# Introduction

**Peter F. Assmann and Geoffrey Stewart Morrison**

## Abbreviations

| | |
|---|---|
| CVC | Consonant–vowel-consonant |
| DCT | Discrete cosine transform |
| F1 | First formant |
| F2 | Second formant |
| L1 | First language |
| L2 | Second language |
| $q_1$ | Scaling coefficient of the first mode (parameter in Story and Bunton articulatory synthesizer) |
| $q_2$ | Scaling coefficient of the second mode (parameter in Story and Bunton articulatory synthesizer) |
| VISC | Vowel inherent spectral change |

## 1 Introduction

The term *vowel inherent spectral change* (VISC) was coined in Nearey and Assmann (1986). It refers to the changes in spectral properties over the time course of a vowel which are characteristic of vowel-phoneme identity. It refers not only to the widely-recognized spectral changes found in diphthongs and triphthongs, but also to the less-well-recognized spectral changes which are characteristic of

P. F. Assmann
School of Behavioral and Brain Sciences, University of Texas at Dallas, Richardson, USA

G. S. Morrison (✉)
Forensic Voice Comparison Laboratory, School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, Australia
e-mail: geoff-morrison@forensic-voice-comparison.net

G. S. Morrison and P. F. Assmann (eds.), *Vowel Inherent Spectral Change*, Modern Acoustics and Signal Processing, DOI: 10.1007/978-3-642-14209-3_1,

vowel-phonemes which have traditionally been called monophthongs in some dialects of some languages, particularly in North American English. Although the importance of VISC in so-called monophthongs has been recognized at least since Joos (1948, p. 101, see also his Fig. 25) and Potter and Steinberg (1950), static spectral models of vowels have predominated in phonetics research. We think, however, that this situation is changing as a larger number of researchers in more sub-branches of phonetics research are now explicitly investigating VISC or incorporating VISC as a component of their wider experimental design. In order to draw attention to and encourage wider understanding and consideration of VISC, we felt it was time to produce a volume dedicated to the topic.[1] The present volume incorporates ten chapters (besides this introduction) written by 18 authors, some of whom have been working on VISC for decades and others for whom it is newer either as a focus of investigation or as an adjunct to their established research interests. We divide the volume into four parts: *VISC Perception*, *VISC Production*, *VISC in Different Populations of Speakers*, and *VISC Applied*. Below we briefly describe these sections and the chapters within them. We hope that the chapters give a representative coverage of the breadth of work being conducted on VISC at the present time and that they will also supply readers with relatively in-depth knowledge of historical, theoretical, and empirical work on VISC.

Our initial inspiration to work on this topic came from our former PhD supervisor, Terrance M. Nearey (Fig. 1), and we dedicate this volume to him. We thank the following individuals who reviewed earlier versions of the chapters: Robert A. Fox, James M. Hillenbrand, Michael Kiefte, Keith Kluender, Richard S. McGowan, Philip Rose, Christian E. Stilp, Stephen A. Zahorian.

*Postscript*: As this volume was going to press in November 2012 we received the sad news that James Jenkins (one of the chapter authors) had passed away. He had a long and distinguished career, and his contributions to science will be missed. On behalf of the authors of all the other chapters in this volume, we extend our sympathies to Winifred and to his other family and friends.

## 2 Summary of Sections and Chapters

The first and longest section in this volume deals with *VISC perception*, and particularly theories relating the dynamic spectral properties of vowels and the perception of vowel phonemes.

Hillenbrand (2013 Chap. 2) begins by providing a historical overview of research indicating that VISC exists in many nominal monophthongs, that it is important for the perception of vowel-phoneme identity (static targets are neither

[1] We also organized a special session on VISC at the 157th Meeting of the Acoustical Society of America in Portland, Oregon, in May 2009, at which early versions of a number of the chapters from this volume were presented.

**Fig. 1** Terry and Salvador at the International Congress of Phonetic Sciences 2003 Barcelona

necessary nor sufficient), and that models of dynamic spectral properties result in higher correct-classification rates and higher correlation with human-listeners' responses than models based on static measurements.

Morrison (2013a Chap. 3) then reviews studies addressing the classic hypotheses as to the perceptually relevant aspects of VISC: onset + offset, onset + slope, and onset + direction. He concludes that models based on the onset + offset hypothesis outperform the other two in terms of correct-classification rates and correlation with human-listeners' perceptual responses, including their pattern of vowel confusions.

Nearey (2013 Chap. 4) first summarizes the patterns of VISC that occur in several dialects of North American English. He then uses statistical models of formant trajectories, including in consonant–vowel-consonant (CVC) contexts, to address the question of whether VISC can be distinguished from the effects of coarticulation with adjacent consonants. Results indicate that more complex models than earlier two-point onset + offset models are needed, but that VISC is an independent aspect of vowel identity, not just a reflex of consonantal coarticulation.

Strange and Jenkins (2013 Chap. 5) review the studies that led them to posit their *Dynamic Specification* model of vowel perception. They observed that correct-identification rates are higher for vowels in consonant context than for vowels in isolation, and for silent-center/edge-only vowels than for center-only/silent-edge vowels. This led them to propose that the most important cues to vowel identity were in the spectro-temporal patterns of the consonant-vowel and vowel-consonant formant transitions. They also review a number of cross-language studies and propose that languages with more crowded vowel spaces use dynamic spectral patterns to maintain contrasts (adaptive dispersion).

Finally in this section on VISC perception, Kluender, Stilp, and Kiefte (2013 Chap. 6) discuss VISC within a broader information-theoretic efficient-coding framework for understanding speech perception. With information measured as cochlea-scaled spectral entropy (unpredictable change), the relatively low-frequency components of vowels (e.g., first and second formants, F1 and F2) provide high psychoacoustic potential information, which is further increased when there is spectral change (e.g., VISC). They also discuss VISC as a means of adaptive dispersion, helping to maintain psychoacoustic contrasts between vowel phonemes—if sensorineural systems optimize sensitivity to change then phonemes are better described in terms of how they contrast with other phonemes in the inventory rather than in terms of prototypes.

The second section, *VISC production*, consists of one chapter by Story and Bunton (2013 Chap. 7). They use a software-implemented articulatory synthesizer based on three-dimensional magnetic resonance imaging of human vocal tracts. The model is parsimonious, requiring three functions of distance from the glottis to describe the vocal tract of a particular speaker, and two parameters ($q_1$ and $q_2$) to describe the instantaneous configuration of the vocal tract. There is a nearly one-to-one mapping between these parameters and F1 and F2. Making the values of $q_1$ and $q_2$ time dependent leads to an articulatory synthesizer which can map vowel-inherent changes in vocal tract shape over the time course of a vowel to vowel-inherent changes in the acoustic spectrum. Human listeners' correct-identification rates for the vowels synthesized with dynamic vocal-tract patterns (and hence dynamic spectral patterns) were much higher than for those with static vocal tracts.

The third section looks at *VISC in different populations of speakers*, in speakers of different dialects of American English, in adult speakers versus child speakers of different ages, and in second-language versus first-language speakers of American English.

VISC has been largely neglected in dialectology. Jacewicz and Fox (2013 Chap. 8) examine cross-dialectal differences in the nominal monophthongs /ɪ/, /ɛ/, and /æ/ in three dialects of American English, central Ohio (Midland), southern Wisconsin (Inland North), and western North Carolina (South), and diachronic changes in those dialects. They conclude that VISC plays a central part in the differentiation of American English dialects and that there are systematic cross-generational changes in VISC patterns associated with ongoing sound changes (vowel shifts).

Assmann et al. (2013 Chaps. 1 and 9) summarize the results of a developmental study of VISC in children ranging in age from five through 18 years. They report reliable patterns of VISC throughout the age range studied, and note that these patterns are largely preserved across a wide range of formant frequency variation associated with age and sex differences. Statistical pattern recognition tests indicate that vowels are well-classified when formant frequency measurements are taken at two sample points around 20 and 70 % of the vowel duration. The optimum locations for onset + offset sampling of the formant trajectory do not vary substantially as a function of sex and age, and adding a third sample point does not lead to significantly better classification scores.

Rogers et al. (2013 Chap. 10) examine the production and perception of VISC by second-language (L2) learners of English (primarily first-language, L1, American Spanish learners of American English). They summarize a series of experiments that use natural and modified syllables to compare vowel identification performance by monolingual English speakers with early and late L2 learners of English. Compared to L1 listeners, both early and late L2 listeners' perception is more sensitive to manipulations that disrupt time-varying formant changes in vowels, and these difficulties may contribute to the overall greater difficulties they face when attending to speech in noisy environments. VISC enhances phonetic contrast, and in production L1 and L2 speakers may adopt different strategies in using VISC to maintain separation between otherwise neighboring vowels.

The last section, *VISC applied*, consists of one chapter by Morrison (2013b Chap. 3) in which he reviews the use of VISC in forensic voice comparison. He provides theoretical motivation as to why measurements of VISC might be good features for forensic voice comparison, and critically reviews procedures which have been used to extract information from VISC. He concludes that parametric curve fitting is the best procedure, many of the others having theoretical or practical flaws. He also empirically demonstrates that a procedure based on fitting parametric curves to formant trajectories outperforms an onset + offset model.

## References


Assmann, P.F., Nearey, T.M., Bharadwaj, S.V.: Developmental patterns in children's speech: Patterns of spectral change in vowels. In: Morrison, G.S., Assmann, P.F. (eds.), Vowel Inherent Spectral Change, Chap. 9. Springer, Heidelberg (2013)

Hillenbrand, J.M.: Static and dynamic approaches to understanding vowel perception. In: Morrison, G.S., Assmann, P.F. (eds.), Vowel Inherent Spectral Change, Chap. 2. Springer, Heidelberg (2013)

Jacewicz, E., Fox, R.A.: Cross-dialectal differences in dynamic formant patterns in American English vowels. In: Morrison, G.S., Assmann, P.F. (eds.), Vowel Inherent Spectral Change, Chap. 8. Springer, Heidelberg (2013)

Joos, M.A.: Acoustic phonetics. Language Supplement. **24**, 1–136. Stable http://www.jstor.org/stable/522229 (1948)

Kluender, K.R., Stilp, C.E., Kiefte, M.: Perception of vowel sounds within a biologically realistic model of efficient coding. In: Morrison, G.S., Assmann, P.F. (eds.), Vowel Inherent Spectral Change, Chap. 6. Springer, Heidelberg (2013)

Morrison, G.S.: Theories of vowel inherent spectral change: A review. In: Morrison, G.S., Assmann, P.F. (eds.), Vowel Inherent Spectral Change, Chap. 3. Springer, Heidelberg (2013a)

Morrison, G.S.: Vowel-inherent spectral change in forensic voice comparison. In: Morrison, G.S., Assmann, P.F. (eds.), Vowel Inherent Spectral Change, Chap. 11. Springer, Heidelberg (2013b)

Nearey, T.M.: Vowel inherent spectral change in the vowels of North American English. In: Morrison, G.S., Assmann, P.F. (eds.), Vowel Inherent Spectral Change, Chap. 4. Springer, Heidelberg (2013)

Nearey, T.M., Assmann, P.F.: Modeling the role of vowel inherent spectral change in vowel identification. J. Acoust. Soc. Am. **80**, 1297–1308 (1986). doi:10.1121/1.394433

Potter, R.K., Steinberg, J.C.: Toward the specification of speech. J. Acoust. Soc. Am. **22**, 807–820 (1950). doi:10.1121/1.1906694

Rogers, C.L., Glasbrenner, M.M., DeMasi, T.M., Bianchi, M.: Vowel-inherent spectral change and the second-language learner. In: Morrison, G.S., Assmann, P.F. (eds.), Vowel Inherent Spectral Change, Chap. 10. Springer, Heidelberg (2013)

Strange, W., Jenkins, J.J.: Dynamic specification of coarticulated vowels: Research chronology, theory, and hypotheses. In: Morrison, G.S., Assmann, P.F. (eds.), Vowel Inherent Spectral Change, Chap. 5. Springer, Heidelberg (2013)

Story, B.H., Bunton, K.: Simulation and identification of vowels based on a time-varying model of the vocal tract area function. In: Morrison, G.S., Assmann, P.F. (eds.), Vowel Inherent Spectral Change, Chap. 7. Springer, Heidelberg (2013)

Springer

http://www.springer.com/978-3-642-14208-6

Vowel Inherent Spectral Change
Morrison, G.S.; Assmann, P.F. (Eds.)
2013, VI, 286 p., Hardcover
ISBN: 978-3-642-14208-6