



ELSEVIER

Speech Communication 41 (2003) 59–69

SPEECH
COMMUNICATION

www.elsevier.com/locate/specom

Sensitivity to change in perception of speech

Keith R. Kluender^{a,*}, Jeffrey A. Coady^a, Michael Kieft^b

^a Department of Psychology, University of Wisconsin, 1202 West Johnson Street, Madison, WI 53706, USA

^b School of Human Communication Disorders, Dalhousie University, 5599 Fenwick Street, Halifax, NS, Canada B3H 1R2

Abstract

Perceptual systems in all modalities are predominantly sensitive to stimulus change, and many examples of perceptual systems responding to change can be portrayed as instances of enhancing contrast. Multiple findings from perception experiments serve as evidence for spectral contrast explaining fundamental aspects of perception of coarticulated speech, and these findings are consistent with a broad array of known psychoacoustic and neurophysiological phenomena. Beyond coarticulation, important characteristics of speech perception that extend across broader spectral and temporal ranges may best be accounted for by the constant calibration of perceptual systems to maximize sensitivity to change.

© 2002 Elsevier Science B.V. All rights reserved.

1. Sensorineural systems respond to change

It is both true and fortunate that sensorineural systems respond to change and to little else. Perceptual systems do not record absolute level be it loudness, pitch, brightness, or color. This fact has been demonstrated in every sensory domain. Physiologically, sensory encoding is always relative. This sacrifice of absolute encoding has enormous benefits along the way to maximizing information transmission. Biological sensors have impressive dynamic range given their evolution via borrowed parts (e.g., gill arches becoming middle ear bones). However, biological dynamic range always is a small fraction of the physical range of absolute levels available in the environment as well

as in the perceptual range essential to organisms' survival. This is true whether one is considering optical luminance or acoustic pressure. The beauty of sensory systems is that, by responding to relative change, a limited dynamic range adjusts to maximize the amount of change that can be detected in the environment.

The simplest way that sensory systems adjust dynamic range to maximize sensitivity to change is via adaptation. Following nothing, a sensory stimulus triggers a strong sensation. However, when sustained sensory input does not change over time, constant stimulation loses impact. This sort of sensory attenuation due to adaptation is ubiquitous, and has been documented in vision (Riggs et al., 1953), audition (Hood, 1950), taste (Urbantschitsch, 1876, cited from (Abrahams et al., 1937)), touch (Hoagland, 1933), and smell (Zwaardemaker, 1895, cited from (Engen, 1982)). Adaptation is the simplest among multiple mechanisms that encourage sensitivity to stimulus change. Common among these is the fact that perception

* Corresponding author. Tel.: +1-608-262-9884; fax: +1-608-262-4029.

E-mail address: krkluend@facstaff.wisc.edu (K.R. Kluender).

of any object or event is always relative—critically dependent on its context.

Because it is relative change that is perceived, perception at any particular time or place depends on temporally or spatially adjacent information. Sensory contrast has been explored most as a mechanism of visual perception. Marr (1976, 1982) has suggested that contrast change is incorporated into the earliest visual processes—the primal sketch. This process involves locating, representing, and interpreting intensity changes of reflected light that typically signal boundaries. The paths that eyes traverse demonstrate this point. When viewing pictures, eye movements

typically track contours, edges and boundaries (Yarbus, 1967).

Instances of sensory contrast are ubiquitous in visual perception. For example, in Fig. 1(a), a gray patch looks darker on a lighter background than the exact same gray patch on a darker background. Similarly, in Fig. 1(b) the series of gray strips is arranged in ascending brightness. Each of the strips is of uniform intensity, but due to contrast, they do not appear uniform. They appear lighter where they abut the darker strip, and darker where they abut the lighter strip. In these cases, contrast serves to exaggerate changes in luminance making the sensory contours more salient.

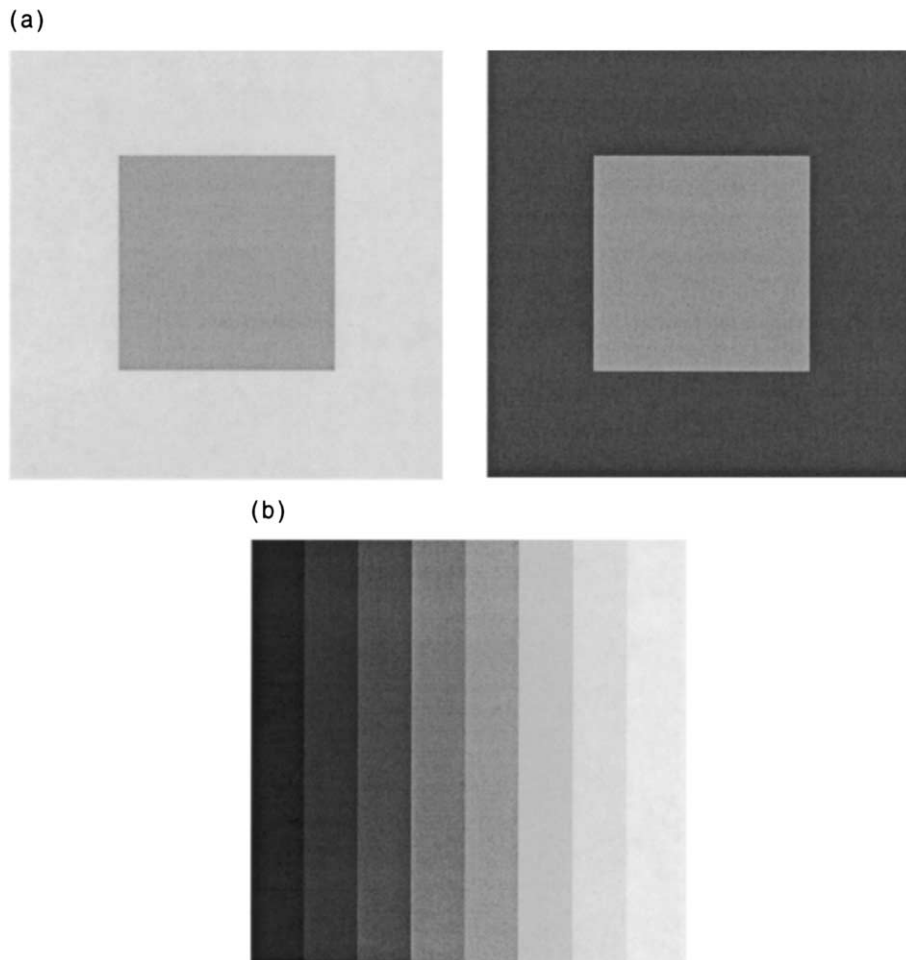


Fig. 1. Some examples of lightness contrast for visual images.

2. Spectral contrast and perceptual accommodation of coarticulated speech

Contrast effects are ubiquitous, and of course, they exist for audition (Cathcart and Dawson, 1928–1929; Christman, 1954). Forms of auditory contrast are important for several aspects of speech perception. Over the past few years, multiple studies conducted at Wisconsin have provided evidence that simple processes of spectral contrast serve, at least in part, as a perceptual complement to coarticulation in the acoustic signal. Coarticulation, the spatial and temporal overlap of adjacent articulatory activities, is reflected in the acoustic signal by severe context dependence. Acoustic information specifying one speech sound varies substantially depending on surrounding sounds. This context always follows the same pattern; adjacent sounds always assimilate toward the spectral characteristics of one another. Owing to mass and inertia of articulators (as well as planning), articulatory movements are compromises between where articulators have been and where they are going. Because the acoustic signal directly reflects these articulatory facts, the spectrum assimilates in the same fashion that speech articulation assimilates in coarticulation.

These basic observations have been made in the past. Lindblom (1963) (see also Öhman, 1965) provided some of the best early evidence concerning how context systematically influences speech production. He reported that the frequency of the second formant (F2) was higher in the productions of [dɪd] and [dʊd] than for the vowels [ɪ] and [ʊ] in isolation, and that F2 for both vowels was even lower in the productions of [bɪb] and [bʊb]. In both contexts, F2s approached the F2 of flanking consonants, which are higher for [d] than for [b]. In a subsequent study, Lindblom and Studdert-Kennedy (1967) demonstrated how perception of coarticulated vowels is complementary to these facts of articulation. They synthesized a series of consonant-vowel-consonant (CVC) syllables with F2 of the vowel varying acoustically from higher to lower and perceptually from [ɪ] to [ʊ]. The vowels were in the context of flanking [w] (low F2) or [y] (high F2) or in isolation without flanking semivowels. Listeners reported hearing [ɪ]

(higher F2) more often in the [wVw] context, and [ʊ] more often in the [yVy] context. Consonant context affected vowel perception in a manner complementary to the assimilative effects of coarticulation. Lindblom and Studdert-Kennedy (1967) wrote: “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” (p. 842, emphasis added).

One of the more widely described cases for perceptual context dependence concerns the realization of the phonemes [d] and [g] as a function of preceding liquid (Mann, 1980) or fricative (Mann and Repp, 1981). Perception of [d] as contrasted with perception of [g], can be largely signaled by the onset frequency and frequency trajectory of the third formant (F3). In the context of a following [a], a higher F3 onset encourages perception of [da] while a lower F3 onset results in perception of [ga]. Onset frequency of the F3 transition varies as a function of the preceding consonant in connected speech. For example, F3-onset frequency for [da] is higher following [al] in [alda] than when following [ar] in [arda]. The offset frequency of F3 is higher for [al] owing to a more forward place of articulation, and is lower for [ar].

Perception of [da] and [ga] has been shown to be complementary to the facts of production much as it is for CVCs. For a series of synthesized consonant-vowel syllables (CVs) varying in onset frequency of the third formant (F3) and varying perceptually from [da] to [ga], subjects are more likely to report hearing [da] (high F3) when preceded by the syllable [ar] (low F3), and hearing [ga] (low F3) when preceded by [al] (high F3) (Mann, 1980; Lotto and Kluender, 1998). In subsequent studies, the effect has been found for speakers of Japanese who cannot distinguish between [l] and [r] (Mann, 1986), for prelinguistic infants (Fowler et al., 1990), and for avian subjects (Lotto et al., 1997). The same pattern of findings for adult human listeners has been replicated for perception of [d] and [g] following fricatives [s] and [ʃ] such that

listeners are more likely to report hearing [d] (high F3) following [f] (lower-frequency noise) and hearing [g] (low F3) following [s] (higher-frequency noise). Although Mann and Repp (1981) also entertained alternative hypotheses, they presented spectral contrast among possible explanations for this result.

More recent studies have attempted to dissociate coarticulation per se from its acoustic consequences by combining synthetic speech targets with nonspeech flanking energy that captures minimal essential spectral aspects of speech. While investigating liquid-stop consonant combinations, Lotto and Kluender (1998) replaced [al] and [ar] precursors with frequency-modulated (FM) sine wave glides. When FM glides matched the center frequency and trajectory of F3 formant transitions of [al] and [ar] offsets, perception of following [d] and [g] patterned as it did for full-spectrum [al] and [ar] precursors. Lower-frequency glide precursors (mimicking [r]) promoted greater perception of higher-frequency F3 [d] while higher-frequency glides promoted greater perception of [g]. In a subsequent condition, FM glides were replaced with constant-frequency sinusoids set to the offset frequency of F3 for [al] and [ar] syllables, and the same pattern of results maintained.

Holt et al. (2000) proceeded to replicate the Lindblom and Studdert-Kennedy findings with CVCs using the vowels [ɛ] and [ʌ] flanked by stop consonants [b] and [d]. First, replicating Nearey (1989) they demonstrated that flanking stops affect perception of intervening vowels in the same way as semivowels [w] and [y]. Subjects reported hearing [ɛ] (higher F2) more often following [b] (lower F2) and hearing [ʌ] (lower F2) more often following [d] (higher F2). Holt et al. then replaced flanking [b] and [d] with FM glides that tracked the center frequency of only F2 for [b] or [g]. Again, the pattern of results for flanking nonspeech FM glides mimicked that for full-spectrum [b] and [d] syllable-initial and syllable-final transitions. Based upon the results for VCCVs (Lotto and Kluender, 1998) and these results for CVCs, they concluded that at least part of perceptual accommodation for coarticulation is not restricted to speech signals. In general, all of the findings are consistent with spectral contrast, whereby the spectral composi-

tion of context serves to diminish or enhance the perceptual efficacy of spectral components for adjacent sounds.

3. Potential mechanisms

In keeping with typical usage, the term “contrast” has been used in a largely descriptive way thus far. There are a large number of experimental precedents for spectral contrast—often called “auditory enhancement” and these precedents provide more specific hypotheses. Summerfield et al. (1984) established the existence of an “after-effect” in vowel perception. When a uniform harmonic spectrum was preceded by a spectrum that was complementary to a particular vowel with troughs replacing peaks and vice versa, listeners reported hearing a vowel during presentation of the uniform spectrum. The vowel percept (for the uniform spectrum) was appropriate for a spectrum with peaks at frequencies where there were troughs in the preceding spectrum.

Summerfield et al. (1984) noted that perceiving vowel sounds in uniform spectra (following appropriate complementary spectral patterns) has a well-known precedent in psychoacoustics. This oft-reported finding is that, if just one member of a set of harmonics of equal amplitude is omitted from a harmonic series and is reintroduced, then it stands out perceptually against the background of the pre-existing harmonics (Schouten, 1940; Green et al., 1959; Cardozo, 1967; Viemeister, 1980; Houtgast, 1972). Consider the series of studies reported by Viemeister (1980). He demonstrated that the threshold for detecting a tone in an harmonic complex is 10–12 dB lower when the incomplete harmonic complex (missing the target tone) is continuous as compared to when the onset of the inharmonic complex is the same as that for the target tone. This was referred to as an “enhancement effect”. Viemeister (1980) then examined a number of properties of this effect, finding that the complex need not be harmonic and that noise maskers or bandpass noise signal also served to enhance the detection of the tone. He also found the effect over a very wide range of intensities for maskers and targets.

Results from follow-up experiments with speech sounds by Summerfield et al. (1984) including dichotic presentation suggested that, at best, only minor roles were played by “central” processes (e.g., selective attention) or by processes of perceptual grouping. Instead, Summerfield et al. (1984, 1987) suggested that the effect may be rooted in peripheral sensory adaptation. This explanation is tenable for the nonspeech cases noted above. One could suggest that neurons adapt (become less sensitive), and the prominence of the added harmonic is due to the fact that neurons tuned to its frequency were not adapted prior to its onset.

In contrast, some researchers (e.g., Houtgast, 1974; Moore and Glasberg, 1983) have suggested that rapid adaptation serves mostly to enhance onsets selectively, with suppression being hypothesized to be a process through which differences in level of adjacent spectral regions in complex spectra (e.g., formants in speech signals) are preserved and/or enhanced. Viemeister and Bacon (1982) showed that, not only was an “enhanced” target tone more detectable; the tone also served as a more effective masker of a following tone. They suggested that suppression must be included in an adaptation scenario to place it in closer accord to this finding. Different frequency components of a signal serve to suppress one another, and Viemeister and Bacon (1982) suggest that what is relevant is not so much adaptation to absolute amplitude of specific frequency components, but rather that individual spectral channels attenuate their ability to suppress neighboring channels. This explanation is consistent with studies of two-tone suppression which has been cast as an instance of lateral inhibition in hearing (Houtgast, 1972). Investigators have argued that suppression helps to provide sharp tuning (e.g., Wightman et al., 1977; Festen and Plomp, 1981).

With respect to speech perception, Houtgast (1974) has argued that this process serves to sharpen the neural projection of a vowel spectrum in a fashion that effectively provides formant extraction. One way to conceptualize these processes is that they serve to provide simultaneous spectral contrast, enhancing prominences versus spectral regions of lesser energy. Summerfield et al. (1984, 1987) suggest that either simple adaptation or

adaptation of suppression could serve to enhance changes in spectral regions where previously there had been relatively little energy. For speech, changes in the distribution of spectral energy would be enhanced, and Summerfield et al. go on to suggest that this process could be important in perception of formant transitions such that changes in frequency prominences “will make formants *contrast* with what were valleys in the preceding vowel” (p. 213, emphasis added).

There also exist several neurophysiological observations that bear upon enhancement effects. In particular, a number of neurophysiological studies of auditory nerve (AN) recordings (e.g., Smith and Zwillocki, 1971; Smith, 1979; Smith et al., 1985) strongly imply a role for peripheral adaptation. More recently, Delgutte and colleagues (Delgutte, 1980, 1986, 1996; Delgutte et al., 1996; Delgutte and Kiang, 1984) have established the case for a much broader role of peripheral adaptation for perception of speech. He notes that peaks in AN discharge rate correspond to spectro-temporal regions that are rich in phonetic information, and that adaptation increases the resolution with which onsets are represented. This role of adaptation for encoding onset information is consistent with earlier observations noted above. Delgutte and colleagues (1994) note neurophysiological evidence that “adaptation enhances *spectral contrast* between successive speech segments” (p. 3, emphasis added). This enhancement arises because a fiber adapted by stimulus components close to its CF is relatively less responsive to subsequent energy at that frequency, while stimulus components not present immediately prior are encoded by fibers that are unadapted—essentially the same process offered by psychoacousticians but now grounded to physiology. Delgutte also notes that adaptation takes place on many timescales. In general, adaptation effects are sustained longer with increasing level in the auditory system. Some of the effects described above (particularly those that are temporally extended) may be less likely to have very peripheral origin.

Inspired by Summerfield and his colleagues’s studies (1984, 1987), Coady and Kluender (2001) recently conducted experiments to investigate whether the same processes were responsible for

the findings with very simple nonspeech flanking stimuli (e.g., Lotto and Kluender, 1998; Holt et al., 2000) and those earlier studies which used rich spectra that were complementary to those for vowel sounds. Although there is an extended history of using sine waves and FM glides as nonspeech proxies for formants, such sounds have only limited resemblance to speech formants. While it is true that spectrograms illustrate formants as bands of energy and formant transitions as bands of energy traversing frequency, such pictorial descriptions can be misleading. For example, if fundamental frequency (f_0) is constant, individual harmonics of the fundamental do not change frequency; only relative amplitudes of harmonics change. Individual frequency components of the speech spectrum change frequency no more than f_0 changes. On the other hand, if one considers the auditory periphery as a series of frequency channels, an FM glide can be viewed as introducing sequential changes in amplitude across successive channels. In an early report concerning effects of simple nonspeech context (FM and constant-frequency sine waves), Lotto and Kluender (1998) used the term “frequency contrast” to describe their effects. This may be a misnomer because the frequency of spectral components did not change in their speech stimuli. Describing effects as frequency contrast implied effects due to frequency differences rather than as differences in relative amplitude or spectral envelope.

Coady and Kluender set out to bridge this gap between simple sine waves as weak renditions of speech and the empirical precedents of Summerfield and his colleagues by investigating whether “afterimage” effects translated into the same sorts of changes in phonetic perception brought about by simpler nonspeech precursors. They elected to use VCV sequences for which characteristics of the initial vowel have dramatic effects on the spectral characteristics of the following consonant in production (Öhman, 1965). And, like the findings reported above for perception of vowels in CVCs, perception of spectral information for the consonant in VCVs is complementary to the facts of production

First, Coady and Kluender (2001) created a series of CV syllables varying acoustically in F2-

onset frequency and perceptually from [ba] to [da] using the cascade branch of the Klatt (1980) synthesizer. Then, two sets of precursor stimuli were synthesized by selectively filtering portions of a harmonic spectrum based on the default glottal model found in (Klatt, 1980) (approximately -6 dB/octave). One set of precursors was modeled after the vowels [e] and [o], and was created by passing the harmonic spectrum through bandpass filters with center frequencies appropriate for each of the first three formants of the vowels [e] and [o] (Fig. 2). The second set of precursors was created by passing the same harmonic spectrum through bandstop filters with the same center frequencies, resulting in troughs where peaks would occur for the vowels. For example, while the [e] precursor had a spectral peak centered at 1920 Hz, its complementary stimulus had a trough at this frequency.

It was hypothesized that these vowel complements should affect perception of the following stop in a fashion consistent with the other adaptation effects that have been discussed above. A trough in place of a formant should serve to enhance the perception of spectral energy following the vowel complement, and therefore show a perceptual effect opposite that observed for normal vowels. Precursor vowel complements should alter perception in a fashion opposite that of normal (noncomplement) vowels because troughs in energy should increase excitability within a frequency range for which excitability was attenuated when a frequency prominence (formant) was present in a normal vowel.

As expected, listeners were more likely to hear [d] following [o] in [oCa] and more likely to hear [b] in [eCa], consistent with Holt’s findings (1999). Further, as predicted by the spectral contrast account, effects for spectral complements were the reverse of this pattern. Listeners were more likely to hear [b] following the spectral complement of [o] and more likely to hear [d] following the spectral complement of [e]. While significant, effects for spectral complements of the vowels were smaller than those for the more vowel-like precursors. These results suggest that similar underlying processes may account both for the very simple nonspeech precursors and for the spectrally rich vowel

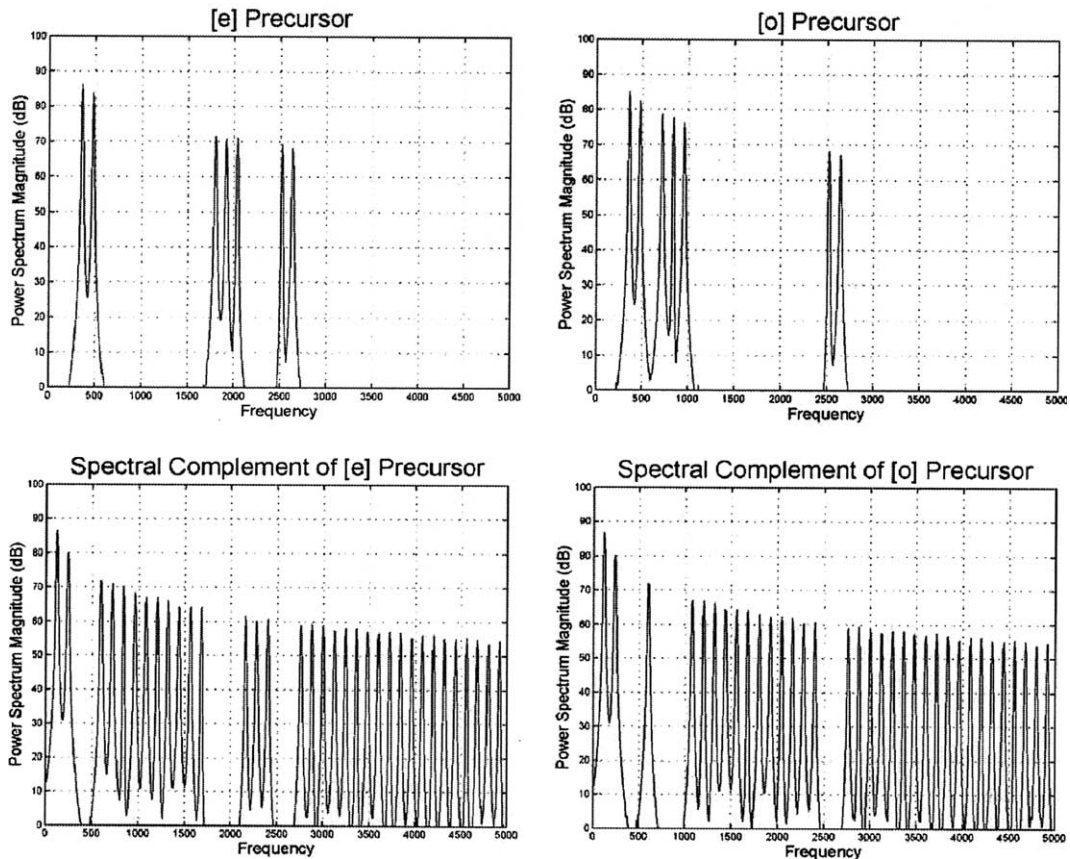


Fig. 2. Spectra of 3-band renditions of [e] and [o] and spectral complements of [e] and [o] used by Coody and Kluender (2001).

complements. Although more complex or domain-limited theories of speech perception have been proposed to explain perception of coarticulated speech, the above patterns of performance in the context of both simple and complex stimuli suggest that spectral contrast is an important part of the explanation.

4. Broader spectral and temporal effects

The above demonstrations of spectral contrast have been relatively focused in both time and frequency. The temporal extent of these effects is on the order of less than 1/2 s, and the particular spectral components of interest have been fairly local (e.g., individual spectral prominences). In keeping with the fundamental principle that per-

ceptual systems respond primarily to change, it follows that introduction of stationary signal characteristics should alter perception such that reliable, redundant, and uninformative contextual acoustic properties are effectively ignored. For example, visual perceivers maintain brightness and color constancy in the face of changes in level or spectral composition of illumination, respectively. There have been a number of recent demonstrations that spectrally broad and temporally extended characteristics of the signal also serve to modify perception of speech, and all of these findings further emphasize the importance of stability and change for perception of speech.

Watkins and Makin (1994, 1996a,b) demonstrated that long-term spectral characteristics of a precursor sentence affect perception of a following target stimulus. Precursor sentences were passed

through specially designed spectral difference envelope filters with a frequency response described by the spectral envelope of one vowel minus the spectral envelope of another. One filter had the spectral envelope of [ɛ] minus the spectral envelope of [ɪ]. The other had the spectral envelope of [ɪ] minus the spectral envelope of [ɛ]. The former spectral difference filter served to pass spectral characteristics of representative of [ɛ] in the precursor while attenuating spectral characteristics of [ɪ]. The reverse was true for the latter filter. The task for listeners was to identify vowels (drawn from a series of vowels varying perceptually from [ɛ] to [ɪ]) following the filtered sentences. Following the precursor filtered by the spectral envelope of [ɛ] minus the spectral envelope of [ɪ], listeners were more likely to hear [ɪ] following the precursor. Similarly, precursors filtered by the spectral envelope of [ɪ] minus [ɛ] resulted in more [ɛ] responses. Listeners were most likely to report hearing the vowel with a spectrum most distinctive from the long-term spectrum of the precursor.

More recently, Kiefte and Kluender (2001) further examined how reliable temporal and spectral characteristics can alter the way speech sounds are perceived. They first demonstrated how spectral tilt (relative balance between low and high-frequency energy) could alter perception of monophthongal vowels. For [u] and [i], the spectrum for [u] is biased toward lower-frequency energy (first formant (F1) and F2) while the spectrum for [i] has relatively more energy at higher frequencies (F2 and F3). Tilt was defined as the slope across peak amplitudes in the regions of F1 and F3, and a single-pole filter was employed to apply tilt characteristic of [u] or characteristic of [i]. Applying the global spectral tilt of [u] to a series of vowel sounds ranging from [u] to [i] encouraged greater perception of [u], and applying the spectral tilt of [i] resulted in listeners hearing more stimuli as [i]. As such, this is a simple demonstration that both local (formant frequencies) and global (tilt) properties of the spectrum contribute to vowel perception.

Similar to Watkins and Makin (1994, 1996a,b), Kiefte and Kluender (2001) then presented listeners with the same monophthong vowel sounds [i-u] preceded by carrier sentences that were processed

using the same filters characterizing gross spectral tilt for [u] and [i]. When long-term spectral tilt was the same for both the carrier sentence and the target vowel, the effect of tilt on vowel perception was effectively negated, and judgments of vowel sounds could be predicted entirely on the basis of formant frequencies. Patterns of performance indicated that extended reliable (redundant) spectral properties were effectively factored out of perception. When long-term spectral tilt changed between precursor and target, the effect of tilt for the vowel perception was robust, again illustrating the importance of spectral change.

One working hypothesis with respect to spectral change is that spectral and temporal extent of contrast effects may be systematically related. Temporally local changes (relatively brief time base in tens of milliseconds) have relatively limited spectral extent (within about a critical band or across a limited area of the basilar partition). As the time base broadens, as was the case for precursor passages, spectral extent of effects also becomes broader. This hypothesis is motivated by facts of auditory physiology and facts about the statistics of natural sound sources. In the auditory nervous system, early levels of processing (AN and ventral cochlear nucleus (VCN)) have relatively narrow frequency tuning and integrate energy over a relatively brief time base. With ascending levels of the auditory system, there are increasing degrees of integration across channels (e.g., Abeles and Goldstein, 1972) and the time base for integration grows about an order of magnitude longer (e.g., Schreiner and Langner, 1984; Creutzfeldt et al., 1980).

This neural organization corresponds well to typical statistical structure of acoustic ecology. It is generally the case that detailed spectral changes (e.g., formant onsets) are relatively short lived, and that broader spectral composition (e.g., ambient background energy) is best defined over a longer time base. For example, as one samples the spectrum for a longer period of time, the averaged spectrum becomes smoother, and the defining characteristics become more global. A more formal way to describe this is that, as the duration of sampling gets longer, fewer cepstral coefficients account for a larger share of the variance in the

cumulative spectrum. Further study is required, and for now, this putative relationship between temporal and spectral extent within the auditory system is offered only as a working hypothesis.

Kiefte and Kluender (2001) did carry out one additional experiment to investigate the importance of long-term statistical characteristics of the signal. By the hypothesis above, global spectral properties may be effective over a longer time base primarily because it is only global properties that survive cumulative sampling (and averaging) across longer spans of a signal that continuously changes with respect to spectrally local properties. If this is the case, then it may be possible to have relatively local spectral properties affect perception to the extent that those local properties are reliable across a broader time base. Kiefte and Kluender filtered a carrier sentence with a single pole corresponding exactly to the F2 of the following vowel. Perceptual data indicated that this manipulation of the carrier effectively cancelled the influence of F2 for perception of the target vowel. Instead, listeners appeared to rely entirely upon global spectral characteristics (tilt) for identification of the target vowel. This initial finding is consistent with the hypothesis based upon statistical structure of environmental sounds because, despite F2 being spectrally local, spectral characteristics were entirely reliable (redundant across extended sampling). It bears note, however, that such a finding remains consistent with organization of the auditory nervous system. While greater spectral and temporal integration becomes possible with increasing levels of the auditory system, it is not the case that integration requires the loss of all spectral detail at higher levels. Neurons sensitive to either spectrally narrow or spectrally broad stimulus characteristics are each found in auditory cortex (Abeles and Goldstein, 1972). Apparently, there are ample parallel pathways to accommodate information at multiple grains of analysis.

5. Conclusions

Much is often made of the fact that speech production is dynamic and that it is necessary to

include some sensitivity to dynamics in one's model of speech perception. This is undoubtedly true, but it is rarely noted that it is true in a relatively obvious way given the fact that perceptual systems respond nearly exclusively to change. Perception of speech is not distinctive by virtue of sensitivity to kinematic aspects of the signal. It would be exceptional if it were *not* sensitive to change. Characteristics of the speech signal that change as a function of time are exactly what the auditory system has evolved to encode for all sounds.

Here, only a sampling of the ways sensitivity to change is revealed for speech perception was presented. Across all of the findings described above, data are best explained by adhering to a principle long appreciated for perception most generally. Perceptual systems respond to change—and very little else. Many examples of perceptual systems responding to change can be portrayed as instances of contrast. In part, contrast explains some of the most central and enduring challenges to theories of speech perception, this being the perception of coarticulated (acoustically assimilated) speech. Beyond this however, important characteristics of speech perception that extend across broader spectral and temporal ranges may best be accounted for by the constant calibration of perceptual systems to maximize sensitivity to change.

Acknowledgements

Research supported by NIDCD DC04072.

References

- Abeles, M., Goldstein, M.H., 1972. Responses of single units in the primary auditory cortex of the cat to tones and to tone pairs. *Brain Res.* 42, 337–352.
- Abrahams, H., Krakauer, D., Dallenbach, K.M., 1937. Gustatory adaptation to salt. *Amer. J. Psych.* 49 (3), 462–469.
- Cardozo, B.L., 1967. Ohm's law and masking. In: IPO Annual Progress Report, Vol. 2. Institute for Perception Research, Eindhoven, The Netherlands, pp. 59–64.
- Cathcart, E.P., Dawson, S., 1928–1929. Persistence (2). *Br. J. Psych.* 19 (2), 343–356.
- Christman, R.J., 1954. Shifts in pitch as a function of prolonged stimulation with pure tones. *Amer. J. Psych.* 67, 484–491.

- Coady, J.A., Kluender, K.R., 2001. The role of spectral contrast in the perception of stop consonants following vowels and their spectral complements. *J. Acoust. Soc. Amer.*, Part 2 109 (5), 2315.
- Creutzfeldt, O., Hellweg, F.-C., Schreiner, C., 1980. Thalamocortical transformation of responses to complex auditory stimuli. *Exp. Brain Res.* 39, 87–104.
- Delgutte, B., 1980. Representation of speech-like sounds in the discharge patterns of auditory nerve fibers. *J. Acoust. Soc. Amer.* 68, 843–857.
- Delgutte, B., 1986. Analysis of French stop consonants with a model of the peripheral auditory system. In: Perkell, J.S., Klatt, D.H. (Eds.), *Invariance and Variability of Speech Processes*. Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 131–177.
- Delgutte, B., 1996. Auditory neural processing of speech. In: Hardcastle, W.J., Laver, J. (Eds.), *The Handbook of Phonetic Sciences*. Blackwell, Oxford, pp. 507–538.
- Delgutte, B., Kiang, N.Y.S., 1984. Speech coding in the auditory nerve IV: Sounds with consonant-like dynamic characteristics. *J. Acoust. Soc. Amer.* 75, 897–907.
- 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.), *Auditory Basis of Speech Perception*. European Speech Communication Association, pp. 1–9.
- Engen, T., 1982. *The Perception of Odors*. Academic Press, New York.
- Festen, J.M., Plomp, R., 1981. Relations between auditory functions in normal hearing. *J. Acoust. Soc. Amer.* 70, 356–369.
- 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.
- Green, D.M., McKey, M.J., Licklider, J.C.R., 1959. Detection of a pulsed sinusoid in noise as a function of frequency. *J. Acoust. Soc. Amer.* 31, 1146–1152.
- Hoagland, H., 1933. Quantitative aspects of cutaneous sensory adaptation I. *J. Gen. Phys.* 16, 911–923.
- Holt, L.L., 1999. Auditory constraints on speech perception: an examination of spectral contrast. Unpublished Ph.D. dissertation, University of Wisconsin-Madison.
- Holt, L.L., Lotto, A.J., Kluender, K.R., 2000. Neighboring spectral content influences vowel identification. *J. Acoust. Soc. Amer.* 108 (2), 710–722.
- Hood, J.D., 1950. Studies in auditory fatigue and adaptation. *Acta Oto-Laryn.*, Suppl. 92.
- Houtgast, T., 1972. Psychophysical evidence for lateral inhibition in hearing. *J. Acoust. Soc. Amer.* 51, 1885–1894.
- Houtgast, T., 1974. Auditory analysis of vowel-like sounds. *Acustica* 31, 320–324.
- Kiefe, M.J., Kluender, K.R., 2001. Spectral tilt versus formant frequency in static and dynamic vowels. *J. Acoust. Soc. Amer.*, Part 2, 109 (5), 2294–2295.
- Klatt, D.H., 1980. Software for a cascade/parallel formant synthesizer. *J. Acoust. Soc. Amer.* 67 (3), 971–995.
- Lindblom, B.E.F., 1963. Spectrographic study of vowel reduction. *J. Acoust. Soc. Amer.* 35, 1773–1781.
- Lindblom, B.E.F., Studdert-Kennedy, M., 1967. On the role of formant transitions in vowel recognition. *J. Acoust. Soc. Amer.* 42, 830–843.
- Lotto, A.J., Kluender, K.R., 1998. General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *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. Amer.* 102, 1134–1140.
- Mann, V.A., 1980. Influence of preceding liquid in stop-consonant perception. *Percept. Psychophys.* 28, 407–412.
- Mann, V.A., 1986. Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of English "l" and "r". *Cognition* 24, 169–196.
- Mann, V.A., Repp, B.H., 1981. Influence of preceding fricative on stop consonant perception. *J. Acoust. Soc. Amer.* 69, 548–558.
- Marr, D., 1976. Early processing of visual information. *Phil. Trans. Royal Soc. London B* 275 (942), 483–519.
- Marr, D., 1982. *Vision*. W.H. Freeman, New York.
- Moore, B.C.J., Glasberg, B.R., 1983. Suggested formulas for calculating auditory-filter bandwidths and excitation patterns. *J. Acoust. Soc. Amer.* 74, 750–753.
- Nearey, T.M., 1989. Static, dynamic, and relational properties in vowel perception. *J. Acoust. Soc. Amer.* 85 (5), 2088–2113.
- Öhman, S.E.G., 1965. Coarticulation in VCV utterances: Spectrographic measurements. *J. Acoust. Soc. Amer.* 39 (1), 151–168.
- Riggs, L.A., Ratliff, F., Cornsweet, J.C., Cornsweet, T.N., 1953. The disappearance of steadily fixated visual test objects. *J. Opt. Soc. Amer.* 43 (6), 495–501.
- Schouten, J.F., 1940. The residue, a new component in subjective analysis. *Proc. Kon. Akad. Wetensch* 43, 356–365.
- Schreiner, C.E., Langner, G., 1984. Coding of temporal patterns in the central auditory system. In: Edelman, G.M., Gall, W.E., Cowan, W.M. (Eds.), *Auditory Function: Neurobiological Bases of Hearing*. Wiley, New York, pp. 337–361.
- Smith, R.L., 1979. Adaptation saturations and physiological masking in single auditory-nerve fibers. *J. Acoust. Soc. Amer.* 65, 166–178.
- Smith, R.L., Zwislocki, J.J., 1971. Responses of some neurons of the cochlear nucleus to tone-intensity increments. *J. Acoust. Soc. Amer.* 50, 1520–1525.
- Smith, R.L., Brachman, M.L., Frisina, R.D., 1985. Sensitivity of auditory-nerve fibers to changes in intensity: A dichotomy between decrements and increments. *J. Acoust. Soc. Amer.* 78, 1310–1316.

- Summerfield, Q., Haggard, M.P., 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. Amer.* 81, 700–707.
- Urbantschitsch, V., 1876. Beobachtungen über anomalien des geschmacks der tastempfindungen und der speichelsecretion in folge von erkrankungen der paukenhöhle. F. Enke, Stuttgart.
- Viemeister, N.F., 1980. Adaptation of masking. In: van den Brink, G., Bilsen, F.A. (Eds.), *Psychophysical, Physiological, and Behavioral Studies in Hearing*. University Press, Delft, pp. 190–197.
- Viemeister, N.F., Bacon, S.P., 1982. Forward masking by enhanced components in harmonic complexes. *J. Acoust. Soc. Amer.* 71, 1502–1507.
- Watkins, A.J., Makin, S.J., 1994. Perceptual compensation for speaker differences and for spectral-envelope distortion. *J. Acoust. Soc. Amer.* 96 (3), 1263–1282.
- Watkins, A.J., Makin, S.J., 1996a. Some effects of filtered contexts on the perception of vowels and fricatives. *J. Acoust. Soc. Amer.* 99 (1), 588–594.
- Watkins, A.J., Makin, S.J., 1996b. Effects of spectral contrast on perceptual compensation for spectral-envelope distortion. *J. Acoust. Soc. Amer.* 99 (6), 3749–3757.
- Wightman, F., McKee, T., Kramer, M., 1977. Factors influencing frequency selectivity in normal and hearing-impaired listeners. In: Evans, E.F., Wilson, J.P. (Eds.), *Psychophysics and Physiology of Hearing*. Academic Press, London, pp. 295–310.
- Yarbus, A.L., 1967. *Eye Movements and Vision* (L.A. Riggs, Trans.) Plenum Press, New York.
- Zwaardemaker, H., 1895. *Die Physiologie des Geruchs*. Engelmann, Leipzig.