Phonetica 2000;57:170–180

Received: November 4, 1999
Accepted: February 22, 2000

General Auditory Processes Contribute to Perceptual Accommodation of Coarticulation

Lori L. Holt[a] Keith R. Kluender[b]

[a]Department of Psychology and The Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, Pa., and [b]Department of Psychology, University of Wisconsin, Madison, Wisc., USA

Abstract

The ability of listeners to recover speech information, despite dramatic articulatory and acoustic assimilation between adjacent speech sounds, is remarkable and central to understanding perception of fluent speech. Lindblom [J. Acoust. Soc. Am. *35*: 1773–1781, 1963] shared with the field some of the most compelling early descriptions of the acoustic effects of coarticulation, and with Studdert-Kennedy [J. Acoust. Soc. Am. *42*: 830–843, 1967] provided perceptual data that remain central to theorization about processes for perceiving coarticulated speech. In years that followed, hypotheses by others that have intended to explain the ability to maintain perceptual constancy despite coarticulation have relied in some way or another upon relatively detailed reference to speech articulation. A number of new findings are reported here that suggest general auditory processes, not at all specific to speech, contribute significantly to perceptual accommodation of coarticulation. Studies using nonspeech flanking energy, capturing minimal spectral aspects of speech, suggest simple processes (that can be portrayed as contrastive) serve to 'undo' assimilative effects of coarticulation. Data from nonhuman animal subjects suggest broad generality of these processes. At a more mechanistic explanatory level, psychoacoustic and neurophysiological data suggestive of underlying sensory and neural mechanisms are presented. Lindblom and Studdert-Kennedy's early hypotheses about the potential for such mechanisms are revived and supported.

Introduction

'Lack of invariance' between phonemes and attributes of the acoustic signal poses a central dilemma in understanding the nature of speech perception. The problem is that there seem to exist few (or no) unitary attributes in the acoustic signal that uniquely specify particular phonemes. The primary culprit of the acoustic variability across productions of a given phoneme is coarticulation – the spatial and temporal overlap of adjacent articulatory activities

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0031–8388/00/0574–0170 $17.50/0

Accessible online at:
www.karger.com/journals/pho

Lori L. Holt
Department of Psychology, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213 (USA)
Tel. +1 412 268-4964, Fax +1 412 268-2798
E-Mail lholt@andrew.cmu.edu

The influence of coarticulation is reflected in the acoustic signal by severe context dependence. Lindblom [1963] documented one of the earlier and most influential demonstrations of context dependence in speech acoustics in an investigation of the influence of consonant context on vowel spectra. The acoustic patterns of vowels are affected by coarticulation with adjacent consonants such that formant frequencies may diverge considerably from patterns of isolated vowels. Lindblom [1963] found that second formant (F_2) frequencies of vowels in [bʊb] and [bɪb] context were significantly lower in frequency than productions of the same vowels ([ʊ] and [ɪ]) in isolation, whereas vowel F_2 frequencies were higher in productions of [dʊd] and [dɪd]. In each context, F_2 frequency of the vowels was shifted away from the F_2 frequency observed for the same vowel in isolation and toward the frequency of F_2 for adjacent consonants (lower for [bVb] and higher for [dVd]).

This example of acoustic context dependence consequent to coarticulation is typical of many observations of acoustic effects of coarticulation that have followed Lindblom's early observations. In general, coarticulation is the compromise between where articulators have been and where they are going in connected speech [Stevens and House, 1963; Öhman, 1966; Mann, 1980]. This is reflected as assimilation in the acoustic patterns of speech. As a result of coarticulation, the acoustic realization of a particular phoneme can vary substantially across phonetic context, always in the direction of neighboring sounds. Such effects can be dramatic. Lindblom [1963], for example, reported F_2 frequency deviations as great as 70% for a single vowel articulated across several consonant contexts.

The ability of listeners to recover speech information despite this dramatic context dependence is central to understanding speech perception. In the laboratory, this capacity has been indexed by observing perception of otherwise identical speech tokens in different phonetic contexts. For example, Lindblom and Studdert-Kennedy [1967] examined perception of consonant-vowel-consonant (CVC) syllables using three synthetic vowel series. The first consisted of a steady-state series varying perceptually from [ʊ] to [ɪ] via manipulation of F_2 frequency. The second and third series were comprised of these same vowels, identical in midpoint formant frequencies, embedded in [wVw] and [jVj] contexts[1]. Recall that, relative to F_2 frequencies of isolated vowels, Lindblom [1963] observed lower F_2 formant frequencies for vowels produced in [bVb] context and higher vowel F_2s in [dVd] context.

Complementary to Lindblom's [1963] measures of production, listeners perceived more syllables as [ɪ] (with lower F_2 frequencies) in the [wVw] context, and more as [ʊ] (with higher F_2 frequencies) in the [jVj] context [Lindblom and Studdert-Kennedy, 1967]. Neighboring consonant context influenced vowel perception in a fashion complementary to the assimilative effects of coarticulation. These oft-cited findings have since been extended by Nearey [1989], who found the same pattern of results for [bVb] and [dVd] syllables and vowel sounds ranging from [ɒ] to [ʌ] and from [ʌ] to [ɛ].

Examples of perceptual accommodation of coarticulation have been documented across a wide variety of other phonetic contexts including identification of stop consonants and fricatives [e.g. Mann, 1980; Mann and Repp, 1980, 1981; Repp, 1983]. In

[1] These stimuli were very similar to those Lindblom [1963] had explored earlier in that labial consonant contexts ([bVb] and [wVw]) have lower F_2 onsets than do palatoalveolar contexts (i.e. [dVd] and [jVj]). Acoustically, the main distinction between stops [b, d] and semivowels [w, j] lies in the rate of change of formant transitions.

these studies, too, perception adjusts in a direction opposite that of coarticulatory assimilation.

How is this accomplished? Lindblom and Studdert-Kennedy [1967] offered several potential theoretical interpretations of their results. They noted that their data were congenital to articulation-based theoretical perspectives like Motor Theory [Liberman et al., 1957] and Analysis-by-Synthesis [Stevens and House, 1963; Stevens and Halle, 1967]. In most later examinations, authors have interpreted their data more categorically as evidence that context effects in speech perception originate in properties of speech distinct from its auditory characteristics [e.g. Mann, 1980; Repp, 1982; Williams, 1986; Fowler et al., 1990]. The theoretical thrust of these accounts, whether embodied by modular processes [Liberman et al., 1957; Liberman and Mattingly, 1985], Direct Realist emphasis on distal events [Fowler, 1986, 1996], or reference to tacit knowledge [Repp, 1982, 1983], suggests mechanisms of speech perception intimately linked with speech production.

Evidence for General Auditory Processes

Human Listeners

Lindblom and Studdert-Kennedy [1967] devoted most of their discussion to potential explanations in terms of general auditory processes. A number of new findings described below suggest that this emphasis was well-warranted and prescient. General auditory mechanisms appear to contribute significantly to perceptual accommodation of coarticulation. The first line of evidence involves close correspondence between effects of neighboring speech and nonspeech context upon speech perception. These studies have used nonspeech flanking energy, capturing minimal spectral aspects of speech, to probe the relative specificity of context effects in speech perception. Holt et al. [1996] used this method to examine the effect of context observed by Lindblom and Studdert-Kennedy [1967]. Two sets of stimuli were constructed. The first set patterned those of earlier studies [Lindblom and Studdert-Kennedy, 1967; Nearey, 1989]. Synthetic speech stimuli varied perceptually from [bʌb] to [bɛb] and [dʌd] to [dɛd]. Along these series, each stimulus step increased in vowel F_2 frequency, rendering a change in perceived vowel identity from [ʌ] to [ɛ]. The primary distinction between [bVb] and [dVd] series was frequency of F_2 onset and offset. For [bVb] stimuli, F_2 onset and offset was relatively low in frequency (800 Hz); whereas for [dVd] stimuli, F_2 was higher (2,270 Hz). Pseudo-spectrograms illustrate these stimuli in figure 1.

A pair of hybrid nonspeech-speech stimulus series, comprised of glide-vowel-glide stimuli, also was constructed. The vowel segment was created using synthesis parameters identical to those used to create the midpoint of the CVC speech stimuli. The critical distinction was that, for these hybrid stimuli, nonspeech frequency-modulated (FM) sine-wave glides, instead of consonant transitions, served as flanking context. These glides modeled only the trajectory of F_2 center frequency for transitions in the speech stimuli. Although nonspeech glides share minimal spectral qualities with F_2 transitions of CVC stimuli, they fall far short of perceptual or acoustic equivalence. Voiced speech sounds possess rich harmonic structure with energy at each multiple of the fundamental frequency (F_0). The FM glides, in contrast, had energy only at the nominal F_2 center frequency, with no fine harmonic structure and no energy mimicking

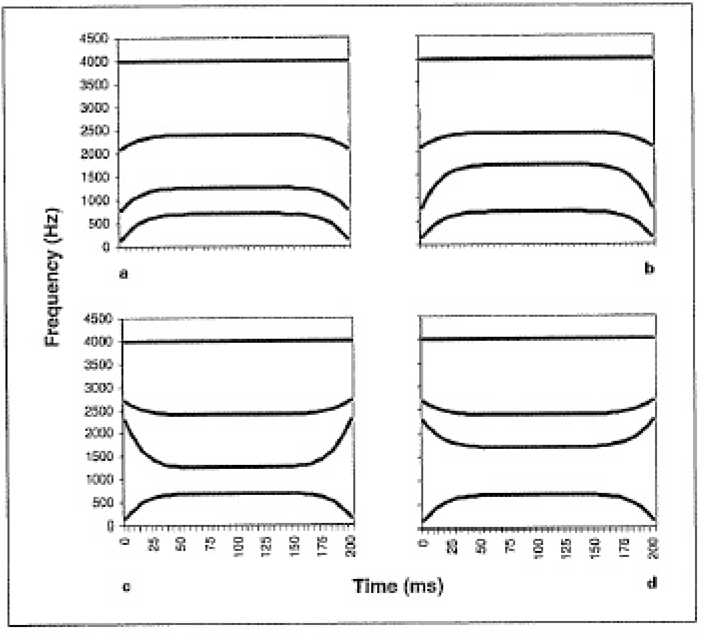

Fig. 1. CVC series endpoints. Representative pseudo-spectrograms of Holt et al. [1996]: stimulus-series endpoints. The top row (**a**, **b**) represents [bVb] stimuli. The bottom row (**c**, **d**) shows [dVd] endpoints. The left column (**a**, **c**) corresponds to low-F_2 ([ʌ]) stimulus endpoints and the right column (**b**, **d**) depicts high-F_2 ([ɛ]) endpoints. Eight intermediate stimuli (not shown), with F_2 midpoint frequency increasing in 50-Hz steps, comprised the remainder of the [bVb] and [dVd] series.

F_1 or F_3. In addition, formant transitions of speech stimuli are not much like FM tones, because component frequencies of speech do not vary (at constant F_0). Instead, relative amplitudes of harmonics change with changing shapes of the spectral envelope. Given these differences, simple FM glides capture only minimal spectral characteristics of energy in the region of F_2. Nevertheless, data from this experiment provide evidence that these minimal similarities are sufficient to elicit similar context effects on vowel perception.

Parallel to the results of Lindblom and Studdert-Kennedy [1967], listeners' vowel identification in CVC context was shifted such that, in the context of [dVd], vowels were more often identified as [ʌ]. In the context of [bVb], vowels were identified more often as [ɛ]. Remarkably, these results with speech were mirrored when energy adjacent to vowels was simple FM glides. As shown in figure 2, listeners labeled vowels flanked by glides modeling F_2 of [bVb] more often as [ɛ] than they did the same vowels

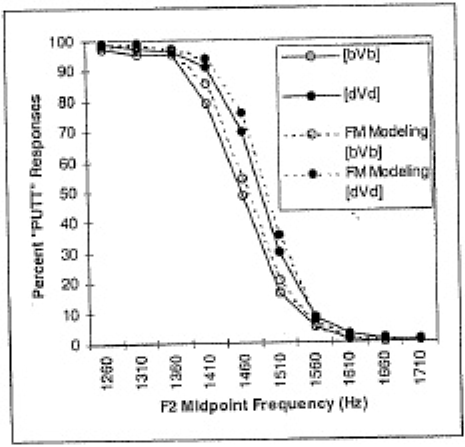

Fig. 2. Mean identification of vowels in consonant and glide contexts. Mean identification functions for vowels in the context of [bVb], [dVd], and two glide contexts that modeled these consonant contexts. Percent 'PUTT' responses as a function of midpoint F_2 frequency are plotted by context.

flanked by glides modeling F_2 of [dVd]. The correspondence of speech and nonspeech-speech hybrid context effects suggests that general auditory processes contribute to perceptual accommodation of coarticulation.

In order to evaluate whether listeners were somehow interpreting flanking energy as speech, Holt et al. [1996] conducted another study in which listeners were asked to identify flanking glides as [b] or [d]. Listeners were near perfect when asked to label consonants of speech CVCs. However, when complying with the unusual request to label FM glides as consonants, listeners varied in the extent to which they were consistent in labeling glides as [b] or [d]. Some consistently labeled glides 'backwards' (descending glides as [b]), while others consistently labeled descending glides as [d]. Some listeners were consistently inconsistent in labeling. No matter how listeners labeled FM glides, there was no correlation between labels and vowel identification.

The effect of nonspeech context generalizes to other phonetic contexts. For VCVs such as [iba], [ida], [uba], and [uda], initial vowel context ([i] or [u]) exerts considerable influence on the perception of the following consonant ([b] or [d]). Based upon the results with CVCs, Holt [1999] predicted that FM glides mimicking only F_2 of naturally produced [i] or [u] would elicit effects on perception of a [ba] to [da] series like that found for preceding [i] and [u]. As was the case for CVC contexts, preceding FM glides produced a significant shift in vowel identification in the same direction as full-formant vowel stimuli despite an intervening silent interval corresponding to vocal tract closure.

Similarly, Lotto and Kluender [1998] examined [VC CV] disyllables. Perception of a series of CVs as [da] or [ga] can be altered by the presence of a preceding VC [Mann, 1980]. When listeners identify members of a synthetic speech series varying acoustically in F_3 onset frequency and perceptually from [ga] to [da], listeners report more [ga] percepts when the CV is preceded by the VC [al]. Conversely, listeners identify the same stimuli more often as [da] when following [ar]. Lotto and Kluender [1998] demonstrated that sine-wave FM glides modeling F_3 transitions of [al] or [ar], and even constant-frequency tones set at F_3 offset frequencies of [al] or [ar], induced the same pattern of [da]-to-[ga] identification responses as did natural and synthetic speech tokens of [al] and [ar]. The effects of nonspeech context have thus been demonstrated for vowels in CVCs and well as for consonants in VCVs and VCCVs.

Nonhuman Animal Listeners

Finally, another line of evidence demonstrates that these effects extend across species. Lotto et al. [1997] trained Japanese quail *(Coturnix coturnix japonica)* to peck at a lighted key in response to endpoints of the same [da] to [ga] series used by Lotto and Kluender [1998]. Birds trained to peck to [ga] pecked most vigorously to novel intermediate members of the [da]-to-[ga] series that were preceded by [al]. Correspondingly, quail trained to peck to [da] pecked most robustly when novel stimuli were preceded by [ar]. Japanese quail exhibited shifts in responses contingent upon VC like those found for identification of CVs by human listeners. Considering that quail should have had no selective pressure to develop specialized mechanisms for processing human speech, these data suggest broad generality of the responsible perceptual processes. In addition, quail had no experience with coarticulated speech, so behavior cannot be explained based on learned covariance of acoustic attributes of coarticulated speech. Similar to this notion of experience playing a role, Lindblom and Studdert-Kennedy [1967] noted the potential contribution of 'expectancy' in perception of coarticulated speech; although, they did not favor such an explanation. Based upon the quail findings, it is safe to conclude that, while experience with coarticulated speech could play a role (see 'The Role of Learning'), it is not a necessary condition. General perceptual processes, extensive enough to be in force for avian subjects, appear sufficient to elicit effects of context in speech perception.

Ecology of Sound-Producing Sources

Taken together, these multiple lines of evidence suggest that some general property of spectral processing in the auditory system contributes significantly to perceptual compensation of coarticulation. In light of the cross-species correspondence in perception of coarticulated speech, one may be encouraged to consider perceptual processes quite generally. Most broadly, one can begin considering perception of speech and other auditory events with an appreciation for the ecology of sound-producing sources. Perceptual systems must operate in accord with physical regularities in the environment. Most theorists (vision and other senses included) assume that perceptual systems are not general signal analysis devices. Instead, they evolved to use information as it is structured in the world. If one desires to bring general principles of auditory perception to bear upon speech perception, one must begin to think about general physical principles that structure the auditory world.

In this case, general principles with respect to physical constraints on sound-producing devices – and presumed to have shaped the evolution of auditory systems – may be extended to the vocal tract as a sound source. At a minimum, there must be at least an abstract accordance between the process of perception and the ecology of sound-producing sources. For perception of speech and other acoustic events, principles that govern a listener's maintenance of a cohesive auditory world must be in agreement with (though not strictly isomorphic to) principles that govern output from a sound source, and general constraints upon sound-producing events (speech articulation being one example) should be approximated in the operation of sensory processes [Kluender, 1991; Lotto et al., 1997]. This view is roughly consistent with Roger Shepard's [1984] theory of internalized constraints he calls 'psychophysical complementarity'.

What can one say, generally, about the ecology of sound-producing sources? It is known that most wordly physical structures, in contrast to electronic gadgets, produce sound that can change only so much so fast. Due to inertia and mass, physical systems tend to be assimilative. The configuration of a system at time t is significantly constrained by its configuration at time $t-1$. Connected speech follows this patterns.

Perceptual Contrast

Most generally, an effective way to 'neutralize' assimilatory effects is perceptual contrast. One frequently cited example for audition is frequently contrast [Cathcart and Dawson, 1928, 1929; Christman, 1954]. One sound is perceived as higher following a sound that is lower, and vice versa. Effects of adjacient [b] and [d] in CVCs are consistent with a contrast account, as listeners report more percepts of [ɛ] (higher F_2) flanked by lower-frequency [b] and more percepts of [ʌ] flanked by higher-frequency [d] [Holt et al., 1996]. Similarly, listeners are more likely to report hearing [d] (higher F_2) following [u] (lower F_2), and [b] following [i] [Holt, 1999]. Finally, for effects of [al] and [ar] on perception of following [da] or [ga], to the extent that frequency composition of the offset is higher (F_3 for [al]), listeners are more likely to report hearing [ga] (lower F_3 onset frequency) and vice versa [Lotto and Kluender, 1998].

For all the examples above, and for all reported cases, the influence of coarticulation is assimilative, and context effects in speech perception can be depicted as contrastive. A general mechanism of perceptual contrast may serve as a valuable tool in maintaining perceptual constancy across context dependence produced by sound-source assimilation. Contrast in itself, however, is a designation that does not implicate any specific auditory mechanism(s). Across perceptual modalities, contrast is an important mechanism for exaggerating differences between neighboring objects and events. The best-known examples are in the visual domain: enhancement of edges produced by lateral inhibition [Hartline and Ratliff, 1957], lightness judgments [Koffka, 1935], judgment of line orientation [Gibson, 1933]. Context effects in behavior are as varied as tempo of behavior [Cathcart and Dawson, 1928/1929], weight lifting [Guilford and Park, 1931]. Mechanisms of contrast exist for every perceptual modality [von Bekesy, 1967; Warren, 1985]. Across domains, contrast is a familiar observation 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, may play an important role in perception of coarticulated speech.

Relationship to Psychoacoustics and Auditory Neurophysiology

However alluring the broad concept of contrast may be, its ubiquity betrays the fact that it falls short as a rigorous explanation. Fortunately, there exist precedents in auditory perception that lend both greater precision to the construct and reveal potential explanation in underlying processes. At least one class of psychoacoustic findings, known collectively as 'auditory enhancement', bears note for its similarity to the present findings. Auditory enhancement refers, generally, to the observation that if one member of a set of equal-amplitude harmonics is omitted and then reintroduced, it stands out perceptually from the rest of the components. Most relevant to the present data, Summerfield et al. [1984, 1987] have related auditory enhancement to speech perception.

When a uniform harmonic spectrum composed of equal-amplitude harmonics is preceded by a spectrum complementary to a particular vowel (with troughs replacing formants and vice versa), listeners report hearing a vowel during presentation of the uniform spectrum [Summerfield et al., 1984]. Moreover, the vowel percept is appropriate for a vowel with formants at frequencies where there were troughs in the preceding spectrum. In a similar vein, a uniform harmonic spectrum precursor intensifies perception of vowels defined by very modest formants of 2–5 dB [Summerfield et al., 1987]. For each of these cases of auditory enhancement, results may be described as contrast between two complex sounds. The 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. Cast in this way, these findings are strikingly similar to effects of context upon vowel identification reported here. Speech and nonspeech-speech hybrids modeled after [dVd] (which have predominantly high-frequency F_2 composition), for example, generate more low F_2 frequency vowel identifications ([ʌ]). This perceptual shift is toward frequencies less well represented in adjacent consonants and nonspeech analogues.

Neural adaptation or adaptation of suppression may serve to enhance changes in spectral regions where previously there had been relatively little energy. As with so many other central issues for speech perception, Lindblom and Studdert-Kennedy [1967, p. 840] were ahead of their time in noting how some effects of acoustic context can be 'exemplified for instance by adaptation and fatigue'. Although adaptation, thus far, has been explored only in service of explaining auditory enhancement effects in psychoacoustic studies, there is great potential for extension to speech as Lindblom and Studdert-Kennedy [1967] conjectured. Delgutte [1996] and 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, or 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.

Often, arguments that effects in speech perception arise from 'general auditory' processes are taken as suggesting that such mechanisms must be peripheral. If, by peripheral, one means to imply that mechanisms exert their influence in the cochlea or at the level of the auditory nerve, then this is almost certainly false for the case of context effects in speech perception. Evidence in other stimulus paradigms suggests that more central regions of the auditory system are likely involved. Effects of context maintain for dichotic presentation, and the influence of preceding spectral energy on speech identification decays gradually over a time-course that extends into hundreds of milliseconds [Lotto, 1996; Holt, 1999]. In addition, neurophysiological investigation of these effects has borne little evidence of neural encoding of speech context effects or auditory enhancement at the level of the auditory nerve [Palmer et al., 1995; Holt and Rhode, 2000]. The mechanisms appear to have a more central origin.

It appears that more than 30 years ago, Lindblom and Studdert-Kennedy [1967] were well ahead of their time. Recent findings described above are consistent with their experimental precedents, and their auditory hypothesis appears secure, supported better than ever by psychoacoustic and neurophysiological findings. In contrast to intervening theorization, general perceptual mechanisms of spectral contrast play a substantial role in compensating for context-dependent patterns of CVC coarticulation.

The Role of Learning

The implication of general auditory processes should not be taken to suggest that there is no more to perceptual accommodation of coarticulation. It is not being claimed that spectral contrast serves to explain all perceptual accommodation of acoustic consequences of coarticulation. The clearest counterexamples are instances for which perception of preceding sounds is affected by following sounds, e.g. [al] to [ar] preceding [ʃ] or [s] [Mann and Repp, 1981]; [u] to [i] preceding [d] or [g] [Ohala and Feder, 1994], [ʃ] to [s] preceding [a] or [u] [Mann and Repp, 1980]. Although backward contrast effects abound in studies of visual perception, there is nothing in our present understanding of auditory neurophysiology that suggests that adaptation and suppression exert such effects.

Aside from auditory processes, experience with speech also is likely to play a significant role in perception. Lindblom and Studdert-Kennedy's [1967] consideration of 'expectancies' may also wear well with time. In coarticulated speech, owing to articulatory constraints, there is a strong correlation between the acoustic properties of preceding and following sounds. In general, when one considers perceptual learning, it is conceptualized in terms of perceptual systems coming to 'expect' this correlated structure though experience with such structure. By way of concrete example, with experience, listeners 'expect' [ɛ] to have lowered frequencies following [b] such that, when judging stimuli varying along a continuum from [ʌ] to [ɛ], listeners are more likely to report hearing [ɛ] following [b]. Coarticulation yields multiple covariances in the signal that are orderly in as much as they reflect dependable regularities in physical constraints upon articulators.

Learning is precisely about behavior coming to reflect covarying aspects of the environment. Holt et al. [1999] recently reported results from two studies using Japanese quail and chinchilla *(Chinchilla laninger)* to investigate the role of F_0 on perception of voicing for stop consonants. Human listeners are more likely to perceive stimuli with higher F_0 values as voiceless – a pattern that follows regularities in production [e.g. Chistovich, 1969; Haggard et al., 1970; Fujimara, 1971; Whalen et al., 1993]. Some investigators [Kingston and Diehl, 1994] have suggested that higher F_0 values enhance perception of voicelessness by exploiting auditory predispositions. A second hypothesis is that this perceptual pattern arises due to experience with F_0/VOT covariation in production. The first hypothesis was tested using chinchilla trained to respond differentially to stops depending upon whether the stops were voiced or voiceless. Absent experience with covariation between F_0 and VOT, there was no effect of F_0 on responses to novel stimuli with intermediate VOTs following training. In a second experiment, three groups of Japanese quail were trained to respond differentially to voiced versus voiceless stops in conditions with three different patterns of F_0/VOT covariation. In training, VOT and F_0 varied in the natural pattern (longer VOT, higher F_0), in an inverse pattern (longer VOT, lower F_0), or in a random pattern (F_0 and VOT uncorrelated). When tested on stimuli with intermediate values of VOT, the third group of quail (no correlation) replicated the chinchilla results with no significant effect of F_0. For the other groups, responses followed the experienced pattern of covariation. These data highlight the potential importance of experienced covariation in speech perception. Such findings do not depreciate the importance of contrast processes in perception of coarticulated speech, but they serve as reminder that there is likely more to the full explanation.

Conclusion

In closing, it is appropriate that Lindblom and Studdert-Kennedy [1967, p. 842] get the last word: 'It is worth reiterating... that mechanisms of perceptual analysis whose operations contribute to enhancing contrast in the above-mentioned sense are precisely the type of mechanisms that seem well suited to their purpose given the fact that the slurred and sluggish manner in which human speech sound stimuli are often generated tends to reduce rather than sharpen contrast.'

References

Bekesy, G. von: Sensory inhibition (Princeton University Press, Princeton 1967).

Cathcart, E.P.; Dawson, S.: Persistence (2). Br. J. Psychol. *19:* 343–356 (1928/1929).

Chistovich, L.A.: Variations of the fundamental voice pitch as a discriminatory cue for consonants. Sov. Physics-Acoustics *14:* 372–378 (1969).

Christman, R.J.: Shifts in pitch as a function of prolonged stimulation with pure tones. Am. J. Psychol. *67:* 484–491 (1954).

Delgutte, B.: Auditory neural processing of speech; in Hardcastle, Laver, The handbook of phonetic sciences (Blackwell, Oxford 1996).

Delgutte, B.; Hammond, B.M.; Kalluri, S.; Litvak, L.M.; Cariani, P.A.: Neural encoding of temporal envelope and temporal interactions in speech; in Ainsworth, Greenberg, Proc. Auditory Basis of Speech Perception (European Speech Communication Association, 1996).

Fowler, C.A.: An event approach to the study of speech perception from a direct-realist perspective. J. Phonet. *14:* 3–28 (1986).

Fowler, C.A.: Listeners do hear sounds, not tongues. J. acoust. Soc. Am. *99:* 1730–1741 (1996).

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

Fujimura, O.: Remarks on stop consonants: synthesis experiments and acoustic cues; in Hammerich, Jakobson, Zwirner, Form and substance: phonetic and linguistic papers presented to Eli Fischer-Jørgensen, pp. 221–232 (Copenhagen, Akademisk Forlag 1971).

Gibson, J.J.: Adaptation, after-effect and contrast in the perception of curved lines. J. exp. Psychol. *16:* 1–31 (1933).

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

Haggard, M.; Ambler, S.; Callow, M.: Pitch as a voicing cue. J. acoust. Soc. Am. *47:* 613–617 (1970).

Hartline, H.K.; Ratliff, F.: Inhibitory interaction of receptor units in the eye of *Limulus*. J. gen. Physiol. *40:* 1357–1376 (1957).

Holt, L.L.: Auditory constraints on speech perception: an examination of spectral contrast; doct. diss. University of Wisconsin-Madison (unpublished, 1999).
Holt, L.L.; Lotto, A.J.; Kluender, K.R.: Perceptual compensation for vowel undershoot may be explained by general perceptual principles. 131st Meet. Acoust. Soc. Am., Indianapolis 1996.
Holt, L.L.; Lotto, A.J.; Kluender, K.R.: Influence of fundamental frequency on stop-consonant voicing perception: a case of learned covariation or auditory enhancement? 138th Meet. Acoust. Soc. Am., Columbus 1999.
Holt, L.L.; Rhode, W.S.: Examining context-dependent speech perception in the chinchilla cochlear nucleus. 23rd Midwinter Meet. Assoc. Res. Otolaryngol. (2000).
Kingston, J.; Diehl, R.L.: Phonetic knowledge. Language *70*: 419–454 (1994).
Kluender, K.R.: Psychoacoustic complementarity and the dynamics of speech perception and production. Perilus *14*: 131–135 (1991).
Koffka, K.: Principles of Gestalt psychology (Harcourt, Brace, New York 1935).
Liberman, A.M.; Cooper, F.S.; Shankweiler, D.P.; Studdert-Kennedy, M.: Perception of the speech code. Psychol. Rev. *74*: 431–461 (1957).
Liberman, A.M.; Mattingly, I.G.: The motor theory of speech perception revised. Cognition *21*: 1–36 (1985).
Lindblom, B.E.F.: Spectrographic study of vowel reduction. J. acoust. Soc. Am. *35*: 1773–1781 (1963).
Lindblom, B.; Studdert-Kennedy, M.: On the role of formant transitions in vowel recognition. J. acoust. Soc. Am. *42*: 830–843 (1967).
Lotto, A.J.: General auditory constraints in speech perception; doct. diss. University of Wisconsin-Madison (unpublished, 1996).
Lotto, A.J.; Kluender, K.R.: General contrast effects of speech perception: effect of preceding liquid on stop consonant identification. Percept. Psychophys. *60*: 602–619 (1998).
Lotto, A.J.; Kluender, K.R.; Holt, L.L.: Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). J. acoust. Soc. Am. *102*: 1134–1140 (1997).
Mann, V.A.: Influence of preceding liquid on stop-consonant perception. Percept. Psychophys. *28*: 407–412 (1980).
Mann, V.A.; Repp, B.H.: Influence of vocalic context on perception of the [sh]-[s] distinction. Percept. Psychophys. *28*: 213–228 (1980).
Mann, V.A.; Repp, B.H.: Influence of preceding fricative on stop consonant perception. J. acoust. Soc. Am. *69*: 548–558 (1981).
Nearey, T.: Static, dynamic, and relational properties in vowel perception. J. acoust. Soc. Am. *85*: 2088–2113 (1989).
Ohala, J.J.; Feder, D.: Listeners' normalization of vowel quality is influenced by 'restored' consonantal context. Phonetica *51*: 111–118 (1994).
Öhman, S.E.G.: Coarticulation in VCV utterances: spectrographic measurements. J. acoust. Soc. Am. *39*: 151–168 (1966).
Palmer, A.R.; Summerfield, Q.; Fantini, D.A.: Responses of auditory-nerve fibers to stimuli producing psychophysical enhancement. J. acoust. Soc. Am. *97*: 1786–1799 (1995).
Repp, B.H.: Phonetic trading relations and context effects: new experimental evidence for a speech mode of perception. Psychol. Bull. *92*: 81–110 (1982).
Repp, B.H.: Bidirectional contrast effects in the perception of VC-CV sequences. Percept. Psychophys. *33*: 147–155 (1983).
Shepard, R.N.: Ecological constraints on internal representation: resonant kinematics of perceiving, imaging, thinking and dreaming. Psychol. Rev. *91*: 417–447 (1984).
Stevens, K.N.; Halle, M.: Remarks on analysis by synthesis and distinctive features; in Wathem-Dunn, Models for the perception of speech and visual form (MIT Press, Cambridge 1967).
Stevens, K.N.; House, A.S.: Perturbations of vowel articulations by consonantal context: an acoustical study. J. Speech Hear. Res. *6*: 111–128 (1963).
Summerfield, Q.; Haggard, M.P.; Foster, J.; Gray, S.: Perceiving vowels from uniform spectra: phonetic exploration of an auditory aftereffect. Percept. Psychophys. *35*: 203–213 (1984).
Summerfield, Q.; Sidwell, A.; Nelson, T.: Auditory enhancement of changes in spectral amplitude. J. acoust. Soc. Am. *81*: 700–707 (1987).
Warren, R.M.: Criterion shift rule and perceptual homeostasis. Psychol. Rev. *92*: 574–584 (1985).
Whalen, D.H.; Abramson, A.S.; Lisker, L.; Mody, M.: F_0 gives voicing information even with unambiguous voice onset times. J. acoust. Soc. Am. *93*: 2152–2159 (1993).
Williams, D.R.: Role of dynamic information in the perception of coarticulated vowels; doct. diss. University of Connecticut (unpublished, 1986).