

VOWEL INHERENT SPECTRAL CHANGE

Editors: Geoffrey Stewart Morrison, Peter F. Assmann

1 *Introduction*

Peter F. Assmann, Geoffrey Stewart Morrison

It has been traditional in phonetic research to characterise monophthongs using a set of static formant frequencies, i.e., formant frequencies taken from a single time-point in the vowel, or averaged over the time-course of the vowel. However, over the last twenty years a growing body of research has demonstrated that, at least for a number of dialects of North American English, vowels which are traditionally described as monophthongs often have substantial spectral change. Vowel inherent spectral change has been observed in speakers' productions, and has also been found to have a substantial effect on listeners' perception. In terms of acoustics, the traditional categorical distinction between monophthongs and diphthongs can be replaced by a gradient description of dynamic spectral patterns. This book includes chapters addressing various aspects of vowel inherent spectral change (VISC), including reviews of theoretical and experimental studies of the perceptually relevant aspects of VISC, historical changes related VISC, cross-dialect and cross-language comparisons, the effects of VISC on second-language speech learning, and the use of VISC in forensic phonetics.

I PERCEPTION AND MODELS

2 *Static and dynamic approaches to understanding vowel perception*

James M. Hillenbrand

The goal of this chapter will be to provide a broad overview of work leading up to the current view that vowel inherent spectral change plays a significant role in the recognition of vowel identity. Although seldom if ever explicitly stated, the view that implicitly guided vowel perception research for many years was the idea that nearly all of the information that was needed to specify vowel quality was to be found in a cross section of the vowel spectrum sampled at a reasonably steady portion of the vowel. There is now a considerable body of evidence showing that VISC (vowel-category-specific spectral changes throughout the course of the vowel) play a secondary but quite important role in the recognition of vowel identity. Evidence comes from a wide variety of experimental techniques: (1) measurement data showing that many nominally monophthongal English vowels show significant spectral change throughout the course of the vowel; (2) perceptual experiments with "silent center" vowels showing that vowel steady-states can be removed with minimal effect on vowel intelligibility; (3) perceptual experiments with both naturally spoken and synthetic speech signals showing that vowels with stationary spectral patterns are not especially well identified; (4) statistical pattern recognition studies showing that vowel categories are separated with far greater accuracy by models that take spectral change into account than otherwise comparable models using features sampled at steady-state; and (5) listening experiments showing that resynthesized vowels are significantly more intelligible when natural spectral change patterns are preserved.

3 *Theories of the perception of vowel inherent spectral change: A review*

Geoffrey Stewart Morrison

In many dialects of North American English, in addition to vowels which are traditionally

described as true and phonetic diphthongs, several traditional monophthongs also have substantial formant movement, and vowel inherent spectral change has been found to be an important factor in vowel identification. The present paper reviews literature pertinent to theories of the perceptually relevant aspects of vowel inherent spectral change. Of the three basic hypotheses, formant onset + offset, formant onset + slope, and formant onset + direction, the weight of evidence indicates that the offset hypothesis is superior. Models which fit curves to whole formant trajectories have, as yet, not been found to outperform simple models based on formant measurements taken at two points (onset and offset) in formant trajectories. The most successful curve-fitting model is interpretable as a parameterisation of the onset + offset hypothesis.

4 Formant trajectories as acoustic correlates to speech perception

Michael Kiefte, Tara Collins, Christian E. Stilp, Keith R. Kluender

Speech-perception research has largely focused on the first two or three formant frequencies as correlates to vowel identification by human listeners. However, while vowels may be well separated by F1 and F2, this separability is very much adversely affected by between-speaker and context variability. Many of these problems can be overcome by considering timevarying vowel-inherent spectral change (VISC; Nearey, 1989). Formant trajectories serve as good discriminants for vowels as formant trajectories are roughly constant within vowel across speaker size, age, and sex as well as across consonantal contexts. While there is much evidence that formant frequency changes across time are excellent correlates to vowel identification, there are alternative proposals that are consistent with observed results. What is more, it assumes that listeners track changes in formant frequency across time which appears to be a difficult problem. For example, speech recognition algorithms have largely shunned formant frequencies as they are very difficult to extract from the speech signal and measurement of spectral peak frequencies is considered highly unreliable. Although many believe that a detailed study of formant tracking in human listeners may provide much needed insight into solving this problem, a preliminary question must first be addressed: Are accurately tracked formant frequencies a necessary prerequisite to vowel perception by human listeners? This chapter describes results from several studies that examine the relationship between changing formant frequencies and perception. Previous studies by the authors show that an alternative representation of vowel acoustics, such as global spectral shape, are precluded by evidence that amplitudes are largely ignored in vowel perception. In addition, it has been shown that, in those cases that other spectral properties appear to have a perceptual effect it is because stimuli have used stationary formants. When formants are changing and dynamic, effects for formant amplitude of spectral tilt disappear indicating that, for the very vowels of interest to speech-perception research—i.e., those with naturally occurring VISC—formant frequency information is by far the most important acoustic property in vowel identification by human listeners. In another set of studies, it is shown that perceptual extrapolation of a formant sweep is mostly dependent on peak frequency and not spectral shape which demonstrates that listeners do indeed follow formant frequency changes as auditory objects. All of these studies carry with them caveats, the most important of which is that, while formant-frequency synthesis parameters are known in these experiments, we cannot know the perceived formant frequencies without further psychoacoustic testing. Further research on formant frequency perception will be described.

5 Perception of vowel sounds with a biologically realistic information theoretic model of speech perception

Keith R. Kluender, Christian, E. Stilp, Timothy R. Rogers, Michael Kiefte,

Two fundamental operating characteristics of sensorineural systems are: sensitivity to relative, not absolute, properties of their input; and, sensitivity to correlations among stimulus attributes. Here, we present three lines of investigation that illustrate how these two fundamental principles provide useful accounts for fundamental characteristics of perception of vowels within and across both talkers and languages. First, we explain how radical differences in physical acoustics of vowel sounds across talkers actually reveal highly systematic and nearly invariant covariance structure. Second, we expand these efforts to vowel inherent spectral change and implications for systematicities across vowel inventories for different languages. Finally, we extend our application of 2nd-order statistics (correlation) to illustrate “locus equations” for vowel sounds in the context of the consonants of English. In several instances, traditional problems in vowel perception are dissolved by this account. All of the foregoing fits well within a biologically realistic information theoretic model for speech perception that is consistent with contemporary efficient encoding models of visual processing.

6 *Dynamic specification theory across languages: An alternative view of vowel spectral change*

Winifred Strange, James J. Jenkins

This chapter will review the research by Strange and her colleagues on the role of onset and offset temporal dynamics in specifying the identity of coarticulated vowels in English and other languages. They review will summarize the empirical data from their first studies of vowel perception in citation-form CVC syllables, through studies of perception of coarticulated vowels in sentence materials in English and German. Strange’s Dynamic Specification theory will be compared and contrasted with Nearey’s Target + Offglide VISC model of perception of North American English vowels. A tentative model of possible ways that speakers maintain acoustic differentiation of coarticulated vowels in languages with small and large vowel inventories will be offered.

II DIACHRONY AND SYNCHRONY

7 *The contribution of dynamic formant differences in vowels to diachronic sound change*

Jonathan Harrington

The aim of the chapter will be to show that there are many kinds of sound changes to vowels that cannot be expressed just in terms of a push-chain rearrangement of positions in the vowel quadrilateral, but that diachronic vowel change can also involve dynamic features. An outline will be given of how dynamic features such as the skew and curvature of formants as a function of time can be parameterised acoustically. Examples will then be given of how these dynamic features can both trigger diachronic vowel change and determine the outcome of the phonetic differences between vowels, once a sound change has taken place.

8 *Cross-dialectal differences in dynamic formant patterns in American English*

Ewa Jacewicz, Robert A. Fox

There is a long phonetic tradition which views a vowel as a linguistic category whose position in the acoustic vowel space (in a two-dimensional F1 x F2 plane) can be adequately characterized by the formant values at a putative steady-state portion of the vowel (or at its midpoint). This approach is particularly apparent in sociolinguistic research comparing dialects or looking at language change (vowel shifts). However, our work on

cross-dialectal variation in vowels indicates that it is not necessarily this single position, per se, but the nature of the dynamic spectral change over the course of vowel's duration that constitutes the most robust differences among the regional variants of English. Our chapter will examine the basic acoustic variation in dynamic patterns of vowel formants in three regional varieties of American English spoken in southeastern Wisconsin (affected by the Northern Cities Shift), western North Carolina (Appalachian English affected by the Southern Vowel Shift) and central Ohio (not considered to be affected currently by any vowel shift). The data come from a very large corpus of recorded utterances (in a variety of phonetic and prosodic contexts) produced by speakers from very narrowly defined dialect areas. Apart from expected spectral variation as a function of context and speaking style, we found that the most consistent cross-dialectal differences among vowels lie in the way the spectral change (measured in terms of vector length, trajectory length as well as rate and direction of formant change) affects their monophthongal and diphthongal properties. Depending on the amount and the direction of the spectral change, the same vowel "category" may display three distinct acoustic patterns which are dialect-specific and reflect systematic changes taking place within the vowel system of a given dialect.

III ACQUISITION AND APPLICATION

9 Developmental patterns in children's speech: Time-varying spectral change in vowels

Peter F. Assmann, Terrance M. Nearey

The aim of this chapter is to compare the pattern of time-varying spectral change in vowels spoken by children and adults. Children's speech differs from adult speech in several important ways. First, children have smaller larynges and supra-laryngeal vocal tracts than adults, with the result that their formants and fundamental frequencies are higher. Second, the temporal and spectral properties of children's speech are inherently more variable, a consequence of developmental changes in motor control. Both of these sources of variability raise interesting questions for theories of talker normalization and vowel specification. This chapter evaluates the implications of acoustic variability in children's vowels for models of vowel specification that incorporate time-varying spectral change.

10 Vowel inherent spectral change and the second-language learner

Catherine L. Rogers, Merete Møller Glasbrenner

The number of research studies directly investigating the use of vowel-inherent spectral change is relatively small, compared to other types of studies of native vs. non-native speech perception. Nevertheless, the development of the appropriate use of this cue by first and second language learners and its integration with other cues to vowel identity can be seen as a method of investigating the development of the subtler aspects of vowel perception that are needed for native-like perception and, potentially, for the development of a robustness to the kinds of cue degradation that may occur in noise or reverberation. Thus, even subtle differences in the efficiency of cue use by non-native listeners may explain why even highly proficient relatively early learners of English as a second language have been shown to require slightly more favorable signal-to-noise ratios than monolingual native listeners for understanding speech in noise at a given criterion level. While other studies of second-language learners have shown specific evidence of differences in cue use or cue weighting by second-language learners, the results of our research comparing the use of dynamic spectral and temporal cues by native and non-native listeners have suggested that dynamic cue use by relatively early learners of English as a second language (Spanish

L1) is not particularly different from that of native English speakers; instead, the early learners' perception appears to be somewhat less *robust* to degradation or removal of the speech cues. These results dovetail with some results of the use of dynamic cues by children learning their first language and for elderly or hearing-impaired listeners. Rather than thinking of these effects as a lack of ability to *integrate* spectral cues, an alternative perspective is that these listeners are less able to use these cues independently of one another and to flexibly switch their listening strategy when one cue is masked or degraded and thus recover from the degradation. We will use this perspective to examine past and current literature from studies of the use of dynamic cues in first and second language learning and to develop a theory or perspective from which to interpret and explain these data.

11 *Vowel inherent spectral change in forensic voice comparison*

Geoffrey Stewart Morrison

Two-parameter models of vowel inherent spectral change, such as dual-target or target-plus-slope models, have been found to be adequate for vowel-phoneme identification. More sophisticated curve-fitting models do not appear to outperform such two-parameter models. This suggests that if only simple cues such as initial and final formant values are necessary for signaling phoneme identity, then speakers may have considerable freedom in the path taken between the initial and final formant values. If the constraints on formant trajectories are relatively lax with respect to vowel-phoneme identity, then with respect to speaker identity there may be considerable information contained in the details of formant trajectories. Differences in physiology and idiosyncracies in the use of motor commands may mean that different individuals consistently produce different formant trajectories between the beginning and end of the same vowel phoneme. If within-speaker variance is substantially smaller than between-speaker variance then formant trajectories may be exploited for forensic speaker comparison. This chapter reviews a number of forensic-voice-comparison studies which have extracted information relevant to speaker identity from formant trajectories. For the purposes of forensic voice comparison, models using parametric curves are found to outperform simple two-parameter models. The performance of different parametric curve models for extracting information and different methods for forensic likelihood ratio calculation are compared.