 285) at 75 dB SPL(A).

3.3. Procedure

One to three listeners were tested concurrently in sound-attenuated booths during a single hour-long experimental session. On each trial, listeners heard two syllables (context followed by target) over headphones. The listeners' task was to identify the target syllable (as 'ba' or 'da' in Experiment 1a or as 'da' or 'ga' in Experiment 1b) by pressing a labeled button on an electronic response box.

---

[4] Though vowels used in auditory physiological investigations are often steady-state (with constant formant frequencies across the entire duration of the vowel) it is much more typical for vowels in fluent speech to have time-varying formant frequencies, or transitions.

The experiments were divided into two blocks corresponding to diotic and dichotic presentation. Each listener completed both blocks and order of block presentation was counterbalanced across listeners. In the dichotic block, context and target were presented to opposite ears, with ear of context presentation randomized across trials. In the diotic block, both context and target were presented to both ears on each trial. In each block, listeners responded to 10 repetitions of each of the context/target combinations. In all, the experiment lasted approximately 45 min.

3.4. Results

3.4.1. Experiment 1a

Identification responses averaged across listeners are presented in Fig. 3. These graphs illustrate the percent of time listeners identified /ba/–/da/ series members as ‘ba’ (based on a two-alternative forced-choice task) as a function of preceding vowel context. A shift in the identification curves indicates an influence of context on consonant labeling. Listeners’ labeling responses for the diotic listening condition (shown in Fig. 3a) demonstrate that preceding vowel has a significant influence on identification of a following /ba/–/da/ syllable ($F_{(1,33)} = 137.76$, $P < 0.0001$). Following /i/, consonants were more likely to be identified as ‘ba’. The same stimuli were labeled more often as ‘da’ when preceded by /u/. This finding replicates the results reported by Holt (1999).

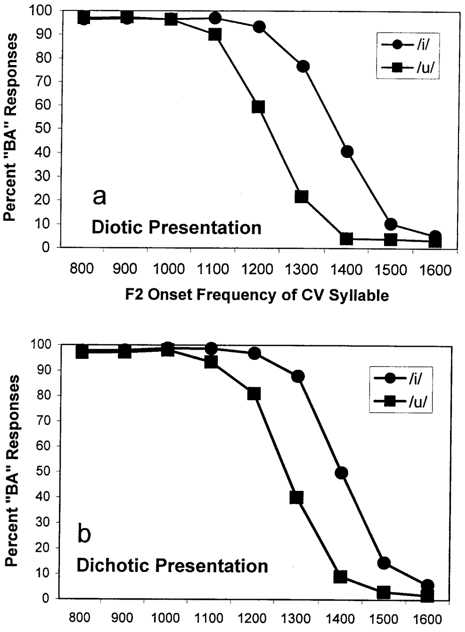

Fig. 3. Panel a illustrates the effect of context for diotic presentation of context and target syllables. Panel b shows results for dichotic presentation. Each graph depicts mean percent ‘ba’ identifications for syllables preceded by /i/ (circles), and /u/ (squares) as a function of /ba/–/da/ F2 onset frequency.

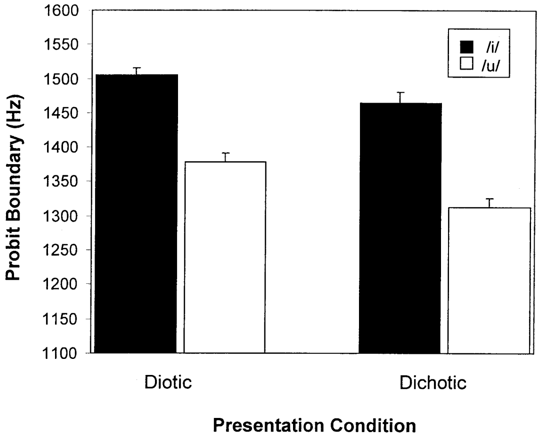

Fig. 4. Probit boundaries for identification of /ba/–/da/ syllables preceded by /i/ (dark bars) and /u/ (light bars) for dichotic and diotic presentation conditions. Taller bars (higher-frequency boundaries) indicate more ‘ba’ responses. A difference in bar height reflects an influence of preceding context on consonant identification.

Most germane to hypotheses of the present study, however, is listeners’ labeling behavior in the dichotic condition. These data are shown in Fig. 3b where listeners’ average percent ‘ba’ responses for this condition are plotted against /ba/–/da/ series-member F2 onset frequency. As is clear from the graph, preceding vowels maintain their influence on /ba/–/da/ identification in the same direction as observed for diotic presentation. Even when vowel and CV are presented to opposite ears, the preceding vowel influences phonetic categorization ($F_{(1,33)} = 367.47$, $P < 0.0001$).

Fig. 4 summarizes these results with probit analyses. Probit analysis is a class of non-linear models of estimation for discrete decision variables. In the present analysis, a cumulative normal curve was fit to the identification functions depicted in Fig. 3 by transforming the percent of ‘ba’ responses to z scores and finding the best fitting line through linear regression. The boundary was taken to be the F2 onset frequency corresponding to 50% (z score of 0) on this line. This analysis provides a means of assessing the identification shift elicited by preceding context. The height of the histogram bars represents the cross-over F2 onset frequency at which listeners’ identification boundary lies. Taller bars designate a greater proportion of ‘ba’ responses and a distinction in bar height between unfilled and solid bars is indicative of an effect of context. The solid bars correspond to identification of the /ba/–/da/ series members in the presence of a preceding /i/ whereas unfilled bars indicate identification in the context of /u/. The mean probit boundary difference for dichotically presented stimuli was slightly larger ($M = 151.6$ Hz) than that elicited by stimuli presented diotically ($M = 126.7$ Hz),

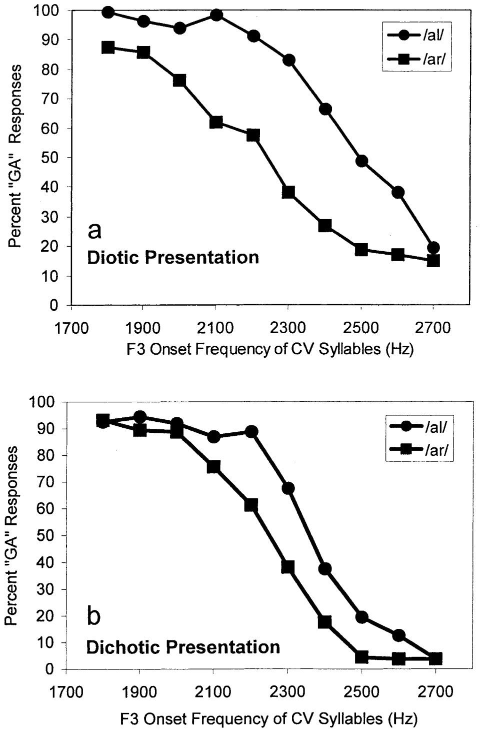

Fig. 5. Panel a illustrates the effect of context for diotic presentation of context and target syllables. Panel b shows results for dichotic presentation. Each graph depicts mean percent 'ga' identifications for syllables preceded by /al/ (circles), and /ar/ (squares) as a function of /ga/–/da/ F3 onset frequency.

although this difference was only marginally significant ($t_{(33)} = -1.48$, $P = 0.07$).

3.4.2. *Experiment 1b*

The results of Experiment 1b are presented in the same format as those of Experiment 1a in Figs. 5 and 6. Fig. 5 illustrates average identification curves in terms of percent 'ga' identifications as a function of F3 onset frequency with the preceding syllable (/al/ or /ar/) as parameters for diotic (panel a) and dichotic (panel b) conditions. As can be surmised from the figures, there is a substantial overall influence of preceding context across conditions ($F_{(1,15)} = 48.44$, $P < 0.0001$) and the context effect is significant in both diotic ($F_{(1,15)} = 30.41$, $P < 0.0001$) and dichotic ($F_{(1,15)} = 18.71$, $P < 0.0001$) conditions. Under each presentation mode, stimuli are more often identified by listeners as /ga/ when preceded by /al/.

However, whereas there was only a marginal difference in the magnitude of the context effect as a function of diotic versus dichotic presentation in Experiment 1a, in Experiment 1b presentation mode did have an influence on the size of the context effect. This can be seen most clearly from inspection of the probit boundaries presented in Fig. 6. The mean probit boundary difference for dichotically presented stimuli was significantly ($t_{(15)} = 2.50$, $P = 0.01$) smaller ($M = 130.04$ Hz) than that elicited by stimuli presented diotically ($M = 253.78$ Hz). Thus, it seems that, for this context effect, diotic presentation, perhaps by virtue of the fact that stimulus components may interact within the auditory periphery, enhances effects of context. However, the fact that the effect maintains across dichotic presentation rules out any account that relies solely upon peripheral neural mechanisms to explain phonetic context effects.

3.5. *Discussion*

The foremost finding of Experiments 1a and 1b is that the effect of preceding context upon consonant identification persists for both diotic and dichotic presentation. In general, with only a marginal trend toward significance, the stimuli of Experiment 1a appear to be a bit more resistant to the effects of diotic versus dichotic presentation. For Experiment 1b, the effect of context was larger for diotically presented stimuli. Overall, it appears that spectral contrast is *not* restricted to monaural interactions. These results are in contrast to the decidedly monaural character of auditory enhancement effects (Summerfield and Assmann, 1989). It appears that spectral contrast may be distinct from auditory enhancement; although, it is possible that these processes work in tandem to enhance spectral differences in speech sounds.

Considering that the effect remains with dichotic presentation, these data also suggest that low-level peripheral mechanisms such as a cochlear masking or auditory-nerve adaptation cannot be solely responsible for these shifts in perception. However, the results of Experiment 1b (for which the context effect was larger for diotically presented stimuli) suggest that there may be some contribution of monaural processes.

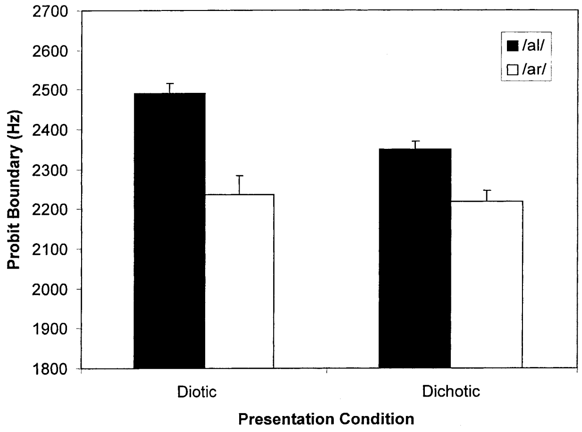

Fig. 6. Probit boundaries for identification of /ga/–/da/ syllables preceded by /al/ (dark bars) and /ar/ (light bars) for dichotic and diotic presentation conditions. Taller bars (higher-frequency boundaries) indicate more 'ga' responses. A difference in bar height reflects an influence of preceding context on consonant identification.

4. Experiments 2a and 2b

The second pair of experiments explores the time course of the two phonetic context effects introduced in Experiment 1 as a means of further examining the putative similarities of spectral contrast in speech perception to auditory enhancement.

Auditory enhancement has a characteristic time course. In their masking study, Viemeister and Bacon (1982) found no appreciable auditory enhancement beyond about 100 ms of silent gap. Experiments 2a and 2b examine the time course over which the effect of preceding context on consonant identification occurs. To this point, each of the stimulus sets used in the spectral contrast experiments described has employed a 50-ms silent interval between preceding context and consonant-series member. Observing the influence of inter-syllable silent interval duration allows examination of the time course over which context exerts an influence. This manipulation provides a coarse sense of the level of the auditory system at which these effects occur. One would expect more peripheral mechanisms to operate over a shorter time course; the influence of such mechanisms would presumably fade for longer inter-syllable silent intervals. Given that Experiment 1 suggested the processes involved are not solely peripheral, it is likely that the effect of preceding context on consonant identification will not dissipate in tens of milliseconds, but rather will gradually decline over a longer time course. This is expected because, as one observes effects of interactions at more central levels of the auditory system, there is a longer temporal window over which auditory events interact and influence one another (Popper and Fay, 1992).

5. Method

5.1. Listeners

Twenty-nine and seventeen native English-speaking undergraduate students participated in Experiments 2a and 2b, respectively. All listeners reported normal hearing. None of the listeners had participated in Experiment 1a or 1b. The experiment protocol was approved by an Institutional Review Board and experiments were conducted in accord with the mandates of the Declaration of Helsinki.

5.2. Stimuli

5.2.1. Experiment 2a

Context and target stimuli were identical to those used in Experiment 1a, but the inter-syllable silent interval varied. Inter-syllable silent intervals of 50, 100, 150, 200, and 400 ms separated vowel from /ba/–/da/ series members. All stimuli were presented diotically.

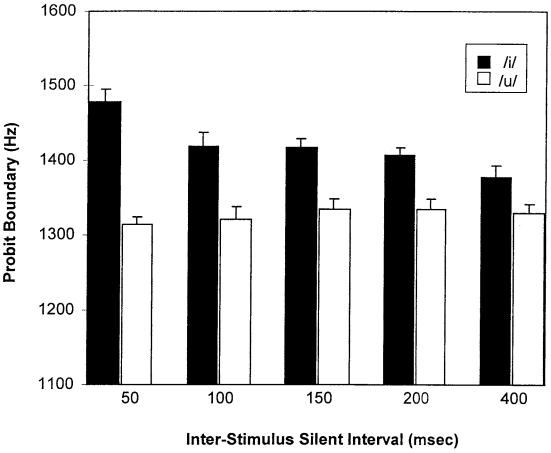

Fig. 7. Probit boundaries for identification of /ba/–/da/ syllables preceded by /i/ (dark bars) and /u/ (light bars) as a function of inter-stimulus silent interval. Taller bars (higher-frequency boundaries) indicate more 'ba' responses. A difference in bar height reflects an influence of preceding context on consonant identification. The error bars depict standard error.

5.2.2. Experiment 2b

The /ga/–/da/ series and the /al/ and /ar/ context stimuli used in Experiment 1b were used again in Experiment 2b. However, the silent interval separating context stimulus and consonant-series member varied such that there were six inter-stimulus silent intervals (25, 50, 100, 175, 275, and 400 ms).

In addition, the stimulus presentation details differed from Experiment 1b. Sampling rate of the stimuli was 10 kHz. Stimuli were converted from digital to analog (Ariel DSP-16) and low-pass filtered (4.8-kHz cutoff frequency, Frequency Devices #677). The RMS energy of each syllable was matched in amplitude prior to being amplified (Stewart HA4) and presented diotically over headphones (Beyer DT-100) at 70 dB SPL(A). Thus, the presentation details were matched for Experiments 2a and 2b.

5.3. Procedure

The methods were essentially like those of Experiment 1. However, each listener participated in five blocks (six for Experiment 2b) of trials corresponding to the inter-syllable silent intervals.[5] Within each block, order of stimulus presentation was randomized. As before, listeners' task was to identify the target (second)

---

[5] A follow-up study using a design mixed across inter-syllable silent interval produced the same qualitative results as this blocked design.

syllable using an electronic response box. The order of blocks was counterbalanced across listeners.

5.4. Results

5.4.1. Experiment 2a

Owing to the large number of identification curves generated from this experiment (two for each of five ISI levels) and considering that the question of interest is how increasing inter-stimulus silent intervals influence the magnitude of the context effect, the results are presented as probit boundaries. Individual identification boundaries for each of the listeners were determined for each vowel context (/i/ and /u/) of each inter-syllable silent interval condition (50, 100, 150, 200, and 400 ms) using probit analysis. Results of probit analysis averaged across listeners are plotted in Fig. 7. The solid bars correspond to identification of the /ba/–/da/ series members in the presence of a preceding /i/ whereas unfilled bars indicate identification in the context of /u/. Taller bars designate a greater proportion of 'ba' responses and a distinction in bar height between unfilled and solid bars illustrates an effect of context. As shown by the plot, the size of the context effect decreases with greater inter-syllable intervals. Thus, the effect is sensitive to the temporal adjacency of neighboring stimuli. However, the context effect is quite robust and the decay with increasing inter-stimulus interval is rather gradual. In fact, planned *t*-tests of the probit boundaries for pairs of vowel contexts are statistically significant ($P < 0.01$) at all intervals, including 400 ms.

5.4.2. Experiment 2b

The results of Experiment 2b mirror the overall observations of Experiment 2a. Probit boundary analyses are presented in Fig. 8. As in Experiment 2a, the effect of preceding context lessens as context and consonant-series member are temporally separated by greater intervals of silence. However, the effect persists for quite long intervals. For each of the silent intervals less than 275 ms, there is a highly significant effect of preceding context on consonant identification ($P < 0.01$). At 400 ms, the effect of context is not significant ($P > 0.05$).

5.5. Discussion

Whereas Viemeister and Bacon (1982) found no appreciable auditory enhancement for their masking study beyond about 100 ms of silent gap, the present data indicate that spectral contrast persists for hundreds of milliseconds. Thus, the neural processes underlying spectral contrast and auditory enhancement appear to have distinct time courses.

Whether these results rule out purely peripheral mechanisms as underlying phonetic context effects is not fully clear from these data alone. The typical duration over which short-term peripheral adaptation exerts its influence is on the order of tens of milliseconds (e.g. Smith, 1977). However, there is also evidence that peripheral adaptation may persist for time constants of at least a second (Young and Sachs, 1973; Relkin and Doucet, 1991). Direct physiological investigation with stimuli similar to those investigated here will be important in deciding this issue, but it seems likely that central auditory mechanisms are playing some role.

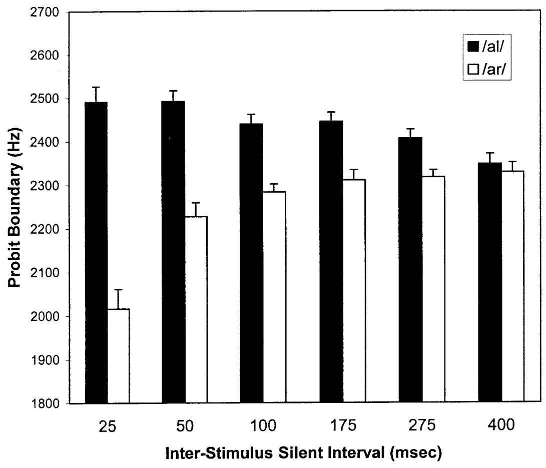

Fig. 8. Probit boundaries for identification of /ga/–/da/ syllables preceded by /al/ (dark bars) and /ar/ (light bars) as a function of inter-stimulus silent interval. Taller bars (higher-frequency boundaries) indicate more 'ga' responses. A difference in bar height reflects an influence of preceding context on consonant identification.

6. General discussion

Our goals for the present work were twofold. First, we aimed to present a review of recent data that phonetic context effects are a function of the spectral properties of context and target syllables. We further argued that spectral contrast serves as an effective descriptive framework within which to interpret these findings. However, spectral contrast, in and of itself, does not suggest any particular mechanism. Our second and most important goal was to provide behavioral data that begin to delineate the mechanism or mechanisms that underlie spectral contrast in speech perception. In particular, we aimed to examine the relationship between spectral contrast and a class of psychoacoustic effects known as auditory enhancement. On the surface, these two classes of findings appear to be quite similar and it is tempting to posit that the two may arise from common neural mechanisms.

The results of the current studies, however, seriously call into question this proposition. Whereas auditory enhancement has been shown to be a purely monaural phenomenon that does not appear when context sounds are presented to the ear contralateral to the target (Summerfield and Assmann, 1989), spectral contrast is evident both diotically and dichotically. In addition, the time course of spectral contrast is quite different from that observed in auditory enhancement experiments. Viemeister and Bacon (1982) found that auditory enhancement effects disappeared with silent gaps of 100 ms. The results of the current experiments indicate that spectral contrast effects persist for hundreds of milliseconds.

On the whole, the current observations strongly suggest that auditory enhancement and spectral contrast do not originate from identical auditory mechanisms. However, these findings do not rule out the possibility that the mechanisms underlying auditory enhancement and spectral contrast may work together in producing effects of contrast in audition. Holt and Lotto (2001) have argued that spectral contrast, as a descriptive tool, may not be so much a single mechanism as a general operating characteristic that captures a pervasive characteristic of perceptual systems originating from the necessity to segregate events and objects perceptually. As such, auditory enhancement and spectral contrast may originate from distinct mechanisms yet both serve the common purpose of promoting contrast.

The present results also strongly suggest that solely peripheral mechanisms do not give rise to spectral contrast, although they may play a facilitative role. Effects of precursor context are present for dichotic presentation of context precursor and target syllable. In addition, effects of the precursor stimulus persisted for longer than would typically be attributed to peripheral processing. Whereas monaural mechanisms may not be totally culpable for these effects may enhance spectral contrast.

These results provide a foundation of empirical work upon which investigation of the neural mechanisms might begin. Already, a number of recent auditory neurophysiological investigations have begun examining the processing of sound sequences by the central auditory system (cf., Brosch et al., 1999; Brosch and Schreiner, 2000) and a few studies have surveyed the effect of context upon speech sounds (cf., Mandava et al., 1995). The neural mechanisms uncovered by these investigations may play a role in encoding speech in a context-dependent manner. Further investigation of how adjacent speech sounds influence auditory processing will be important in defining the precise mechanisms that govern phonetic context effects in speech perception.

Previous neurophysiological work on auditory enhancement has focused on peripheral mechanisms at the auditory-nerve or cochlear-nucleus levels (Palmer et al., 1995). The data presented here introduce the possibility that more central stages of processing are influential in spectral contrast in speech perception. In agreement with this supposition, recent work examining chinchilla neural response to speech context stimuli has shown little evidence for spectral contrast at the level of the auditory nerve and ventral cochlear nucleus (Holt and Rhode, 2000; Holt et al., 2000b).

Most broadly, this work can be cast as a template for understanding speech perception. For years, integration of behavioral research with neurophysiological investigation has proven advantageous in understanding psychacoustic phenomena. Behavioral work with human listeners spurs hypotheses as to putative neural mechanisms that can then be tested directly in animal models. However, this approach has been much less commonly applied to phenomena of speech perception. This is likely due to the long-standing assumption that speech phenomena have their origins in specialized neural mechanisms dedicated to speech processes and present only in humans. However, data accumulating from animal behavioral experiments over the last several decades have provided substantive evidence that many of the phenomena of speech perception may have bases in general auditory function. From these experiments, we have learned that animals respond to speech categorically (Morse and Snowdon, 1975; Kuhl and Miller, 1975; 1978; Waters and Wilson, 1976), exhibit phonetic context effects (Dent et al., 1997; Lotto et al., 1997) and are sensitive to acoustic trading relations (Kluender and Lotto, 1994; Holt et al., 2001).

As such, there is a great deal to be gained from a template of research that encourages auditory physiologists to exploit the findings of human and animal behavioral work in speech perception to precipitate hypotheses as to the underlying mechanisms of audition that give rise to them. Communication among researchers who employ each of these methods can inspire physiological investigation of some of the more complex aspects of audition (such as context effects) and encourage advances in our understanding of the neural bases of speech perception.

Acknowledgements

This work was supported in part by a National Science Foundation Predoctoral Fellowship to L.L.H. Preparation of this paper was supported, in part, by funding from the National Organization for Hearing Research to L.L.H.

References

Borsky, S., Tuller, B., Shapiro, L.P., 1998. 'How to milk a coat': The effects of semantic and acoustic information on phoneme categorization. J. Acoust. Soc. Am. 103, 2670–2676.

Brosch, M., Schreiner, C.E., 2000. Sequence sensitivity of neurons in cat primary auditory cortex. Cereb. Cortex 10, 1155–1167.

Brosch, M., Schulz, A., Scheich, H., 1999. Processing of sound sequences in macaque auditory cortex: Response enhancement. J. Neurophysiol. 82, 1542–1559.

Cant, N.B., 1992. The Mammalian Auditory Pathway: Neuroanatomy. Springer, New York.

Cathcart, E.P., Dawson, S., 1928. Persistence: a characteristic of remembering. Br. J. Psychol. 18, 262–275.

Connine, C.M., 1987. Constraints on interactive processes in auditory word recognition: The role of sentence context. J. Mem. Lang. 26, 527–538.

Delgutte, B., 1996. The Handbook of Phonetic Sciences. Blackwell, Oxford, pp. 505–538.

Delgutte, B., Hammond, B.M., Kalluri, S., Litvak, L.M., Cariani, P.A., 1996. Neural encoding of temporal envelope and temporal interactions in speech. In: Ainsworth, W., Greenberg, S. (Eds.), Proceedings of the ESCA Workshop on the Auditory Basis of Speech Perception, Vol. 1. Keele University, pp. 1–11.

Dent, M.L., Brittan-Powell, E.F., Dooling, R.J., Pierce, A., 1997. Perception of synthetic /ba/–/wa/ speech continum by budgerigars (*Melopsittacus undulatus*). J. Acoust. Soc. Am. 102, 1891–1897.

Fowler, C.A., 1996. Listeners do hear sounds, not tongues. J. Acoust. Soc. Am. 99, 1730–1741.

Fowler, C.A., Best, C.T., McRoberts, G.W., 1990. Young infants' perception of liquid coarticulatory influences on following stop consonants. Percept. Psychophys. 48, 559–570.

Fowler, C.A., Brown, J.M., Mann, V.A., 2001. Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans. J. Exp. Psychol. Hum. Percept. Perform. 877–888.

Ganong, W.F., 1980. Phonetic categorization in auditory word processing. J. Exp. Psychol. Hum. Percept. Perform. 6, 110–125.

Gibson, J.J., 1933. Adaptation, after-effect and contrast in the perception of curved lines. J. Exp. Psychol. 16, 1–31.

Green, D.M., McKay, M.J., Licklider, J.C.R., 1959. Detection of a pulsed sinusoid in noise as a function of frequency. J. Acoust. Soc. Am. 31, 1446–1452.

Guilford, J.P., Park, D.G., 1931. The effect of interpolated weights upon comparative judgments. Am. J. Psychol. 43, 589–599.

Hartline, H.K., Ratliff, F., 1954. Spatial summation of inhibitory influences in the eye of *Limulus*. Science 120, 781.

Holt, L.L., 1999. Auditory constraints on speech perception: An examination of spectral contrast. Dissertation, Department of Psychology University of Wisconsin, Madison, WI.

Holt, L.L., Lotto, A.J., 2001. Auditory context effects for nonequivalent sources. J. Acoust. Soc. Am. 109, 2311.

Holt, L.L., Rhode, W.S., 2000. Examining context-dependent speech perception in the chinchilla cochlear nucleus, 2000. Midwinter Meeting of Association for Research in Otolaryngology, St. Petersberg Beach, FL.

Holt, L.L., Lotto, A.J., Kluender, K.R., 2000a. Neighboring spectral content influences vowel identification. J. Acoust. Soc. Am. 108, 710–722.

Holt, L.L., Ventura, V., Rhode, W.R., Behesta, S., Rinaldo, A., 2000b. Context-dependent neural coding in the chinchilla cochlear nucleus. J. Acoust. Soc. Am. 108, 2641.

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? J. Acoust. Soc. Am. 109, 764–774.

Klatt, D.H., 1980. Software for a cascade/parallel formant synthesizer. J. Acoust. Soc. Am. 67, 971–990.

Kluender, K.R., Lotto, A.J., 1994. Effects of first formant onset frequency on [-voice] judgments result from general auditory processes not specific to humans. J. Acoust. Soc. Am. 95, 1044–1052.

Koffka, K., 1935. Principles of Gestalt Psychology. Harcourt Brace, New York.

Kuhl, P.K., Miller, J.D., 1975. Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. Science 190, 69–72.

Kuhl, P.K., Miller, J.D., 1978. Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. J. Acoust. Soc. Am. 63, 905–917.

Lindblom, B.E.F., Studdert-Kennedy, M., 1967. On the role of formant transitions in vowel recognition. J. Acoust. Soc. Am. 42, 830–843.

Lotto, A.J., Kluender, K.R., 1996. Effects of amplitude on voicing contrast may not be explained by VIIIth nerve synchrony capture. J. Acoust. Soc. Am. 99, 2591–2603.

Lotto, A.J., Kluender, K.R., 1998. General auditory processes may account for the effect of preceding liquid on perception of place of articulation. Percept. Psychophys. 60, 602–619.

Lotto, A.J., Kluender, K.R., Holt, L.L., 1997. Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). J. Acoust. Soc. Am. 102, 1134–1140.

Mandava, P., Rupert, A.L., Moushegian, G., 1995. Vowel and vowel sequence processing by cochlear nucleus neurons. Hear. Res. 87, 114–131.

Mann, V.A., 1980. Influence of preceding liquid on stop-consonant perception. Percept. Psychophys. 28, 407–412.

Mann, V., 1986. Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of English 'l' and 'r'. Cognition 24, 169–196.

Mann, V.A., Repp, B.H., 1980. Influence of vocalic context on perception of the [sh]-[s] distinction. Percept. Psychophys. 28, 213–228.

Mann, V.A., Repp, B.H., 1981. Influence of preceding fricative on stop consonant perception. J. Acoust. Soc. Am. 69, 548–558.

Morse, P.A., Snowdon, C.T., 1975. An investigation of categorical speech discrimination by Rhesus monkeys. Percept. Psychophys. 17, 9–16.

Palmer, A.R., Summerfield, Q., Fantini, D.A., 1995. Responses of auditory-nerve fibers to stimuli producing psychophysical enhancement. J. Acoust. Soc. Am. 97, 1786–1799.

Popper, A.N., Fay, R.R., 1992. The Mammalian Auditory Pathway: Neurophysiology. Springer, New York.

Relkin, E.M., Doucet, J.R., 1991. Recovery from prior stimulation. I: Relationship to spontaneous firing rates of primary auditory neurons. Hear. Res. 55, 215–222.

Repp, B.H., 1982. Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. Psychol. Bull. 92, 81–110.

Ryugo, D.K., 1992. The Auditory Nerve: Peripheral Innervation. Cell Body Morphology, and Central Projections. In: Webster, D.B., Popper, A.N., Fay, R.R. (Eds.), The Mammalian Auditory Pathway Neuroanatomy, Vol. 1. Springer, New York, pp. 23–65.

Smith, R.L., 1977. Short-term adaptation in single auditory nerve fibers: Some post-stimulatory effects. J. Neurophysiol. 40, 1098–1112.

Summerfield, Q., Assmann, P.F., 1989. Auditory enhancement and the perception of concurrent vowels. Percept. Psychophys. 45, 529–536.

Summerfield, Q., Haggard, M., Foster, J., Gray, S., 1984. Perceiving vowels from uniform spectra: Phonetic exploration of an auditory aftereffect. Percept. Psychophys. 35, 203–213.

Summerfield, Q., Sidwell, A., Nelson, T., 1987. Auditory enhancement of changes in spectral amplitude. J. Acoust. Soc. Am. 81, 700–707.

Viemeister, N.F., 1980. Adaptation of masking, in Psychophysical. In:

van den Brink, G., Bilsen, F.S. (Eds.), Physiological and Behavioral Studies in Hearing, Delft U.P., Delft, pp. 190–198.

Viemeister, N.F., Bacon, S.P., 1982. Forward masking by enhanced components in harmonic complexes. J. Acoust. Soc. Am. 71, 1502–1507.

von Bekesy, G., 1967. Sensory Perception. Princeton University Press, Princeton, NJ.

Warren, R.M., 1985. Criterion shift rule and perceptual homeostasis. Psychol. Rev. 92, 574–584.

Waters, R.S., Wilson, W.A., 1976. Speech perception by Rhesus monkeys: The voicing distiction in synthesized labial and velar stop consonants. Percept. Psychophys. 19, 285–289.

Young, E., Sachs, M.B., 1973. Recovery from sound exposure in auditory-nerve fibers. J. Acoust. Soc. Am. 54, 1535–1543.