*INTERSPEECH 2019: Show & Tell Contribution*
September 15–19, 2019, Graz, Austria

# Formant pattern and spectral shape ambiguity of vowel sounds, and related phenomena of vowel acoustics – Exemplary evidence

*Dieter Maurer*[1], *Heidy Suter*[1], *Christian d'Hereuse*[1], *Volker Dellwo*[2]

[1]Institute for the Performing Arts and Film, Zurich University of the Arts, Switzerland
[2]Department of Computational Linguistics, University of Zurich, Switzerland

dieter.maurer@zhdk.ch, chdh@inventec.ch, heidy.suter@zhdk.ch, volker.dellwo@uzh.ch

## Abstract

In the specialist literature on vowel acoustics, there is an extensive and often controversial debate on whether the primary acoustic cues of vowel quality are contained in the formant patterns or, alternatively, in the spectral shape. Yet, recent studies have shown that neither formant patterns nor spectral shapes are vowel quality-specific but that they are ambiguous because of a complex interaction between pitch and vowel-related spectral characteristics. In order to give insight into the phenomenon of formant pattern and spectral shape ambiguity of vowel sounds and its role for vowel acoustics, exemplary series of speech and of vowel sounds are presented in an online documentation, most of them selected from the Zurich Corpus. The presentation includes sound playbacks and results of an acoustic analysis (FFT spectra, LPC curves, spectrograms, $f_0$ contours, formant patterns) and of a vowel recognition test. A Klatt synthesiser is also included for resynthesis and synthesis purposes. The presentation intends (i) to support researchers in their evaluation of existing and future studies, questioning whether the actual variation and pitch-dependency of the vowel spectrum is taken into account when attempting to generalise experimental results, and (ii) to support students in their acquisition of state-of-the-art knowledge of vowel acoustics.

**Index Terms**: vowel, formant, spectral shape

## 1. Introduction

In the literature, the primary cues of vowel quality are generally understood as being either contained in formant patterns (*F*-patterns), or, alternatively, in the spectral shape in terms of a derivation of the spectral envelope through some kind of smoothing operation ([1, 2]). Secondary cues that potentially affect vowel sounds and vowel recognition include phonation type, speaker characteristics such as size and gender differences and related fundamental frequency ($f_0$) normalisation, vocal effort, duration, vowel-inherent spectral change, sound context and transitions, formant amplitude, spectral contrast and spectral tilt, and auditory spectral averaging process (for detailed references, see the online presentation, link below). However, such a differentiated view only appertains to the specialist literature on vowel acoustics. In other scientific fields, the acoustic cues of vowel sounds are straight forwardly considered to be primarily contained in the *F*-patterns in terms of peaks in the spectral envelope, patterns that further depend on age, size or gender of a speaker and corresponding differences in vocal tract size. This simplification and generalisation is even observed in textbooks designed for introductory courses to Phonetics.

Yet, contrastively, recent studies culminated in a falsification of both theses of either *F*-patterns or spectral envelopes as being vowel quality-related: neither *F*-patterns nor spectral envelopes as such relate to single vowel qualities but – although in a non-systematic way – to different qualities. First and foremost, this ambiguity phenomenon is a consequence of a complex interaction between pitch and spectral characteristics in perceived vowel quality. This is true for inter- and intra-speaker comparisons of natural vowel sounds as well as for synthesised sounds, and the ambiguity is not limited to adjacent vowel qualities. (For detailed references, see also the online presentation, link below).

Obviously, the demonstration and communication of the ambiguity phenomenon and its acoustic and perceptual context is of high importance and high interest not only to phonetics but also for many other disciplines that deal with speech acoustics. In a previous work based on sounds of previous single studies, we have already given a summary of many of the aspects in question, including extensive illustrations ([3]). However, the sounds these illustrations were based on were recorded under varying conditions and with varying sound qualities, and the rights for online playback could not retrospectively be obtained from all speakers. To study and provide sounds to the scientific community that are recorded under systematically controlled conditions, with extended variation of production parameters and permission for online audio playback, we have created a large new sound corpus termed the Zurich Corpus of Vowel and Voice Quality (henceforth: Zurich Corpus, [4]). In its first version, it consists of about 34 600 sounds of the long Standard German vowels /i–y–e–ø–ɛ–a–o–u/ of 70 speakers (men, women and children, nonprofessionals as well as professional actors/actresses and singers), produced with varying basic production parameters such as phonation type, vocal effort, $f_0$ and vowel context. The corpus includes results of an acoustic analysis and of a listening test. (For details on the method, see [4].)

Against this general background and on the basis of the Zurich Corpus and additional field recordings of speech, this presentation gives exemplary evidence for the argument that neither *F*-patterns nor spectral envelopes can be considered to be the primary cues for vowel recognition in general. Furthermore, it also illustrates the acoustic and perceptual context of the ambiguity phenomenon. Thereby, for the various disciplines related to speech acoustics, it addresses (i) researchers to allow them to re-evaluate results of existing studies and to create experimental settings for future experiments, taking into

account the actual variation and pitch-dependency of the vowel spectrum, and (ii) students to support the acquisition of state-of-the-art knowledge of vowel acoustics.

## 2. Content and form of documentation

The presentation starts with the demonstration of formant pattern and spectral shape ambiguity of vowel sounds for all long Standard German vowels, and it subsequently embeds this phenomenon in the context of a general – but non-systematic – pitch-dependency of vowel-related spectral characteristics and of other aspects of spectral variation. For details on the content, see Table 1 and the following paragraphs. For the online presentation, see:

**http://is2019.phones-and-phonemes.org**

Table 1: *Content of presentation.*

**Part I – Formant pattern and spectral shape ambiguity**

1. Natural sounds and their resynthesis; sounds of different vowels at different $f_o$ with similar *F*-patterns and/or spectral shapes:
   - Ambiguity for /ɛ–e–i/
   - Ambiguity for /ɛ–ø–y/
   - Ambiguity for /a–o–u/
2. Synthesised sounds related to open-tube filter patterns, varying $f_o$:
   - Ambiguity for /ə–ø–y/

**Part II – Context of the ambiguity phenomenon**

3. $f_o$ contour and upper $f_o$ ranges of speech:
   - Everyday speech samples and speech during artistic performance with upper $f_o$ exceeding 350 Hz for men and 500 Hz for women
4. Vowel recognition of natural isolated high-pitched sounds:
   - Sound of all long vowels at $f_o$ = 700–800 Hz
   - Sounds of the corner vowels /i–a–u/ at $f_o$ = 1 kHz
5. Pitch-dependency of the vowel spectrum in natural vocalises:
   - Sounds of all long vowels of a man, a woman and a child with $f_o$ variation (C-major scale) of 22–34 semitones
6. Other aspects of spectral variability for natural sounds of all long vowels:
   - Sounds with different modes of phonation (voiced, breathy, whispered and creaky phonation)
   - Sounds with different vocal effort
7. Non-systematic relation between vowel-related spectral peaks or spectral envelopes for natural sounds of all long vowels:
   - Sounds with different numbers of spectral peaks
   - Sounds with "flat" or "sloping" spectral portions in their vowel-specific frequency range
   - Differences related to vowel quality, range of $f_o$ variation, formant levels, and harmonic configuration

***Part I – Formant pattern and spectral shape ambiguity:*** The sound series in Part I give exemplary evidence for the ambiguity in question: Firstly, comparisons of natural sounds of different front or back vowels are presented, with similar estimated *F*-patterns and spectral envelopes as a direct consequence of $f_o$ differences of the sounds compared (the ambiguity can be replicated by the Klatt resynthesis tool included, see below). Secondly, comparisons of synthesised sounds at different $f_o$ levels but related to equal open-tube filter patterns that are commonly attributed to schwa vowels of men, women or children are presented, demonstrating that, in perception, assumed neutral or centralised articulatory configurations are not consistently related to the neutral vowel schwa.

***Part II – Context of the ambiguity phenomenon:*** The sound series in Part II illustrate the acoustic and perceptual context of the ambiguity phenomenon: wide ranges of observable $f_o$-contours of speech; recognisable high-pitched vowel sounds with upper $f_o$ levels that surpasses $F_1$ of most vowel qualities as reported in literature; changes in the vowel-related spectral envelopes for vowel sounds produced with extensive $f_o$ variation (vocalises); changes in the vowel-related spectral envelope caused by changes in phonation mode and in vocal effort; aspects of the non-systematic relationship between perceived vowel quality, $f_o$, spectral peaks and spectral envelope.

***Form of online presentation of sound series:*** Systematically organised and commented compilations of sounds are presented online with audio playback feature, spectral characteristics (FFT spectra, LPC curves, spectrograms, $f_o$-contours, calculated *F*-patterns) and recognition rate results of the listening test performed. A Klatt synthesiser (browser-inherent tool) is also directly linked to the natural sounds. It allows for a sound resynthesis based on calculated $f_o$ and *F*-patterns of natural sounds and for a sound synthesis based on manipulated values. The presentation will also serve as a basis for future expansion: future versions will include additional sound series.

## 3. Relevance

To the best of our knowledge, the Zurich Corpus represents the most extensive online documentation designed to study the effect of the variation of basic production parameters on vowel quality-related acoustic and perceptual characteristics. As demonstrated in this presentation, it provides an excellent basis for exploring the ambiguity of *F*-patterns and spectral shapes in their relation to vowel quality as well as its broader acoustic and perceptual context. – Up to now, the awareness and discussion of the ambiguity phenomenon was restricted to a limited number of specialists, and there is a lack of direct access to compilations of sound examples which allow for an understanding of the phenomenon. The present online documentation gives insight into the phenomenon for a wider audience of researchers (and also of students) of various fields dealing with speech acoustics: It allows to acquire basic knowledge, to extensively listen to exemplary sound series and to directly compare *F*-patterns and spectral shapes, and to crosscheck the perceptual role of $f_o$ and *F*-patterns in vowel synthesis. This kind of insight encourages a re-evaluation of existing and, most importantly, a conceptual and methodological valuation of future studies that attempt to address vowel acoustics in a general perspective, i.e. including the observable variation of vowel-related spectral characteristics.

## 4. Acknowledgements

This work was supported by the Swiss National Science Foundation SNSF, Grants 100016_143943 / 100016_159350.

## 5. References

[1] J. M. Hillenbrand and R. A. Houde, "A narrow band pattern-matching model of vowel perception", *The Journal of the Acoustical Society of America*, vol. 113, no. 2, 1044–1055, 2003.

[2] R. Swanepoel, D. J. Oosthuizen, and J. J. Hanekom, "The relative importance of spectral cues for vowel recognition in severe noise", *The Journal of the Acoustical Society of America,* vol. 132, no. 4, 2652–2662, 2012.

[3] D. Maurer, *Acoustics of the Vowel – Preliminaries*. Bern: Peter Lang, 2016.

[4] D. Maurer, C. d'Heureuse, H. Suter, V. Dellwo, D. Friedrichs, and T. Kathiresan, T., "The Zurich Corpus of Vowel and Voice Quality, Version 1.0," in *INTERSPEECH 2018 – 19th Annual Conference of the International Speech Communication Association, September 2-6, Hyderabad, India, Proceedings,* 2018, pp. 1417–1421.