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

Interspeech paper

D. Maurer, H. Suter, C. d'Heureuse, V. Dellwo, "Formant pattern and spectral shape ambiguity of vowel sounds and related phenomena of vowel acoustics – Exemplary evidence," in INTERSPEECH 2019 – 20th Annual Conference of the International Speech Communication Association, September 15-19, Graz, Austria, Proceedings, 2019, pp. 2368–2369 (including corrections and additions until 2020-03-11).

New in the version of 2019-12-09: The results of the listening tests in the first chapter are given for each sound pair and sound triplet in details (main examples only).
New in the version of 2020-03-11: New sound series were added to Chapter 3.1.3.
New in the version of 2020-06-19: New sound series were added to Chapters 6.3 and 7.1.1

Introduction

Background, aim and content

In introductory courses to phonetics and in many speech acoustics-related disciplines, the primary acoustic features related to vowel quality are understood to be contained in the formant patterns of the sounds. In the phonetic specialist literature, however, there is an extensive and often controversial debate on whether these cues of vowel quality are contained in either the formant patterns or, alternatively, in the spectral shape.

Yet, earlier studies indicated and recent studies made evident that both theses of either *F*-patterns or spectral envelopes being vowel quality-related are empirically contradicted:

> Neither *F*-patterns nor spectral envelopes as such relate to single vowel qualities but – although in a non-systematic way – to different qualities in terms of **formant pattern and spectral shape ambiguity of vowel sounds**.

First and foremost, the ambiguity phenomenon is a consequence of an inherent pitch-dependency of vowel-related spectral characteristics. This is true for inter- as well as intra-speaker comparisons of vowel sounds and for natural as well as synthesised sounds.

Up to now, the awareness and discussion of the ambiguity phenomenon and its acoustic and perceptual context was restricted to a limited number of specialists, and there is almost a complete lack of direct access to compilations of sound examples which allow for an understanding of the phenomenon. The present online documentation gives exemplary evidence for this phenomenon and its acoustic and perceptual context. The documentation contains systematically organised and commented compilations of sounds with audio playback feature, spectral characteristics (FFT spectra, LPC curves, spectrograms, f_0-contours, calculated formant patterns) and recognition rate results of the listening tests performed. A Klatt synthesiser (browser-inherent tool) is also included and is directly linked to the natural sounds. It allows for a sound resynthesis based on calculated f_0 and *F*-patterns of natural sounds and for a sound synthesis based on manipulated values. The documentation will also serve as a basis for further expansion: future versions will include additional sound series.

For a wider audience of researchers (and also of students) of various fields dealing with speech acoustics, the present online documentation thus gives insight into:

- the ambiguity phenomenon
- its acoustic and perceptual context

The documentation allows:

- to extensively listen to exemplary sound series
- to directly compare the sound-related *F*-patterns, spectral shapes and entire sound spectra with the fine structure of the harmonics
- to crosscheck the perceptual role of f_o and *F*-patterns in vowel resynthesis
- to acquire basic knowledge of the ambiguity phenomenon and of its acoustic and perceptual context

The documentation

- starts with the demonstration of formant pattern and spectral shape ambiguity of vowel sounds, 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 vowel-related spectral variation

The documentation does not only address the field of phonetics but all disciplines related to speech acoustics. It aims at allowing researchers 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 spectral characteristics of vowel sounds. Besides, it also supports students in their acquisition of state-of-the-art knowledge of vowel acoustics.

For each aspect of interest, the documentation gives a chapter with primary examples and additional chapters entitled “Additions”. These additions provide supplementary evidence; In some cases, however, they rely on less restricted conditions of selection than the main examples. – Future versions of the documentation will include further additions, which will be dated.

Please note for sound playback

Use state of the art headphones to listen to the sounds; otherwise, sound quality might be significantly impaired. The use of PC loudspeakers will not suffice.

Table of contents, user guide

For the table of contents, see the bookmarks of this document:
- PDF downloaded and opened in the Acrobat Reader, bookmarks on the left **(recommended)**
- PDF displayed in the browser, bookmarks according to browser-specific layout

For details regarding the use of the documentation, please refer to
- Video (mp4)
- Assistant of the Help menu (sound archive)

- Other items of the Help menu in the sound archive

Speakers, vowel sounds, speech extracts, acoustic analyses, graphic representations, numerical indications, listening test

Speakers, vowel sounds, and sound production parameters: This documentation presents vowel sounds of the eight long Standard German vowels /i–y–e–ø–ɛ–a–o–u/ produced in isolation (V context) or in consonant–vowel–consonant–vowel context (sVsV context, s = /s/ or /z/) by speakers of different gender and age groups (men, women, children) are presented. Sound production includes an extensive variation of production parameters, above all of phonation type, vocal effort and f_o. All speakers and sounds are selected from the Zurich Corpus of Vowel and Voice Quality (for details see Maurer et al., 2018). – Note that the sounds assigned to the vowel quality /a/ include the entire range of back and front vowels /ɑ–a/, due to the varying regional dialects and variants of the speakers.

Please note:

- In the text, IPA symbols /i, y, e, ø, ɛ, a, o, u/ are used to represent vowel qualities of natural sounds. In the sound archive, vowel qualities are indicated orthographically as letters of the German alphabet /i, ü, e, ö, ä, a, o, u/.
- In addition, for synthesised sounds and for the results of the listening tests, IPA symbols /ɔ–ə/ are used to represent open o and schwa according to the standard of the Zurich Corpus.
- In one experimentation investigating open-tube resonance characteristics, /œ/ is also used.
- In the results of the listening tests, recognised vowel categories are given in these symbols, and sounds recognised as lying within a vowel boundary are given by two symbols with no blank in between.
- The presentation of the vowels within a chapter (the order in which they are listed) is organised according to the topic in question.

Speakers and speech extracts: Speakers and speech extracts from various sources are presented: Field recordings, recordings published on the Internet and TV or DVD recordings. – Playback function is only activated for recordings by authors who have given written consent. – Many of the speech extracts are deliberately short in order to be able to present a large number of samples but, at the same time, keep the listening and evaluation effort at a minimum.

Acoustic analysis: Acoustic analysis of isolated vowel sounds (V context) was conducted on the middle 0.3 sec vowel nucleus for a frequency range of 0–5.5 kHz concerning f_o contour, average f_o frequency, average spectrum, spectrogram, average formant patterns (frequencies, bandwidths, levels) and formant tracks. (In addition, the average spectrum was also calculated for a frequency range of 0–11 kHz.) – Concerning formant pattern estimation, LPC analysis (Burg algorithm, window length=25 ms, time steps=5 ms, pre-emphasis=50 Hz) was conducted in parallel for three parameter settings according to three commonly used age and gender-specific standards of 12 (standard for men), 10 (standard for women) and 8 (standard for children) poles for the frequency range of 0–5.5 kHz. – The same analysis was conducted on sVsV sounds for the middle 0.3 sec of the first or second vowel sound, depending on their duration (for details of automatic procedure see Maurer et al., 2018). – The speech extracts were analysed for f_o contour, spectrogram (0–5.5 kHz) and long-term average spectrum (LTAS; 0–5.5 and 0–11 kHz). – The acoustic analysis was conducted with a script using the Praat functionalities (Boersma & Weenink, 2019). – Note that the overall f_o ranges indicated for the speech extracts were determined acoustically in terms of approximations by listening to the sounds. (Please ignore possible f_o calculation errors in the f_o contours displayed that exceed the verified ranges given for the sounds. Among other reasons, these errors are due to background noise/music, the sound of an audience or to automated pitch calculation.)

Graphic representations and numerical indications: For vowel sounds, the graphs include a display of the entire sound wave, sound nucleus analysed, f_o contour, spectrum, spectrogram and formant tracks. In addition, three LPC filter curves (for the three parameter settings mentioned) of the middle window of the sound nucleus was overlaid to the spectrum in order to illustrate the correspondence between spectral peaks and calculated formants. – Numerical average values of f_o and formant patterns were added to the sound information. – The graphs for read or sung material depict sound wave, f_o contour, spectrogram and LTAS.

Listening test: If not noted, recognised vowel qualities are given in terms of the result of a standard listening test in which five professionally trained speakers (singers, actors or voice teachers) listened to every single vowel sound in order to assign a perceived vowel quality (for test details, see Maurer et al., 2018). However, for some sound sample, sample-specific tests were performed and they are described in details in the corresponding chapters.

If not explicitely mentioned, the details of the vowel recognition results can be viewed in the sound legend (see the legend displayed below a sound spectrum on the online presentation, vowel qualities in brakets; vowel recognition is given in phonetic order).

For further details regarding method and the Zurich Corpus the examples were selected from, see:

The Zurich Corpus of Vowel and Voice Quality (Maurer et al., 2018)

Terms, notation and abbreviations

Spectral shape and spectral envelope are used as synonyms according to Hillenbrand and Houde (2003).

Resynthesis is used as a term for a Klatt synthesis (see below) based on calculated LPC values for formant patterns of natural sounds, independent of whether the resynthesis is also based on the calculated f_0 of the natural sounds or whether f_0 is manually altered. Accordingly, synthesis is used as a term for a Klatt synthesis that is not based on calculated LPC values for formant patterns of natural sounds.

As mentioned above:

- In the text, IPA symbols /i–y–e–ø–ɛ–a–o–u/ are used to represent vowel qualities.
- In the sound archive, vowel qualities are indicated orthographically as letters of the German alphabet /i–ü–e–ö–ä–a–o–u/.

Abbreviations used in the text relate to Titze et al. (2015):

f_0	fundamental frequency
F-pattern	formant pattern
$F_{(i)}$	formant frequency (F_1, F_2, …)
$B_{(Fi)}$	formant bandwidth (B_{F1}, B_{F2}, …)
$L_{(Fi)}$	formant level (L_{F1}, L_{F2}, …)

Abbreviations used in the Klatt synthesis form:

F0	fundamental frequency
F(i)	formant frequency (F1, F2, …)
B(i)	formant bandwidth (B1, B2, …)
L(i)	formant level (L1, L2, …)

Klatt synthesis

A Klatt synthesiser (Klatt, 1980, Klatt & Klatt, 1990) is included in this documentation. The corresponding source code was rewritten and improved by Christian d'Heureuse. For details, see

Klatt synthesiser, source code of Christian d'Heureuse, with references and comments

The main purpose to include a Klatt synthesiser is to allow:

- to resynthesise a natural sound based on calculated values for formants and f_0 and to test whether the vowel quality of the natural sound is maintained in resynthesis
- to resynthesise a natural sound based on calculated values for formants but **to alter f_0 and to test whether the vowel quality of the natural sound can be changed by f_0 alteration in terms of formant pattern ambiguity**
- to relate such testing to a comparison of two or several sounds of different vowels with similar formant patterns and spectral shapes

Note that:

- **formant pattern ambiguity in resynthesis corresponds to spectral shape ambiguity** since, in resynthesis, the entire filter curve in terms of a spectral envelope is the same for the sounds compared

Further indications concerning the Klatt functionality implemented:

- For a direct resynthesis, Klatt synthesis in Cascade mode is applied and related to the calculated average formant frequencies and f_0 of the sound in question, the calculated average of f_0 being indicated per default value in the corresponding synthesis field preceding the synthesis player; when a sound is displayed, as default for resynthesis, f_0 is set to the sound-related calculated frequency level; however, f_0 can be manually altered; on the basis of a setting for f_0, resynthesis is effected and the sound is played back via the play button.
- For vowel sounds produced with whisper phonation, the default source for Klatt synthesis is a white noise.
- For vowel sounds produced with creaky phonation, default f_0 level is set to 65 Hz for Klatt synthesis independently of the pitch of the natural sound.
- For additional tweaking of synthesis parameters, click on the KlattSyn link. Via the user interface, Klatt synthesis can be applied in cascade or Parallel mode.

How to re-confirm formant pattern and spectral shape ambiguity of a comparison of natural sounds in resynthesis – Short instruction for resynthesis

Comparison of sounds with similar formant patterns and/or similar spectral envelopes but different f_0 (formant pattern and spectral shape ambiguity of vowel sounds) – short instruction for re-confirmation in resynthesis:

- Listen to a given natural vowel sound of a comparison.
- Resynthesise this sound at its calculated f_0; this f_0 level is by default inserted in the corresponding synthesis field (displayed in the Layouts S, M and L); simply click on the synthesis play button.
- Resynthesise this sound at calculated f_0 of the other sound(s) of the comparison: insert the corresponding f_0 level(s) in the synthesis field for f_0 and click on the synthesis play button.
- Test whether the vowel quality is maintained or altered.
- Repeat these steps with all sounds of comparison.

Sound series highlighted in red

In the listing of sound series below, some series are highlighted in red. For a first experience of sounds related to a specific aspect, we recommend to listen to these series first.

Part I – Formant pattern and spectral shape ambiguity

In Part I, firstly, comparisons of natural sounds of different front or back vowels are presented, with similar calculated *F*-patterns and/or spectral envelopes as a direct consequence of f_o differences of the sounds. This ambiguity can be replicated by the Klatt synthesis tool (see Introduction).

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.

1 Natural sounds and their resynthesis; sounds of different vowels at different f_o with similar *F*-patterns and/or spectral shapes

Major intention: To provide exemplary evidence for formant pattern and spectral shape ambiguity of vowel sounds.

Preliminary notes – vowel recognition:

Vowel qualities of the original natural sounds presented in this chapter were tested according to the standard listening test (see Introduction).

Vowel qualities of the resynthesised sounds (main examples only) were tested in a documentation-specific listening test, in which two sounds were presented as one test item, the first sound being the natural sound, the second sound being a resynthesis of the natural sound on a given f_o level (either the average f_o level of the natural sound replicated or a f_o level of another natural sound of comparison). A 1 sec interval was set in between sounds one and two. Test items were presented in random order, and the listeners were asked to assign a vowel quality to the second sound only. (For the sounds given in the additions, vowel recognition was tested for the natural sounds only. For the results, see the legend below the spectra, displayed when clicking on a link.)

The results of both listening tests are given for each single comparison (see links below; note that the results of the first listening test of the natural sounds are also given in the legend to the vowel spectra displayed when activating a link.)

For almost all sounds, vowel recognition of the natural vowel sounds of the main examples is 100 % (5/5 listeners recognised the quality according to the intention of the speaker). For only five sounds, the recognition rate is 80%, and only one sound is recognised in the boundary of o and ɔ.

The results of vowel recognition are less unambiguous for the resynthesised sounds: For some comparisons of sounds of front vowels, increasing f_o in the Klatt resynthesis produces an open–closed shift in the recognised vowel quality according to the difference of the natural reference sounds, but unrounded–rounded confusions occur; vowel recognition in vowel boundaries as well as confusions of /ø, ɛ, ə/ also occur; for comparisons of the vowel qualities /a–o–u/, the resynthesised sounds of /u/ at high f_o levels are sometimes impaired, and the resynthesised sounds of /u/ at low f_o levels do sometimes not match with the vowel quality of the low-pitched natural vowel sound.

The differences of results of the two listening tests and the sometimes inconsistent results for the resynthesised sounds may be the effect of different causes: The overall medium sound quality produced by Klatt synthesiser, above all when including middle and higher f_o, the Cascade mode applied (not accounting for the actual levels of spectral peaks observed for natural sounds) and, for high-pitched sounds of /u/, the spectral difference between sounds of /o/ and /u/ often being related to the level of the second harmonic; the static character of the resynthesis applied both with regard to f_o and *F*-patterns; the automatic calculation of average formant frequencies and bandwidths with no corrections of formant tracks and crosschecks of the reliability of average bandwidths, and the formant estimation at middle and higher f_o levels as methodically not substantiated.

However, the resynthesis allows for a crosscheck whether a documented similarity of *F*-patterns for natural sounds can be replicated in terms of a general open–closed vowel quality shift with increasing f_o: The shift direction is the main criterion for the present resynthesis approach.

1.1 Ambiguity for sounds of /ɛ–e–i/

1.1.1 Main examples, ambiguity of adjacent vowel qualities, comparisons of sounds of single speakers, range of f_o variation ≥ c. 1 octave

This chapter presents comparisons of sounds of two adjacent vowels /ɛ–e/ or /e–i/ produced by single speakers for which both the calculated vowel-related formant frequencies F_1–F_2–F_3 as well as the vowel-related spectral envelope < 3 kHz for adults and 3.5 kHz for children are similar. For the comparisons, the upper f_o range of the sounds is limited to 400 Hz for men, 450 Hz for women, and 500 Hz for children, and the f_o differences of a sound pair is limited to approximately 1 octave. These limitations are to reflect the range of the average, everyday speaking voice of women and children, and a range of chest and "mixed" voice for men. (For the non-systematic aspect of a lack of ambiguity for f_o below approximately 200 Hz, see Chapter 7.)

As a further condition of the comparison, Klatt resynthesis (Cascade mode) confirms the ambiguity phenomenon: If the sounds are resynthesised based on the calculated f_o and calculated formant patterns of the natural sounds (with speaker group-specific parameters of LPC analysis), the recognised vowel quality of the natural sound is in most cases maintained or it corresponds to the vowel boundary (or boundaries) of the natural sound replicated; but if f_o of the sound(s) of the contrasting vowel quality is taken as the basis for the Klatt resynthesis, the vowel quality changes in open–closed direction with increasing f_o. (For an example, see Chapter 1.1.1, Series A, and 1.1.2, Series A.)

↗ **Series A: Sounds of /e–i/ produced by a woman; f_o differences of the sounds = 167–310 Hz.**

Natural:	1 = /e e e e e/	2 = /i i i i i/
RS: fo of original:	1 = /e e e e e/	2 = /i i i i i/
RS: fo higher/lower:	1 = /i i i i i/ (f_o increased)	2 = /e e e e ei/ (f_o decreased)

Legend. *Natural: 1, 2* = vowel recognition of the two natural sounds of comparison (five assignments per sound of the five listeners, see listening tests); *RS: f_o of original* = Klatt resynthesis at average f_o of the natural sounds; *RS: f_o higher/lower* = Klatt resynthesis at average f_o of the opposed sounds; 1 = vowel recognition of the first sound of a pair; 2 = vowel recognition of the second sound.

Example for replication in Klatt resynthesis (valid for all other main examples with the limitations mentioned):

- If the natural sound of /e/ (sound 1) is resynthesised on the basis of the related calculated formant pattern (speaker group-specific parameters, for women = P05) and at original f_o of 167 Hz, the vowel quality of /e/ is maintained (see the detailed result of the listening test for *RS: f_o* of original: 1 = /e e e e e/, the vowel qualities given by the five listeners separately:).
- But if, in synthesis, f_o is set to 310 Hz, the vowel quality changes to a more closed vowel, here to /i/ (see the detailed result of the listening test for *RS: fo higher/lower:* 1 = /i i i i i/).
- If the natural sound of /i/ (sound 2) is resynthesised on the basis of the related calculated formant pattern and at original f_o of 310 Hz, the vowel quality of /i/ is maintained (f_o of original: 2 = /i i i i i/.
- But if, in synthesis, f_o is set to 167 Hz, the vowel quality changes to a more open vowel, here to /e/ (*RS: fo higher/lower:* 2 = /e e e e ei/).

Series B: Sounds of /ε–e/ produced by a woman; f_0 of the sounds = 181–412 Hz.

Natural:	1 = /ε ε ε ε ε/	2 = /e e e e e /
RS: fo of original:	1 = /ε ε ε ε ε/	2 = /ε e e e e/
RS: fo higher/lower:	1 = /ε e e e e/	2 = /ε ε ε ε ε/

Series C: Sounds of /ε–e/ produced by a woman; f_0 of the sounds = 183–388 Hz.

Natural:	1 = /ε ε ε ε ε/	2 = /e e e e e /
RS: fo of original:	1 = /ε ε ε ε ε/	2 = /ε e e e e/
RS: fo higher/lower:	1 = /e e e e e/	2 = /ε ε ε ε ε/

Series D: Sounds of /ε–e/ produced by a woman; f_0 of the sounds = 197–326 Hz.

Natural:	1 = /ε ε ε ε ε/	2 = /e e e e e /
RS: fo of original:	1 = /ɔ ε ε ε ε/	2 = /ε εe e e e/
RS: fo higher/lower:	1 = /εe e e e e/	2 = /ε ε ε ε e/

Series E: Sounds of /e–i/ produced by a man; f_0 of the sounds = 165–346 Hz.

Natural:	1 = /e e e e e/	2 = /i i i i i/
RS: fo of original:	1 = /e e e e ei/	2 = /ü i i i i/
RS: fo higher/lower:	1 = /üi i i i i/	2 = /e e e e e/

Series F: Sounds of /e–i/ produced by a child; f_0 of the sounds = 217–429 Hz.

Natural:	1 = /e e e e e/	2 = /i i i i i/
RS: fo of original:	1 = /o e e e e/	2 = /i i i i i/
RS: fo higher/lower:	1 = /ü i i i i/	2 = /e e e e e/

1.1.2 Main examples, ambiguity of adjacent and non-adjacent vowel qualities, range of f_0 variation > 1 octave

This chapter presents comparisons of sounds of adjacent and non-adjacent vowels /ε–e–i/ produced by single speakers or by several speakers of an age and gender-related speaker group for which the vowel-related spectral envelope < 3 kHz for adults and 3.5 kHz for children are similar. – For these comparisons, the upper range of f_0 and the range of f_0 variation are not limited. – The conditions of Klatt resynthesis replication accord to Chapter 1.1.1.

Series A: Sounds of /ε–e–i/ produced by men; f_0 of the sounds = 99–263–539 Hz. – Note: For the Klatt resynthesis of the sound of /ε/, please set F1 to 500 Hz and B1 to 90 Hz in the Klatt user interface, according to the vowel spectrum.

Natural:	1 = /ε ε ε ε ε/	2 = /e e e e e/	3 = /i i i i i/
RS: fo of original:	1 = /ε ε ε ε e/	2 = /e e e e e/	3 = /y i i i i/
RS: fo higher/lower:	1 = /e e e e e/–/y y i i i/	2 = /ε ε ε ε ε/–/y y i i i/	3 = /ε ε ε ε ε/–/e e e e e/

Legend. *Natural 1, 2, 3* = vowel recognition of the three natural sounds of comparison (five assignments per sound of the five listeners, see listening tests); *RS: f_0 of original* = Klatt resynthesis at average f_0 of the natural sounds; *RS: f_0 higher/lower* = Klatt resynthesis at average f_0 of the opposed sounds; 1 = vowel recognition of the first sound of a triplet; 2 = vowel recognition of the second sound; 3 = vowel recognition of the third sound. Note for a triplet: For sound 1, both f_0 levels of the opposed sounds are higher (263 Hz and 539 Hz in this example); for sound 2, one f_0 level is lower and the other higher (99 Hz and 539 Hz in this example); for sound 3, both f_0 levels of the opposed sounds are lower (99 Hz and 263 Hz in this example).

Series B: Sounds of /ε–e–i/ produced by women; f_0 of the sounds = 204–421–646 Hz.

Natural:	1 = /ε ε ε ε ε/	2 = /e e e e e/	3 = /i i i i i/
RS: fo of original:	1 = /ε ε ε ε ε/	2 = /ε e e e e/	3 = /e i i i i/
RS: fo higher/lower:	1 = /ε e e e e/–/e e ei i i/	2 = /ε ε ε ε ε/–/i i i i i/	3 = /ε ε ε ε ε/–/e e e e e/

Series C: Sounds of /ε–e–i/ produced by women; f_0 of the sounds = 256–359–663 Hz.

Natural:	1 = /ε ε ε ε ε/	2 = /εe e e e e/	3 = /i i i i i/
RS: fo of original:	1 = /ε ε ε ε ε/	2 = /ε ε e e e/	3 = /i i i i i/
RS: fo higher/lower:	1 = /ε ε e e e/–/e i i i i/	2 = /ε ε ε ε ε/–/i i i i i/	3 = /ε ε ε ε ε/–/e e e e e/

1.1.3 Additions I – Additional sound comparisons

This chapter presents sound comparisons of the same kind as given in the previous two chapters. However, no limitations were applied concerning the upper range of f_o, the range of f_o variation, the vowel context, the number of speakers and of the speaker groups for the sounds compared. (Nevertheless, in most cases, sounds produced in V context by speakers of one age and gender-related speaker group are compared.) Moreover, the similarity of the formant patterns and/or the spectral envelopes are sometimes less pronounced. – The conditions of Klatt resynthesis replication accord to Chapter 1.1.1, sometimes limited to the open-closed or closed-open shift direction.

Series A: Sounds of /e–i/ produced by a man; f_o of the sounds = 174–392 Hz; vowel context V and sVsV.

Series B: Sounds of /e–i/ produced by a woman; f_o of the sounds = 179–350 Hz.

Series C: Sounds of /e–i/ produced by a woman; f_o of the sounds = 221–443 Hz.

Series D: Sounds of /e–i/ produced by a child; f_o of the sounds = 224–435 Hz.

Series E: Sounds of /ɛ–e–i/ produced by women; f_o of the sounds = 170–261–532 Hz.

Series F: Sounds of /ɛ–e–i/ produced by women; f_o of the sounds = 177–257–582 Hz.

1.2 Ambiguity for sounds of /ɛ–ø–y/

1.2.1 Main examples, ambiguity of adjacent vowel qualities, comparisons of sounds of single speakers, range of f_o variation ≥ c. 1 octave

This chapter presents comparisons of sounds of two adjacent vowels /ɛ–ø/ or / ø–y/ produced by single speakers for which both the calculated vowel-related formant frequencies F_1–F_2–F_3 as well as the vowel-related spectral envelope < 3 kHz for adults and 3.5 kHz for children are similar. – Limitations for f_o and conditions of Klatt resynthesis accord to Chapter 1.1.1.

Series A: Sounds of /ɛ–ø/ produced by a man; f_o of the sounds = 163–322 Hz. – Note: Resynthesis of the sound of /ø/ at f_o of the sound of /ɛ/ relates to schwa; for a corresponding improved comparison of sounds of two men, see the Additions section.

Natural:	1 = /ɛ ɛ ɛ ɛ ɛ/	2 = /ø ø ø ø ø/
RS: fo of original:	1 = /ø ø ə ə ə/	2 = /ə ə ø ø ø/
RS: fo higher/lower:	1 = /ɛ ə ø ø ø/	2 = /ə ə ə ø ø/

Series B: Sounds of /ɛ–ø/ produced by a woman; f_o of the sounds = 170–263 Hz.

Natural:	1 = /ɛ ɛ ɛ ɛ ɛ/	2 = /ø ø ø ø ø/
RS: fo of original:	1 = /ɛ ɛ ɛ ɛ ø/	2 = /ø ø ø ø ø/
RS: fo higher/lower:	1 = /ə ø ø ø e/	2 = /ɛ ə ɛø ø ø/

Series C: Sounds of /ɛ–ø/ produced by a woman; f_o of the sounds = 212–295 Hz.

Natural:	1 = /ɛ ɛ ɛ ɛ ɛ/	2 = /ø ø ø ø ø/
RS: fo of original:	1 = /ɛ ɛ ɛ e e/	2 = /ø ø ø ø ø/
RS: fo higher/lower:	1 = /ɛ ɛ øe e e/	2 = /ə ə ø ø ø/

Series D: Sounds of /ø–y/ produced by a man; f_o of the sounds = 163–330 Hz.

Natural:	1 = /ø ø ø ø ø/	2 = /y y y y y/
RS: fo of original:	1 = /ø ø ø ø øy/	2 = /y y y y y/
RS: fo higher/lower:	1 = /y y y y y/	2 = /ø ø ø ø øy/

Series E: Sounds of /ø–y/ produced by a woman; f_0 of the sounds = 221–423 Hz.

Natural:	1 = /ø ø ø ø ø/	2 = /y y y y y/
RS: fo of original:	1 = /ø ø ø ø ø/	2 = /u y y y y/
RS: fo higher/lower:	1 = /y y y y y/	2 = /ø ø ø ø øy/

Series F: Sounds of /ø–y/ produced by a woman; f_0 of the sounds = 221–444 Hz.

Natural:	1 = /ø ø ø ø ø/	2 = /y y y y y/
RS: fo of original:	1 = /ø ø ø ø ø/	2 = /y y y y y/
RS: fo higher/lower:	1 = /y y y y y/	2 = /ø øy øy y y/

Series G: Sounds of /ø–y/ produced by a woman; f_0 of the sounds = 222–405 Hz.

Natural:	1 = /ø ø ø ø ø/	2 = /y y y y y/
RS: fo of original:	1 = /ə ə ø ø ø/	2 = /y y y y y/
RS: fo higher/lower:	1 = /ø øy y y y/	2 = /ə ø ø ø ø/

Series H: Sounds of /ø–y/ produced by a child; f_0 of the sounds = 226–436 Hz.

Natural:	1 = /ø ø ø ø ø/	2 = /y y y y y/
RS: fo of original:	1 = /ø ø ø ø ø/	2 = /y y y y y/
RS: fo higher/lower:	1 = /y y y y y/	2 = /ø ø ø ø ø/

1.2.2 Main examples, ambiguity of adjacent and non-adjacent vowel qualities, range of f_0 variation > 1 octave

This chapter presents comparisons of sounds of adjacent and non-adjacent vowels /ɛ–ø–y/ produced by single speakers or by several speakers of an age and gender-related speaker group for which the vowel-related spectral envelope < 3 kHz for adults and 3.5 kHz for children are similar. – For these comparisons, the upper range of f_0 and the range of f_0 variation are not limited. – The conditions of Klatt resynthesis replication accord to Chapter 1.1.1.

Series A: Sounds of /ɛ–ø–y/ produced by men; f_0 of the sounds = 84–252–512 Hz. – Note: Sounds of /ø–y/ schwa-like.

Natural:	1 = /ɛ ɛ ɛ ɛ ɛ/	2 = /ø ø ø ø ø/	3 = /øy y y y y/
RS: fo of original:	1 = /ə ɛ ɛ ɛ ɛ/	2 = /ø ø ø ø ø/	3 = /y y y y y/
RS: fo higher/lower:	1 = /ə ə ø ø ø/–/ə ø øy y y/	2 = /ə ə ə ø ø/–/ø y y y y/	3 = /ə ə ə ø ø/–/ø ø ø ø ø/

Series B: Sounds of /ɛ–ø–y/ produced by a man; f_0 of the sounds = 107–339–574 Hz.

Natural:	1 = /ɛ ɛ ɛ ɛ ɛ/	2 = /ø ø ø ø ø/	3 = /y y y y y/
RS: fo of original:	1 = /ɛ ɛ ɛ ɛ ɛ/	2 = /ø ø ø ø ø/	3 = /ø øy y y y/
RS: fo higher/lower:	1 = /ɛ ɛ ɛø ø ø/–/ə ø øy øy y/	2 = /ə ɛ ɛ ɛ ɛ/–/ə ø øy y y/	3 = /ə ɛ ɛ ø y/–/ə ø ø ø ø/

Series C: Sounds of /ɛ–ø–y/ produced by men; f_0 of the sounds = 109–348–622 Hz.

Natural:	1 = /ɛ ɛ ɛ ɛ ɛ/	2 = /ø ø ø ø ø/	3 = /y y y y y/
RS: fo of original:	1 = /ɛ ɛ ɛ ɛ ɛ/	2 = /ɛ ə ø ø e/	3 = /ø y y y y/
RS: fo higher/lower:	1 = /ɛ ə ø ø e/–/ø ø øy øy y/	2 = /ɛ ɛ ɛ ɛ ɛ/–/ø e y y y/	3 = /ɛ ɛ ɛ ɛ ø/–/ø ø ø ø ø/

Series D: Sounds of /ɛ–ø–y/ produced by men; f_0 of the sounds = 124–239–507 Hz.

Natural:	1 = /ɛ ɛ ɛ ɛ ɛ/	2 = /ɛø ø ø ø ø/	3 = /y y y y y/
RS: fo of original:	1 = /ɛ ɛ ɛ ɛ ɛ/	2 = /ø ø ø ø ø/	3 = /øy y y y y/
RS: fo higher/lower:	1 = /ə ø ø ø ø/–/øy øy y y y/	2 = /ɛ ɛ ə ø ø/–/ø y y y y/	3 = /ə ə ə ø ø/–/ø ø ø ø ø/

1.2.3 Additions

This chapter presents sound comparisons of the same kind as given in the previous two chapters However, no limitations were applied, and the similarity of the formant patterns and/or the spectral envelopes are sometimes less pronounced (see Chapter 1.1.3). – The conditions of Klatt resynthesis replication accord to Chapter 1.1.1, sometimes limited to the open-closed or closed-open shift direction.

Series A: Sounds of /ε–ø/ produced by a man; f_o of the sounds = 127–240 Hz.

Series A: Sounds of /ε–ø/ produced by two men; f_o of the sounds = 163–339 Hz.

Series A: Sounds of /ε–ø–y/ produced by women and children; f_o of the sounds = 174–255–501 Hz.

1.3 Ambiguity for sounds of /a, o, u/

1.3.1 Ambiguity of adjacent vowel qualities, comparisons of sounds of single speakers, range of f_o variation ≥ c. 1 octave, main examples

This chapter presents sound comparisons of two adjacent vowels /a–o/ or /o–u/ produced by single speakers for which both the calculated vowel-related formant frequencies F_1–F_2 as well as the vowel-related spectral envelope < 2 kHz are similar. – Limitations for f_o and conditions of Klatt resynthesis accord to Chapter 1.1.1.

Series A: Sounds of /a–o/ produced by a man; f_o of the sounds = 161–317 Hz.

Natural:	1 = /a a a a a/	2 = /o o oɔ oɔ ɔ/
RS: fo of original:	1 = /a a a a a/	2 = /o o ɔ ɔ ɔa/
RS: fo higher/lower:	1 = /o oɔ ɔ ɔ a/	2 = /a a a a a/

Series B: Sounds of /o–u/ produced by a woman; f_o of the sounds = 186–395 Hz.

Natural:	1 = /o o o o o/	2 = /u u u u u/
RS: fo of original:	1 = /uo o o o o/	2 = /u u u u uo/
RS: fo higher/lower:	1 = /u u u u u/	2 = /uo o o o o/

Series C: Sounds of /o–u/ produced by a man; f_o of the sounds = 195–392 Hz.

Natural:	1 = /o o o o o/	2 = /u u u u u/
RS: fo of original:	1 = /o o o o o/	2 = /u u u u u/
RS: fo higher/lower:	1 = /u u u u o/	2 = /uo o o o o/

Series D: Sounds of /o–u/ produced by a child; f_o of the sounds = 198–401 Hz.

Natural:	1 = /o o o o o/	2 = /u u u u u/
RS: fo of original:	1 = /o o o o o/	2 = /u u u u u/
RS: fo higher/lower:	1 = /u u u u u/	2 = /o o o o o/

Series E: Sounds of /o–u/ produced by a woman; f_o of the sounds = 215–401 Hz.

Natural:	1 = /o o o o o/	2 = /u u u u u/
RS: fo of original:	1 = /o o o o o/	2 = /u u u u u/
RS: fo higher/lower:	1 = /u u u u u/	2 = /o o o o o/

Series F: Sounds of /o–u/ produced by a man; f_o of the sounds = 219–392 Hz.

Natural:	1 = /o o o o o/	2 = /u u u u u/
RS: fo of original:	1 = /o o o o oɔ/	2 = /u u u u u/
RS: fo higher/lower:	1 = /u u u uo o/	2 = /o o o o o/

Series G: Sounds of /o–u/ produced by a woman; f_o of the sounds = 223–435 Hz.

Natural:	1 = /o o o o o/	2 = /u u u u u/
RS: fo of original:	1 = /uo o o o o/	2 = /u u u u u/
RS: fo higher/lower:	1 = /u u u u u/	2 = /u uo o o o/

Series H: Sounds of /o–u/ produced by a child; f_0 of the sounds = 256–489 Hz.

Natural:	1 = /o o o o o/	2 = /u u u u uo/
RS: fo of original:	1 = /uo o o o o/	2 = /u u u u o/
RS: fo higher/lower:	1 = /u u u u u/	2 = /o o o o ɔ/

Note: One previously given example of a sound pair of /a–o/ produced by a woman at F_1 of 185–396 Hz was deleted because of a lack of replication in Klatt synthesis.

1.3.2 Ambiguity of adjacent and non-adjacent vowel qualities, range of f_0 variation > 1 octave, main examples

This chapter presents comparisons of sounds of adjacent and non-adjacent vowels /a–o–u/ produced by single speakers or by several speakers of an age and gender-related speaker group for which both the calculated vowel-related formant frequencies F_1–F_2 as well as the vowel-related spectral envelope < 2 kHz are similar. – For these comparisons, the upper range of f_0 and the range of f_0 variation are not limited. – The conditions of Klatt resynthesis replication accord to Chapter 1.1.1 (note the limitation for high-pitched sounds of /u/ mentioned).

Series A: Sounds of /a–o–u/ produced by two men; f_0 of the sounds = 107–327–595 Hz.
Note: Sounds of /a–o/ produced by one man.

Natural:	1 = /a a a a a/	2 = /o o o o o/	3 = /u u u u u/
RS: fo of original:	1 = /a a a a a/	2 = /o o o oɔ ɔa/	3 = /u u u uo uo/
RS: fo higher/lower:	1 = /o o o ɔ ɔa/–/u uo ɔ a a/	2 = /a a a a a/–/u u uo o o/	3 = /ɔ a a a a/–/o o o o oɔ/

Series B: Sounds of /a–o–u/ produced by men; f_0 of the sounds = 129–292–584 Hz.

Natural:	1 = /a a a a a/	2 = /o o o o oɔ/	3 = /u u u u u/
RS: fo of original:	1 = /a a a a a/	2 = /o o o ɔ ɔ/	3 = /u u uo o o/
RS: fo higher/lower:	1 = /o o ɔ a a/–/u ɔ ɔ ɔa a/	2 = /ɔ a a a a/–/u u u o o/	3 = /ɔ a a a a/–/uo o o o o/

Series C: Sounds of /a–o–u/ produced by a man; f_0 of the sounds = 131–291–575 Hz.

Natural:	1 = /a a a a a/	2 = /o o o o o/	3 = /u u u u u/
RS: fo of original:	1 = /ɔ ɔa ɔa a a/	2 = /o o o o oɔ/	3 = /u u u uo o/
RS: fo higher/lower:	1 = /o o ɔ ɔ a/–/u u u o ɔ/	2 = /ɔ a a a a/–/u u u uo o/	3 = /ɔa a a a a/–/u uo o o o/

Series D: Sounds of /a–o–u/ produced by women; f_0 of the sounds = 163–300–622 Hz.

Natural:	1 = /a a a a a/	2 = /o o o o o/	3 = /u u u u u/
RS: fo of original:	1 = /ɔa ɔa a a a/	2 = /o o o ɔ ɔ/	3 = /u u u u uo/
RS: fo higher/lower:	1 = /o o ɔ ɔ ɔ/–/u u uo ɔ ə/	2 = /a a a a ə/–/u u u uo o/	3 = /a a a a ə/–/o o o ɔ ɔ/

1.3.3 Additions

This chapter presents sound comparisons of the same kind as given in the previous two chapters. However, no limitations were applied, and the similarity of the formant patterns and/or the spectral envelopes are sometimes less pronounced (see Chapter 1.1.3). – The conditions of Klatt resynthesis replication accord to Chapter 1.1.1, sometimes limited to the open-closed or closed-open shift direction.

Series A: Sounds of /a–o/ produced by two men; f_0 of the sounds = 98–349 Hz.

Series B: Sounds of /a–o/ produced by two men; f_0 of the sounds = 130–344 Hz.

Series C: Sounds of /a–o/ produced by two women; f_0 of the sounds = 157–333 Hz.

Series D: Sounds of /a–o/ produced by a man; f_0 of the sounds = 164–341 Hz.; vowel context V and sVsV.

Series E: Sounds of /a–o/ produced by two women; f_0 of the sounds = 182–322 Hz.

Series F: Sounds of /a–o/ produced by a woman; f_0 of the sounds = 182–322 Hz; vowel context V and sVsV.

Series G: Sounds of /o–u/ produced by a woman; f_0 of the sounds = 184–353 Hz.

Series H: Sounds of /o–u/ produced by two men; f_0 of the sounds = 185–389 Hz.

Series I: Sounds of /o–u/ produced by a man; f_0 of the sounds = 192–385 Hz.

Series K: Sounds of /o–u/ produced by a woman; f_0 of the sounds = 225–443 Hz.

Series L: Sounds of /a–o–u/ produced by three men; f_0 of the sounds = 96–257–544 Hz.

Series M: Sounds of /a–o–u/ produced by women; f_0 of the sounds = 262–431–654 Hz.

1.4 f_0 and/or pitch and vowel quality-related spectral frequency ranges

When examining the sound spectra of front vowels produced at different f_0 in vocalises of single speakers, as discussed below in Chapter 5, marked shifts of the spectral peak below 1.5 kHz can be observed. However, in many cases, the higher spectral part does not, in parallel, manifest also a marked variation. This observation let us assume that the vowel spectrum is dichotomous: while the lower vowel-related frequency range of the spectrum is f_0 or pitch-dependent, the higher frequency range is not (see also a corresponding remark in the Preliminaries).

This assumption was tested on the bases of the above sound pairs produced by single speakers (see Chapters 1.1.1, 1.2.1 and 1.3.1). For each sound of a natural sound pair of two different vowels, in addition to the two resynthesised sounds, a third sound was synthesised (Klatt synthesis, Cascade mode), for which F_1 – and in some cases of sounds of back vowels also F_2 – was adjusted in parallel to a decrease or increase of f_0. In this way, the vowel quality of the natural sound was attempted to be maintained despite a marked difference f_0 of the natural and the synthesised sound.

The details and the produced sound series are given in a separate document entitled “Spectral frequency ranges of the vowel spectrum and f_0”:

Spectral frequency ranges of the vowel spectrum and f_0

2 Synthesised sounds related to single open-tube filter patterns but different levels of f_o: Ambiguity for /ə–ø–y/

Major intention: To provide exemplary evidence for resonance patterns of open tubes that are not consistently related to the neutral vowel schwa in vowel recognition.

Preliminary note – vowel recognition test:

For the present synthesis experiment, vowel recognition was tested applying an experiment-specific test procedure. Involving the same five listeners and the same hardware as for the standard test, three subtests were conducted as is described below. The assignments of four listeners were given orally and were transcribed by the first author. The assignments of the author was added and is given below as the result for the listener “1”.

In addition to the standard listening test, the vowel /œ/ was included as an option in the test, because two listeners insisted on the recognition of this vowel quality or of a vowel boundary including this quality.

Subset 1: In the first subset, all nine synthesised sounds of the following three chapters were presented in random order. In a first round, the listeners were asked to listen to the sounds in order to become familiar. In a second round, for each single vowel sound presented separately, they were asked to vocally repeat (imitate) the sound (ignoring the pitch) and then to assign a vowel quality or a boundary of adjacent vowel qualities. During this test, the listeners could ask for a repetition of the sound playback as many times as they wanted.

Subset 2: In the second subset, the sound triplet related to a single resonance pattern was presented separately, the three sounds presented in the order of ascending f_o. In a first round, the listeners were asked to listen to a sound triplet. In a second round, they were asked to vocally repeat (imitate) all three sounds of a triplet (ignoring the pitch) and then to assign them to three vowel qualities or vowel boundaries. Again, the listeners could ask for a repetition of the sound playback as many times as they wanted.

Subset 3: The third subset is a repetition of the second test, the sounds of a triplet presented in the order of descending f_o.

For each of the three subsets, the results are given below in Tables 1 to 3: For each listener (1 to 5), each level of f_o and each of the three subtests (single sounds presented in random order, sound triplets presented in ascending and descending order), the recognised vowel qualities are given. In the column “All”, the assignments of all five listeners are given in phonetic order. In the column “Maj.”, the majority of assignments is given in terms of either a single vowel quality or a vowel boundary of two qualities (both qualities given with no blank space in between) if 3/5 listeners in minimum gave identical assignments. If no majority was found, the range of vowel qualities is given (the most open and closed qualities of the range given with “–“ inserted).

2.1 Ambiguity for sounds related to an open–tube filter pattern generally attributed to an average vocal tract size of men

This chapter presents three synthesised sounds that relate to a filter pattern generally assumed to be representative of open-tube resonance characteristics for a vocal tract that is neutral and of the average size of males.

Sounds were synthesised using the Klatt synthesiser (Cascade mode, sample frequency = 44.1 kHz) with F_1–F_2–F_3–F_4–F_5 = 500–1500–2500–3500–4500 Hz. All bandwidths were set to 100 Hz. Sounds were synthesised on three levels of f_o being 1/1, ½ and 1/3 of the first formant frequency, i.e. f_o = 125–250–500 Hz. In consequence, formant frequencies always match with frequencies of harmonics in the sound spectrum.

Table 1: Results of the vowel recognition test, indicating an opened–closed shift with rising f_o. However, vowel recognition depended somewhat from the sound context in the subtests (see also the results in 2.2 and 2.3).

f f_o	F_1	F_2	F_3	F_4	F_5		Vowel recognition per listener							
Hz	Hz	Hz	Hz	Hz	Hz			1	2	3	4	5	All	Maj.
125	500	1500	2500	3500	4500		Single sounds	ə	ə	ə	œ	ə	ə ə ə ə œ	ə
							Sound triplets, f_o ↑	ə	ə	ə	ə	ə	ə ə ə ə ə	ə
							Sound triplets, f_o ↓	ə	ə	ə	ə	ə	ə ə ə ə ə	ə
250							Single sounds	ø	ø	ø	ø	œ	œ ø ø ø ø	ø
							Sound triplets, f_o ↑	ø	ø	ø	ø	œ	œ ø ø ø ø	ø
							Sound triplets, f_o ↓	ø	ø	ø	ø	œ	œ ø ø ø ø	ø
500							Single sounds	øy	ø	øy	y	øy	ø øy øy øy y	øy
							Sound triplets, f_o ↑	y	y	y	y	y	y y y y y	y
							Sound triplets, f_o ↓	y	y	y	y	øy	y y y y øy	y

Related sounds (click on the Play button in the screen displayed):

- Sound A: f_o = 125 Hz, recognised vowel = ə (overall majority).
- Sound B: f_o = 250 Hz, recognised vowel = ø (overall majority).
- Sound C: f_o = 500 Hz, recognised vowel = y (subtest 1 = øy; subtests 2 and 3 = y).

2.2 Ambiguity for sounds related to an open–tube filter pattern generally attributed to an average vocal tract size of women

This chapter presents three synthesised sounds that relate to a filter pattern generally assumed to be representative of open-tube resonance characteristics for a vocal tract that is neutral and of the average size of females.

Sounds were synthesised using a Klatt synthesiser (Cascade mode, sample frequency = 44.1 kHz) with F_1–F_2–F_3–F_4–F_5 = 600–1800–3200–4000–5400 Hz. All bandwidths were set to 100 Hz. Sounds were synthesised on three levels of f_o being 1/1, 1/2 and 1/3 of the first formant frequency, i.e. f_o = 200–300–600 Hz. In consequence, formant frequencies always match with frequencies of harmonics in the sound spectrum.

Table 2: Results of the vowel recognition test, again indicating an opened–closed shift with rising f_o.

f_o	F_1	F_2	F_3	F_4	F_5		Vowel recognition per listener							
Hz	Hz	Hz	Hz	Hz	Hz			1	2	3	4	5	All	Maj.
200	600	1800	3000	4200	5400		Single sounds	ə	ə	ə	œ	ə	ə ə ə ə œ	ə
							Sound triplets, f_o ↑	ə	ə	ə	ə	ə	ə ə ə ə ə	ə
							Sound triplets, f_o ↓	ə	ə	ə	ə	ə	ə ə ə ə ə	ə
300							Single sounds	ø	ə	ø	ø	œ	ə œ ø ø ø	ø
							Sound triplets, f_o ↑	ø	əø	ø	ø	œ	əø œ ø ø ø	ø
							Sound triplets, f_o ↓	ø	ø	ø	ø	ə	ə ø ø ø ø	ø
600							Single sounds	øy	ø	ø	y	øy	ø ø øy øy y	ø–y
							Sound triplets, f_o ↑	y	y	y	y	øy	øy y y y y	y
							Sound triplets, f_o ↓	y	y	y	y	øy	øy y y y y	y

Related sounds (click on the Play button in the screen displayed):

- Sound A: f_o = 200 Hz, recognised vowel = ə (overall majority).
- Sound B: f_o = 300 Hz, recognised vowel = ø (overall majority).
- Sound C: f_o = 600 Hz, recognised vowel = y (subtext 1 = ø–y; subtests 2 and 3 = y).

2.3 Ambiguity for sounds related to an open–tube filter patterns often attributed to an average vocal tract size of children

This chapter presents three synthesised sounds that relate to a filter pattern generally assumed to be representative of open-tube resonance characteristics for a vocal tract that is neutral and of the average size of children.

Sounds were synthesised using a Klatt synthesiser (Cascade mode, sample frequency = 44.1 kHz) with F_1–F_2–F_3–F_4–F_5 = 700–2100–3500–4900–6300 Hz. All bandwidths were set to 100 Hz. Sounds were synthesised on three levels of f_o being 1/1, 1/2 and 1/3 of the first formant frequency, i.e. f_o = 233–350–700 Hz. In consequence, formant frequencies always match with frequencies of harmonics in the sound spectrum.

Table 3: Results of the vowel recognition test. An opened–closed shift with rising f_o was confirmed, but vowel recognition was less clear than was found for the previous two open tube resonance patterns investigated.

f_o	F_1	F_2	F_3	F_4	F_5		**Vowel recognition per listener**							
Hz	Hz	Hz	Hz	Hz	Hz			1	2	3	4	5	All	Maj.
233	700	2100	3500	4900	6300		Single sounds	ε	ε	ə	ε	ə	ə ə ε ε ε	ε
							Sound triplets, f_o ↑	ε	ε	ə	ε	ə	ə ə ε ε ε	ε
							Sound triplets, f_o ↓	ε	ε	ə	ε	ə	ə ə ε ε ε	ε
350							Single sounds	øe	əø	əε	œ	ə	ə əø əε œ øe	ə–øe
							Sound triplets, f_o ↑	øe	ə	ø	ε	ə	ə ə ε ø øe	ə–øe
							Sound triplets, f_o ↓	øe	ø	ø	ε	œ	œ ε ø ø øe	œ–øe
700							Single sounds	øy	y	y	øy	y	øy øy y y y	y
							Sound triplets, f_o ↑	y	y	y	øy	ø	øy øy y y y	y
							Sound triplets, f_o ↓	y	y	y	y	y	y y y y y	y

Related sounds (click on the Play button in the screen displayed):

- Sound A: f_o = 233 Hz, recognised vowel = ε (overall majority).
- Sound B: f_o = 350 Hz, recognised vowel = ø (subtests 1 and 2 = ə–øe; subtest 3 = œ–øe).
- Sound C: f_o = 700 Hz, recognised vowel = y (overall majority).

Part II – Context of the ambiguity phenomenon

In Part II, the acoustic and perceptual context of the ambiguity phenomenon is illustrated: wide ranges of observable f_0-contours of speech; recognisable high-pitched vowel sounds with upper f_0 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_0 variation (vocalises); changes in vowel-related spectral envelope caused by changes in phonation and in vocal effort; and aspects of the non-systematic relationship between f_0, spectral peaks and spectral envelope.

3 f_0 contour and upper f_0 ranges in natural speech: Pitch contours of speech to observe in everyday life and in artistic performance, with upper f_0 exceeding 350 Hz for men and 500 Hz for women and children

Major intention: To provide exemplary evidence for f_0 contour and upper f_0 range of natural speech substantially exceeding the frequencies of the first formant for closed vowels (and sometimes also for mid-open vowels) as generally given in studies of average formant frequencies for vowel sounds produced in isolation or in the context of citation-form words. (For Standard German vowels, see Pätzold and Simpson, 1997.)

The examples given below indicate that formant pattern and spectral shape ambiguity is not just a phenomenon occurring in rare cases of very high-pitched vowel sounds in singing but that it also occurs in the middle frequency range of everyday speech for women and children as is the case for the frequency range of "mixed" voice and of falsetto of men.

Note that for speech extracts without play back mode, the player is disabled and only the graphical illustration of the pitch contours is shown.

3.1 Pitch contours of speech extracts produced by male speakers, with upper f_0 exceeding 350 Hz

3.1.1 Untrained male speakers, journalists, politicians, and TV hosts, main examples

This chapter presents speech extracts of generally untrained male speakers (meaning not professionally trained as singers or actors/actresses) as observed in everyday life, including journalists and TV hosts, with pitch contours exceeding statistical F_1 of the vowels /i, y, u/ for men, that is exceeding 350 Hz.

Series A: Extracts of speech of a French criminologist in a TV discussion; range of f_0 = c. 110–440 Hz; language = fr.
Source: France 5, TV broadcast "C'est dans l'air", 2019-01-12.

Series B: Extracts of infant directed speech; range of f_0 = c. 150–600 Hz; language = en.
Source: BBC News, Special Report (2013), Youtube https://www.youtube.com/watch?v=nza_Rxg44a8 (Retrieved 2019-03-23)

Series C: Extracts of a political speech; range of f_0 = c. 150–470 Hz; language = fr.
Source: https://www.youtube.com/watch?v=WoLZYEO2-FQ (Retrieved 2016-12-17; link not active anymore).

Series D: Extracts of a sports commentator commenting on a football match; range of f_0 = c. 110–440 Hz, language = fr.
Source: http://www.culturepsg.com/news/match/psg-barca-4-0-le-resume-video/14149 (Retrieved 2017-02-15)

3.1.2 Actors performing on stage or in films, including comic and voice-over performances, main examples

This chapter presents speech extracts of actors performing on stage or in films, with pitch contours exceeding statistical F_1 of the vowels /i, y, u/ for men, that is exceeding 350 Hz.

Series A: Extracts of speech of a comic actor hosting an Oscar Award; f_0 = c. 150–900 Hz; language = en.
Source: https://www.youtube.com/watch?v=ez6H7pdckDc (Retrieved 2019-04-18).
Note: 16 extracts; use the next page button on the top–left of the layout to view all examples.

Series B: Extracts of speech of a comic actor performing as a female character; range of f_0 = c. 220–780 Hz, language = fr.
Source: https://www.rts.ch/play/tv/divertissement/video/la-revue-de-marie-therese-2012?id=4545635 (Retrieved 2019-04-18)

Series C: Extracts of speech of an Indonesian comic actor performing in a Drama Gong; range of f_0 = c. 300–600 Hz; language = id.
Source: Proprietary recording.

Series D: Extracts of speech of a French comic actor performing on stage; range of f_0 = c. 130–700 Hz, language = fr.
Source: https://www.youtube.com/watch?v=fuWkguCHpFY (Retrieved 2019-04-18).

Series E: Extracts of speech of a comic actor performing in a TV show; range of f_0 = c. 110–490 Hz, language = en.
Source: https://www.youtube.com/watch?v=BZykxMLbbrM (Retrieved 2019-04-18).

Series F: Extracts of speech of a comic actor performing on stage; range of f_0 = c. 130–620 Hz (except the cries); language = fr.
Source: https://www.youtube.com/watch?v=oofJZpX50FI (Retrieved 2019-04-18).

3.1.3 Additions

This chapter presents similar examples of speech as in the previous two chapters, uploaded after the first publication date of this presentation.

Series A: Extracts of speech of a voice imitator acting in a radio broadcast; range of f_0 = c. 90–580 Hz; language = fr.
Sounds 1 and 2: f_0 = c. 130–580 Hz; imitation of a French female TV host.
Sound 3: f_0 = c. 110–490 Hz; imitation of a German female politician.
Sounds 4 and 5: middle f_0 = c. 200 Hz; imitation of a French male specialist in economics and a French male sport journalist; illustration of men's voices with f_0 levels corresponding to average f_0 as given for female speakers in formant statistics.
Sound 6: f_0 = c. 90–490 Hz; imitation of a French male politician; illustration of men's voice switching from chest to falsetto register.
Sounds 7–10: main f_0 range below 200 Hz; imitation of four French male politicians; illustration of men's voices in a lower range of f_0.
Sources (retrieved 2020-02-26):
https://www.europe1.fr/emissions/La-revue-de-presque/best-of-angela-merkel-quand-on-vient-a-paris-on-veut-voir-les-specialites-francaises-comme-la-greve-canteloup-3950411 (sounds 3–5, 8, 10)
https://www.europe1.fr/emissions/La-revue-de-presque/best-of-cest-presque-lagriculture-3950826 (sound 7)
https://www.europe1.fr/emissions/La-revue-de-presque/emmanuel-macron-sur-les-municipales-a-paris-agnes-buzyn-peut-nous-faire-une-remontada-canteloup-3951360 (sounds 1, 6, 9)
https://www.europe1.fr/emissions/La-revue-de-presque/philippe-douste-blazy-sur-les-mesures-contre-le-coronavirus-cest-anne-hidalgo-qui-a-invente-le-confinement-a-paris-canteloup-3951587 (sound 2)

Series B: Extracts of speech of a well known Turkish male politician; range of f_0 = c. 130–600 Hz; language = tr.
Sounds 1 and 2 (short extract of 1): f_0 = c. 200–350 Hz; speech in a political rally.
Sound 3: f_0 = c. 100–260 Hz; interview given for a journal.
Sound 4: f_0 = c. 330–620 Hz; speech in a meeting of supporters, given in falsetto voice.
Sources (retrieved 2020-03-03):
https://www.sabah.com.tr/video/haber/son-dakika-baskan-erdogandan-ak-parti-il-danisma-meclis-toplantisinda-onemli-aciklamalar-video (sounds 1 and 2)
https://www.youtube.com/watch?v=W8Y4Hr-noDU (sound 3)
https://www.youtube.com/watch?v=AMLaqNDEM-c (sound 4)

3.2 Pitch contours of speech extracts produced by female speakers, with upper f_0 exceeding 500 Hz

3.2.1 Untrained female speakers, journalists, politicians, and TV hosts, main examples

This chapter presents speech extracts of untrained female speakers, including journalists and TV hosts, with pitch contours exceeding statistical F_1 of the vowels /i, y, u/ (400 Hz) and of the vowels /e, ö, o/ (500 Hz; see Pätzold and Simpson, 1997, for a reference) for women, that is exceeding 500 Hz.

Series A: Extracts of speech of a market woman selling grilled chicken in a market in Paris; range of f_0 = c. 220–700 Hz (excluding very high-pitched exclamations); language = fr.
Source: Proprietary recordings.

Series B: Extracts of speech of a woman demonstrating infant drected speech; range of f_0 = c. 200–950 Hz; language = en.
Source: BBC News, Special Report (2013), https://www.youtube.com/watch?v=nza_Rxg44a8 (Retrieved 2019-04-18).

Series C: Extracts of speech of a medical doctor talking on television; range of f_0 = c. 250–520 Hz; language = fr.
Source: France 2 (TV station), broadcast “13 Heures”, 2012-10-03.

3.2.2 Actresses performing on stage or in films, including comic and voice-over performances, main examples

This chapter presents main examples of speech extracts of actresses performing on stage or in films, with pitch contours exceeding statistical F_1 of the vowels /i, y, u/ (400 Hz) and of the vowels /e, ö, o/ (500 Hz; see Pätzold and Simpson, 1997, for a reference) for women, that is exceeding 500 Hz.

Series A: Extracts of speech of a narratress of fairy tales; range of f_0 = c. 150–950 Hz; language = de-CH.
Source: Trudi Gerster: “Zwerge Märli”. Zurich, Tudor Recordings, 2007.
Note: 8 extracts; use the next page button on the top–left of the layout to view all examples.

Series B: Extracts of speech of a comic actress performing on stage; range of f_0 = c. 200–850 Hz; language = fr.
Source: https://www.youtube.com/watch?v=AvM6MuV5XA0 (Retrieved 2019-04-18).

Series C: Extracts of speech of a comic actress performing on stage, range of f_0 = c. 200–880 Hz; language = fr.
Source: https://www.youtube.com/watch?v=ZQG9tdfLDME (Retrieved 2019-04-18).

Series D: Extracts of speech of an actress in an animated television series (voice-over); range of f_0 = c. 220–700 Hz.; language = en.
Source: https://www.dailymotion.com/video/x6nciws (Retrieved 2019-04-18).

Series E: Extracts of speech of an actress in an animated television series (voice-over); range of f_0 = c. 200–650 Hz (except one higher exclamation); language = fr.
Source: https://www.youtube.com/watch?v=CVqIu0fy9k8 (Retrieved 2019-04-18).

Series F: Extracts of speech of an actress in an animated television series (voice-over); range of f_0 = c. 200–700 Hz; language = fr.
Source: https://www.youtube.com/watch?v=CVqIu0fy9k8 (Retrieved 2019-04-18).

3.2.3 Additions

This chapter presents similar examples of speech as in the previous two chapters. However, for most of the examples, the permission for sound playback could not be obtained.

(To come in future versions)

3.3 Pitch contours of speech extracts produced by child speakers, with upper fo exceeding 500 Hz (additions)

3.2.1 Untrained child speakers (additions)

This chapter presents speech extracts of untrained child speakers. (Examples added after the first publication of this documentation).

Series A: Extracts of speech of female child speaker speaking at a festival; main range of f_0 = c. 400–660 Hz; language = unknown; added 2019-09-26.
Source: https://www.youtube.com/watch?v=oa71E4xA0g8, Retrieved 2019-09-22.

3.2.2 Trained child speakers (additions)

(To come in future versions)

4 Vowel recognition of natural isolated high-pitched sounds

> Major intention: To provide exemplary evidence of successful vowel recognition for high-pitched vowels with f_0 exceeding the F_1 frequencies that are generally indicated in studies of average formant frequencies for vowel sounds produced in isolation or in the context of citation-form words. (For Standard German, see Pätzold and Simpson, 1997.)

4.1 Sounds of /y, e, ø, ε, o/ at f_0 = 700–800 Hz, main examples

This chapter presents intelligible high-pitched sounds of the vowel /y, e, ø, ε, o / in the f_0 range of 700–800 Hz. Note also the vowel-related spectral differences.

Series A: Sounds of /y/.

Series B: Sounds of /e/.

Series C: Sounds of /ø/.

Series D: Sounds of /ε/.

Series E: Sounds of /o/.

4.2 Sound of the corner vowels /i, a, u/ at f_0 = c. 1 kHz Hz, main examples

This chapter presents intelligible high-pitched sounds of the corner vowels /i, a, u / at f_0 of c. 1 kHz. Note also the vowel-related spectral differences.

Series A: Sounds of /i/.

Series B: Sounds of /a/.

Series C: Sounds of /u/.

4.3 Additions

This chapter presents similar examples of speech as given in the previous two chapters.

(To come in future versions)

5 f_0/pitch-dependency of the vowel spectrum in natural vocalises: Sounds of /i, y, e, ø, ε, a, o, u/ of a man, a woman and a child with f_0 variation (C-major scale) of 34, 31 and 22 semitones

Major intention: To provide exemplary evidence for the f_0/pitch-dependency of the vowel-related spectrum and the vowel-related formants < 1.5–2 kHz.

5.1 Adult male speaker, range of intended f_0 = 110–784 Hz (34 semitones, according to C-major scale)

For each of the eight long Standard German vowels, this chapter presents a series of sounds produced by a man with extensive f_0 variation (vocalises, f_0 variation according to C-major scale, medium vocal effort).

Series A: Sounds of /i/.

Series B: Sounds of /y/.

Series C: Sounds of /e/.

Series D: Sounds of /ø/.

Series E: Sounds of /ε/.

Series F: Sounds of /a/.

Series G: Sounds of /o/.

Series H: Sounds of /u/.

5.2 Adult female speaker, range of intended f_0 = 131–784 Hz (31 semitones, according to C-major scale)

For each of the eight long Standard German vowels, this chapter presents a series of sounds produced by a woman with extensive f_0 variation (vocalises, f_0 variation according to C-major scale, medium vocal effort).

Series A: Sounds of /i/.

Series B: Sounds of /y/.

Series C: Sounds of /e/.

Series D: Sounds of /ø/.

Series E: Sounds of /ε/.

Series F: Sounds of /a/.

Series G: Sounds of /o/.

Series H: Sounds of /u/.

5.3 Child (male) speaker, range of intended f_0 = 220–784 Hz (22 semitones, according to C-major scale)

For each of the eight long Standard German vowel, this chapter presents a series of sounds produced by a child with extensive f_0 variation (vocalises, f_0 variation according to C-major scale, medium vocal effort).

Series A: Sounds of /i/.

Series B: Sounds of /y/.

Series C: Sounds of /e/.

Series D: Sounds of /ø/.

Series E: Sounds of /ɛ/.

Series F: Sounds of /a/.

Series G: Sounds of /o/.

Series H: Sounds of /u/.

5.4 Additions

This chapter presents comparable examples of vocalises including variation of vocal effort.

(To come in future versions)

6 Other aspects of spectral variability of natural sounds related to single vowel categories

Major intention: To provide exemplary evidence for changes in the vowel-related spectral envelope caused by changes in phonation and in vocal effort.

6.1 Sounds of /i, y, e, ø, ɛ, a, o, u/ produced with different modes of phonation

6.1.1 Sounds produced by an adult male speaker

For each of the eight long Standard German vowels, this chapter presents a series of sounds produced by a man with voiced, breathy, creaky and whisper phonation. f_o of the sounds with voiced phonation = c. 131 Hz, and f_o of the sounds with breathy phonation = c. 165 Hz (values according to C-major scale).

Series A: Sounds of /i/.

Series B: Sounds of /y/.

Series C: Sounds of /e/.

Series D: Sounds of /ø/.

Series E: Sounds of /ɛ/.

Series F: Sounds of /a/.

Series G: Sounds of /o/.

Series H: Sounds of /u/.

6.1.2 Sounds produced by an adult female speaker

For each of the eight long Standard German vowels, this chapter presents a series of sounds produced by a woman with voiced, breathy, creaky and whisper phonation. f_o of the sounds with voiced phonation = c. 220 Hz, and f_o of the sounds with breathy phonation = c. 349 Hz (values according to C-major scale).

Series A: Sounds of /i/.

Series B: Sounds of /y/.

Series C: Sounds of /e/.

Series D: Sounds of /ø/.

Series E: Sounds of /ɛ/.

Series F: Sounds of /a/.

Series G: Sounds of /o/.

Series H: Sounds of /u/.

6.1.3 Sounds produced by a child (male) speaker

For each of the eight long Standard German vowel, this chapter presents a series of sounds produced by a child with voiced, breathy, creaky and whisper phonation. f_0 of the sounds with voiced phonation = c. 262 Hz, and f_0 of the sounds with breathy phonation = c. 262 Hz (values according to C-major scale).

Series A: Sounds of /i/.

Series B: Sounds of /y/.

Series C: Sounds of /e/.

Series D: Sounds of /ø/.

Series E: Sounds of /ɛ/.

Series F: Sounds of /a/.

Series G: Sounds of /o/.

Series H: Sounds of /u/.

6.1.4 Additions

This chapter presents similar sound series as the previous three chapters.

(To come in future versions)

6.2 Sounds of /i, y, e, ø, ɛ, a, o, u/ produced with different vocal effort at similar levels of f_0

6.2.1 Comparison of sounds with low and high vocal effort in the lower f_0 range of men

For each of the eight long Standard German vowel, this chapter presents a series of sounds produced by men in their lower range of f_0 = c. 131–165 Hz (f_0 as intended by the speaker; values in Hz according to C-major scale), with low and high vocal effort.

Series A: Sounds of /i/.

Series B: Sounds of /y/.

Series C: Sounds of /e/.

Series D: Sounds of /ø/.

Series E: Sounds of /ɛ/.

Series F: Sounds of /a/.

Series G: Sounds of /o/.

Series H: Sounds of /u/.

6.2.2 Comparison of sounds with low and high vocal effort in the lower f_o range of women

For each of the eight long Standard German vowel, this chapter presents a series of sounds produced by women in their lower range of f_o = c. 220–262 Hz (f_o as intended by the speaker; values in Hz according to C-major scale), with low and high vocal effort.

Series A: Sounds of /i/.

Series B: Sounds of /y/.

Series C: Sounds of /e/.

Series D: Sounds of /ø/.

Series E: Sounds of /ɛ/.

Series F: Sounds of /a/.

Series G: Sounds of /o/.

Series H: Sounds of /u/.

6.2.3 Comparison of sounds with low vocal effort and shouted vowels in the middle f_o range of men

For each of the eight long Standard German vowel, this chapter presents a series of sounds produced by men in their middle range of f_o = c. 330 Hz (f_o as intended by the speaker; values in Hz according to C-major scale), with low vocal effort and with shouting.

Series A: Sounds of /i/.

Series B: Sounds of /y/.

Series C: Sounds of /e/.

Series D: Sounds of /ø/.

Series E: Sounds of /ɛ/.

Series F: Sounds of /a/.

Series G: Sounds of /o/.

Series H: Sounds of /u/.

6.2.4 Comparison of sounds with low vocal effort and shouted vowels in the middle f_0 range of women

For each of the eight long Standard German vowel, this chapter presents a series of sounds produced by women in their middle range of f_0 = c. 440 Hz (f_0 as intended by the speaker; values in Hz according to C-major scale), with low vocal effort and with shouting.

Series A: Sounds of /i/.

Series B: Sounds of /y/.

Series C: Sounds of /e/.

Series D: Sounds of /ø/.

Series E: Sounds of /ɛ/.

Series F: Sounds of /a/.

Series G: Sounds of /o/.

Series H: Sounds of /u/.

6.2.5. Additions

Comparison of sounds with low and high vocal effort in the lower f_0 range of children

For each of the eight long Standard German vowel, this chapter presents a series of sounds produced by children in their lower range of f_0 = c. 262–294 Hz (f_0 as intended by the speaker; values in Hz according to C-major scale), with low and high vocal effort.

Series A: Sounds of /i/.

Series B: Sounds of /y/.

Series C: Sounds of /e/.

Series D: Sounds of /ø/.

Series E: Sounds of /ɛ/.

Series F: Sounds of /a/.

Series G: Sounds of /o/.

Series H: Sounds of /u/.

Comparison of sounds with low vocal effort and shouted vowels in the middle and higher f_o range of children

For each of the eight long Standard German vowel, this chapter presents a series of sounds produced by children in their middle and higher range of f_o = c. 494–587 Hz (f_o as intended by the speaker; values in Hz according to C-major scale), with low vocal effort and with shouting.

Series A: Sounds of /i/.

Series B: Sounds of /y/.

Series C: Sounds of /e/.

Series D: Sounds of /ø/.

Series E: Sounds of /ɛ/.

Series F: Sounds of /a/.

Series G: Sounds of /o/.

Series H: Sounds of /u/.

7 Non-systematic relation between vowel quality-related spectral peaks or spectral envelopes for natural and synthesised sounds

Major intention: To provide exemplary evidence for aspects of the non-systematic relationship between f_0, spectral peaks and spectral envelope, i.e., sounds with different numbers of vowel-related spectral peaks, sounds with "flat" or sloping vowel-related spectral portions, absence of f_0-dependence of vowel spectra and formant patterns and, consequently, absence of formant pattern and spectral shape ambiguity for f_0 < c. 200–250 Hz for all vowels, and cases of absence of formant pattern and spectral shape ambiguity for a substantial part of sounds of the vowels /ɛ–a/.

7.1 Sounds of /i, y, e, ø, ɛ, a, o, u/ with different numbers of spectral peaks

7.1.1 Main examples

For each of the eight long Standard German vowels, this chapter presents a series of sounds produced by different speakers of different speaker groups with a varying number of spectral peaks.

Series A: Sounds of /i/ with either one or no spectral peak < 1 kHz or one or two spectral peaks in the frequency range of 2–4 kHz.

Series B: Sounds of /y/ with either one or no spectral peak < 1 kHz or one or two spectral peaks in the frequency range of 1.5–3 kHz.

Series C: Sounds of /e/ with either one or two (whispered and breathy sounds) spectral peaks < 1 kHz or one or two spectral peaks in the frequency range of 2–4 kHz.

Series D: Sounds of /ø/ with either two (whispered sounds) or one spectral peak < 1 kHz, or with one or two spectral peaks in the frequency range of 1.5–3 kHz.

Series E: Sounds of /ɛ/ with either two (breathy sounds) or one or no spectral peak < 1 kHz, or with one or two spectral peaks in the frequency range of 2–4 kHz.

Series F: Sounds of /a/ with one to four spectral peaks < 1.5 kHz.

Series G: Sounds of /o/ with one to three spectral peaks < 1.5 kHz.

Series H: Sounds of /u/ with one or two spectral peaks < 1.5 kHz.

7.1.2 Additions

This chapter presents similar sound series as the previous chapter.

Series A: Sounds of /a/ produced by a single female speaker; below 1.5 kHz, vowel spectra show either two peaks (as "expected"), or three, or one, or they harmonic configuration is "flat" or "sloping", or the first harmonic is dominant and only a single second peak is manifest.

Series B: High-pitched sounds of /i/ which manifest no spectral peak > 1.5 kHz.

7.2 Sounds of /i, y, e, ø, ɛ, a, o, u/ with "flat" or "sloping" spectral portions in their vowel-specific frequency range

7.2.1 Main examples

For each of the eight long Standard German vowels except /u/, this chapter presents a series of sounds produced by different speakers of different speaker groups with "flat" or "sloping" spectral portions in their vowel-specific frequency range, lacking a clearly determinable spectral peak (ignoring the first harmonic). – For sounds of the vowel /u/, see the previous chapter.

Series A: Sounds of /i/ with "flat" spectral portions > 1.5 kHz.

Series B: Sounds of /y/ with "flat" or "sloping" spectral portions in either the entire frequency range < 2.5 kHz or the frequency range 1–2.5 kHz.

Series C: Sounds of /e/ with "flat" spectral portions > 1.5 kHz.

Series D: Sounds of /ø/ with "flat" or "sloping" spectral portions in either the entire frequency range < 2.5 kHz or the frequency range 1–2.5 kHz.

Series E: Sounds of /ɛ/ with "flat" or "sloping" spectral portions in either the entire frequency range < 3 kHz or the frequency range 1–3 kHz.

Series F: Sounds of /a/ with "flat" or "sloping" spectral portions < 1.5 kHz.

Series G: Sounds of /o/ with "flat" or "sloping" spectral portions < 1.5 kHz.

7.2.2 Additions

This chapter presents similar sound series as the two previous chapters.

Series A: Additional sounds of /y/ with "flat" or "sloping" spectral portions in either the entire frequency range < 2.5 kHz or the frequency range 1–2.5 kHz.

Series B: Additional sounds of /e/ with "flat" or "sloping" spectral portions in either the entire frequency range < 2.5 kHz or the frequency range 1–2.5 kHz.

Series C: Additional sounds of /ø/ with "flat" or "sloping" spectral portions in either the entire frequency range < 2.5 kHz or the frequency range 1–2.5 kHz.

Series D: Additional sounds of /a/ with "flat" or "sloping" spectral portions < 1.5 kHz.

Series E: Additional sounds of /a/ of a single child speaker demonstrating spectral variability < 2 kHz.

7.3 Sounds of /i, y, e, ø, o, u/ at f_0 < 250 Hz and sounds of /ε, a/ with extensive f_0 variation and no substantial variation of *F*-patterns and spectral envelopes

7.3.1 Sounds of /i, y, e, ø, o, u/ with f_0 variation of c. 1 octave in the frequency range < 250 Hz without substantial variation of *F*-patterns and spectral envelopes (lack of formant pattern ambiguity for f_0 variation in the lower frequency range)

For each of the closed and mid-closed vowels /i, y, e, ø, o, u/, this chapter presents sounds produced by single speakers with f_0 below 250 Hz with maintained vowel quality and similar formant patterns and spectral shapes despite a f_0 variation of approximately one octave or more.

Series A: Sounds of /i/ produced by a man; f_0 differences of the sounds = 107–249 Hz.

Series B: Sounds of /y/ produced by a man; f_0 differences of the sounds = 106–222 Hz.

Series C: Sounds of /e/ produced by a man; f_0 differences of the sounds = 100–198 Hz.

Series D: Sounds of /ø/ produced by a man; f_0 differences of the sounds = 108–219 Hz.

Series E: Sounds of /o/ produced by a man; f_0 differences of the sounds = 108–216 Hz.

Series F: Sounds of /u/ produced by a man; f_0 differences of the sounds = 110–243 Hz.

7.3.2 Sounds of of /ε, a/ with extensive f_0 variation without substantial variation of spectral envelopes (lack of formant pattern ambiguity in relation to vowel quality and spectral shape)

For each of the two vowels /ε, a/, this chapter presents sounds produced by several speakers with similar formant patterns and/or spectral shapes despite a difference in f_0 of approximately 2 octaves.

Series A: Sounds of /ε/ produced by two men and one woman; f_0 differences of the sounds = 116–630 Hz.

Series B: Voiced sounds of /a/ produced by four women; f_0 differences of the sounds = 142–660 Hz.

7.3.3 Additions

This chapter presents similar sound series as the previous chapter.

(To come in future versions)

References (Chapters 1 to 7)

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

Boersma, Paul & Weenink, David (2019). Praat: doing phonetics by computer [Computer program]. Version 6.0.50, retrieved 31 March 2019 from http://www.praat.org/.

Klatt, D. H. (1980). Software for a cascade/paralell formant synthesizer. Journal of the Acoustical Society of America, 67(3), 971–995.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. Journal of the Acoustical Society of America (Vol. 87).

Maurer, D. (2016): Acoustics of the Vowel – Preliminaries. Bern/Frankfurt: Peter Lang.

Maurer, D., d'Heureuse, C., Suter, H., Dellwo, V., Friedrichs, D., Kathiresan, T. (2018) The Zurich Corpus of Vowel and Voice Quality, Version 1.0. INTERSPEECH 2018 – 98th Annual Conference of the International Speech Communication Association, September 2-6, Hyderabad, India, Proceedings, 2018, 1417–1421.

D. Maurer, D., Suter, H., d’Heureuse, C., Dellwo, V. (2019). Formant pattern and spectral shape ambiguity of vowel sounds and related phenomena of vowel acoustics – Exemplary evidence. INTERSPEECH 2019 – 20th Annual Conference of the International Speech Communication Association, September 15-19, Graz, Austria, Proceedings, 2019 (in print).

Pätzold, M., & Simpson, A. P. (1997). Acoustic analysis of German vowels in the Kiel Corpus of Read Speech. Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung Universität Kiel, 32, 215–247.

Titze, I. R., Baken, R. J., Bozeman, K. W., Granqvist, S., Henrich, N., Herbst, C. T., … Kent, R. D. (2015). Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization. The Journal of the Acoustical Society of America, 137(5), 3005–3007.

Appendix

As stated in the Interspeech 2019 paper, in the literature, the primary cues of vowel quality are generally understood as being either contained in formant patterns, or, alternatively, in the spectral shape in terms of a derivation of the spectral envelope through some kind of smoothing operation. 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_o) normalisation, vowel-inherent spectral change, sound context and transitions, formant amplitude, spectral contrast and spectral tilt, and auditory spectral averaging process.

To provide more detailed references, excerpts of Maurer (2018), newly edited and extended, are given below.

Reference: Maurer, D. (2018). Why a phenomenology of vowel sounds is needed. In Belz, M., Mooshammer, C., Fuchs, S., Jannedy, S., Rasskazova, O., & Żygis, M. (2018), Proceedings of the Conference on Phonetics & Phonology in German-speaking countries (P&P 13), pp. 121–124.

Why a phenomenology of vowel sounds is needed (excerpts)

Phonetic summaries generally state that vowel sounds exhibit spectral peaks (termed formants) as the primary acoustic and perceptual cue for the perceived vowel quality, and that these peaks are the consequence of vowel-specific resonance characteristics of the vocal tract. However, different conceptual understandings of formants exist side by side, and there is an extensive and often controversial debate in the literature addressing topics that are considered either as aspects of methodology, or as additional cues, or as aspects that are difficult to understand in the framework of a formant concept. (For overviews, see e.g. Harrington, 2012, Ciocca and Whitehill, 2013, Kiefte et al., 2013; for overviews and exemplary discussions of single aspects, see the corresponding references given below).

Formant concept: As Titze et al. (2015) state: “Unfortunately, the common definition between a formant and a resonance is yet to be established.” Above all, formants are understood in terms of either resonances of the vocal tract, or peaks of the spectral envelope, or filters resulting from an acoustic analysis and related to a corresponding algorithm (e.g. LPC; see Wolfe, n.d.).

Formant estimation: Up to now, no objective method of formant estimation exists, regardless of the algorithm applied: formant patterns are generally estimated by means of an interactive measurement procedure involving general phonetic knowledge, the selection of a software (see e.g., Burris et al, 2014, examining the accuracy and compatibility of different acoustic analysis software packages), an analytical skill of the examiner, context information (size and gender of the speaker), visual crosschecks of calculated values on the basis of the sound spectrum and spectrogram, sometimes related to changes of parameter settings and recalculation of the patterns, and manual interpolations of calculated formant tracks (see e.g., Hillenbrand et al., 1995). Even though LPC analysis has replaced spectrographic measurement, there remains an inherent circularity in the method of formant pattern estimation (Ladefoged, 1967, Hillenbrand et al., 1995). In addition to this general methodological problem, incongruency between the expected and the actual numbers of spectral peaks occurs, understood as “formant merging” or as “spurious formants” (Ladefoged, 2003, pp. 114–115, 119–120). Further, and most important, formant estimation loses methodological substantiation with increasing fundamental frequency (f_o). Some scholars consider the critical f_o frequency level as corresponding to approximately half of the first formant frequency (F_1) of a sound (see Ladefoged, 1967, pp. 80–81, Sundberg, 1987, pp. 124–125; Swerdlin et al., 2010; see also Burriset al., 2014, demonstrating formant measurement is less accurate for sounds of corner vowels produced in citation-form words and at corresponding f_o levels by females and children than for comparable sounds produced men), others assume an f_o level in the F_1 region of the closed vowels, i.e. an f_o level of c. 350 Hz (Monsen & Engebretson, 1983, Fereirra, 2007) as representing that limit. (Note also, that formant bandwidth measurements using acoustic analysis software packages proved to be highly inaccurate.)

Formants and additional cues: The debate on additional cues that potentially affect the acoustics of vowel sounds and the perception and recognition of vowel quality, concerns different types of phonation (see Swerdlin et al., 2010), speaker characteristics (above all size and gender differences; see Johnson, 2008) and f_o (see Cheveigné & Kawahara, 1999, Barreda & Nearey, 2012), vocal effort (see Liénard & Di Benedetto, 1999), duration (Hillenbrand et al., 2000), vowel-inherent spectral change, context and transitions (see Morrison & Assmann, 2013), formant amplitude (see Kiefte et al., 2010), spectral contrast and spectral tilt (see Liu & Eddins, 2008), and auditory spectral averaging process (see Chistovich & Lublinskaya, 1979).

Aspects difficult to understand in the framework of formants: Besides the lack of an objective method to estimate formant patterns, the debate on aspects that are difficult to understand in the framework of a formant concept concerns, above all, the lack of evidence that the data reduction process, implied by this concept, corresponds to the auditory processing of speech sounds, as well as observed nonlinearities in the relation between shifts of formant frequencies and shifts in the perceived vowel quality (Bladon, 1982), and the lack of evidence for a peak picking mechanism of perception as indicated by recognisable vowel sounds with suppressed single formants (Ito et al., 2001) or flat spectra (Carpenter & Morton, 1962, Gooding, 1986, Maurer, 2016, pp. 147–157; Maurer &Suter, 2017a; see also Ito, 2001).

Formants versus spectral shape: Referring to Hillenbrand et al. (1999), Hillenbrand and Houde (2003) and Swanepoel et al. (2012), we conclude that the entire debate on the multitude of aspects mentioned and their often controversial appraisal still have left us with only two main viewpoints, that the major acoustic and perceptual cues are contained in either the formants – more precisely the formant frequency patterns (Carlson et al, 1975, Kasturi et al., 2002) – or, alternatively, in the spectral shape (Bladon, 1982, Zahorian & Jagharghi, 1993; for a relativisation of a complete opposition, see Ito et al., 2001, Molis, 2005), all other aspects of minor or additional effect. Thereby, spectral shape is commonly understood as the pitch-independent envelope of the spectrum derived from some kind of smoothing operation (Hillenbrand et al., 1999, Hillenbrand and Houde, 2003).

Methodological limitations of spectral envelope estimation: With rising f_o, as is true for formant estimation, spectral smoothing becomes also problematic because of spectral undersampling and interrelated distortions. The problem is severe for $f_o \geq 300$ Hz (Cheveigné & Kawahara, 1999, Hillenbrand, 2003).

Formants and f_o: Most scholars conclude for a marginal or very limited effect of f_o on vowel quality of sounds of speakers equal in size and gender (see Cheveigné & Kawahara, 1999, Barreda & Nearey, 2012).However, most of the studies related to this conclusion reported values for f_o variation investigated for the production of natural sounds or for vowel synthesis below 300 Hz. Yet, the few studies which included higher f_o levels, concluded for a substantial effect of f_o on vowel recognition (Fujisaki & Kawashima, 1968; Traunmüller 1981, 1988, Hirahara & Kato, 1992, Maurer & Landis, 1995, 1996, Ménard et al, 2002, Maurer, 2016, pp. 158–169). This finding was either interpreted as calling for some kind of intrinsic normalisation of f_o and formants, possibly also related to paralinguistic variation of vocal effort, or as an indication of pitch-related spectral representation of vowel quality, a perspective adopted here.

"Oversinging" F_1 as generally given in formant statistics: Pätzold and Simpson (1997) reported statistical F_1 for six of the eight long Standard German vowels /i-y-e-ø-o-u/ < 400 Hz for men, and < 450 Hz for women. Summarising studies on vowel recognition in Western classical singing, Sundberg (2013, pp. 86–88) concluded that recognition can be maintained for all vowels up to f_o of C5 (523 Hz). Studies on vowel sounds produced in other artistic styles or involving untrained speakers, however, showed even higher f_o limits for general vowel recognition up to f_o in the range of 660–1046 Hz (dependent on the conditions of vowel production and of the listening tests; see Smith & Scott, 1980, Maurer & Landis, 1996, Maurer et al., 2014, Friedrichs et al., 2015, Maurer, 2016, pp. 158–166). Further, the corner vowels /i, a, u/ were found to be recognisable up to c. 1 kHz (Friedrichs et al., 2017). All these studies show that, at least for a substantial part of vowels of a language, they can be produced and recognised on f_o above statistical F_1 obtained for relaxed speech or for citation-form words.

"Oversinging" the f_o frequency limit for formant and spectral envelope estimation: The finding that vowel sounds can generally be recognised at f_o of c. 600 Hz and even above indicates a discrepancy between perception and methods of acoustic analysis: vowels can be recognised at pitches for which no formant frequency and no spectral envelope estimation is methodologically substantiated; further, the assumption of a direct relation between "spectral undersampling" and degradation of vowel quality perception (Ryalls & Lieberman, 1982) is also contradicted.

Significance of extensive f_o variation in vowel production and perception: There is a strong tendency in the phonetic literature to describe the acoustic characteristics of vowel sounds on f_o levels related to relaxed speech and to citation-form words, and to consider extensive f_o variation as a phenomenon of either size and age-differences of the speakers, or specific (strong) emotions, or singing. Moreover, most investigations on singing concern the Western classical singing style. However, we assume that the significance of f_o variation should be reflected on differently: (i) f_o ranges of speakers different in size and gender substantially overlap. (ii) There is no pitch-related difference of spoken and sung vowels, and in art, the transition between speaking and singing can be fluid (traditional Chinese opera style may serve here as an excellent example). (iii) Western classical opera style cannot be regarded as providing a general reference for vowel production and recognition, because the style-specific need for vocal power and instrumental sound timbre is often superordinated to vowel differentiation. (iv) Roughly spoken, according to Hollien (1972) and his terminology, vocal expressions can be experienced up to f_o = c. 500 Hz for men and c. 800 Hz for women in modal register, and up to f_o = c. 800 Hz for men and even substantially above 1 kHz for women in loft or falsetto register. Thus, voice range profiles of untrained speakers and trained speakers and singers cover often two to three octaves. Further, the intensity range is usually greatest at intermediate f_o levels (Titze, 1992). (v) Noteworthy, expressions with register changes and/or with strong emotional variations, strong vocal efforts and shouting, as well as specific speaking styles (ethnolects, infant directed speech, speech in a large audience, artistic speaking and singing styles etc.) include extensive f_o variation.

Ambiguity of formant patterns and spectral envelopes: If formant patterns and spectral envelopes for sounds with different f_o and/or pitch differ, then what is to be expected are sounds with quasi-identical formant patterns or even quasi-identical spectral envelopes which, however, represent different vowel qualities, the main acoustic difference being their level of f_o and/or their pitch. This kind of ambiguity was already indicated in early studies of vowel synthesis (Potter & Steinberg, 1950, Miller, 1953), as is true (even not discussed explicitly) for the later study of Hirahara and Kato (1992). Recently, Maurer et al. (2017) confirmed the ambiguity in vowel synthesis. Further, and most importantly, the ambiguity was also demonstrated for natural vocalisations, including sounds produced by speakers equal in size and gender or even by single speakers (Maurer & Landis, 2000; Maurer, 2016, pp. 64–65 and187–216).

Noteworthy, Maurer et al. (2017) also demonstrated that open-tube resonance patterns, in their turn, are perceptually not “neutral”, i.e. not exclusively related to the “neutral” Schwa sound, but that they are also ambiguous for vowel recognition if f_o is varied.

f_o versus pitch: Because the two phenomena discussed here can also be observed in cases of a “missing fundamental” (see e.g. Maurer & Suter, 2017b), strictly speaking, we consider the phenomena as related to pitch perception. In most cases, however, both f_o and pitch are concerned.

Non-systematic relation between f_o/pitch and vowel quality-related sound spectrum – the role of the spectral fine structure: As discussed earlier (Maurer & Landis, 1995, 2000, Maurer, 2016, p. 59 and pp. 158–169; see also Traunmüller,1981, 1988, Bladon, 1984, Liénard & Di Benedetto, 1999), the relation between f_o/pitch, spectral peaks and envelope of the sound (if methodologically substantiated), and vowel quality is not systematic. It varies according to the f_o and/or pitch range and the course of the spectral envelope in general, and according to frequencies, levels and harmonic resolution of the spectral peaks in particular, the peaks represented, e.g., in calculated values of formant frequencies, bandwidths and levels. However, roughly speaking, ambiguous spectral peaks and envelopes occur if f_o and/or pitch is varied substantially above c. 200 Hz, and they primarily concern sounds of closed and mid-closed vowel qualities.

References (Appendix)

Barreda, S., & Nearey, T. M. (2012). The direct and indirect roles of fundamental frequency in vowel perception. *The Journal of the Acoustical Society of America*, *131*(1), 466–477.

Bladon, A. (1982). Arguments against formants in the auditory representation of speech. In R. Carlson, & B. Granström (Eds.), *The representation of speech in the peripheral auditory system*, (pp. 95–102). Amsterdam: Elsevier Biomedical Press.

Bladon, R.A.W., Henton, C.G., Pickering, J.B., 1984. Towards an auditory theory of speaker normalization. Lang. Commun. 4, 59–69.

Burris, C., Vorperian, H.K., Fourakis, M., Kent, R.D., Bolt, D.M. (2014). Quantitative and descriptive comparison of four acoustic analsyis systems: vowel measurements. J. Speech. Lang. Hear. Res. 57 (1), 26–45.

Carlson, R., Fant, G., & Granström, B. (1975). Two-formant models, pitch and vowel perception. In G. Fant, & M. A. A. Tathum (Eds.), *Auditory analysis and perception of speech* (pp. 55–82). London: Academic Press.

Carpenter, A., & Morton, J. (1962). The perception of vowel colour in formantless complex sounds. *Language and Speech*, *5*(4), 205–214.

de Cheveigné, A., & Kawahara, H. (1999). Missing-data model of vowel identification. *Journal of the Acoustical Society of America*, *105*(6), 3497–3508.

Ciocca, V., & Whitehill, T. L. (2013). The Acoustic Measurement of Vowels. In M. J. Ball & F. E. Gibbon (Eds.), *Handbook of Vowels and Vowel Disorders* (pp. 113–137). New York, London: Psychology Press, Taylor & Francis Group.

Chistovich, L. A., & Lublinskaya, V. V. (1979). The ‘‘center of gravity’’ effect in vowel spectra and critical distance between the formants: psychoacoustical study of the perception of vowel-like stimuli. *Hearing Research, 1,* 185–195.

Ferreira, A. J. (2007). Static features in real-time recognition of isolated vowels at high pitch. *The Journal of the Acoustical Society of America*, *122*(4), 2389–2404.

Friedrichs, D., Maurer, D., & Dellwo, V. (2015). The phonological function of vowels is maintained at fundamental frequencies up to 880 Hz. *The Journal of the Acoustical Society of America*, *138*(1), EL36–EL42.

Friedrichs, D., Maurer, D., Rosen, S., & Dellwo, V. (2017): Vowel recognition at fundamental frequencies up to 1kHz reveals point vowels as acoustic landmarks. *Journal of the Acoustical Society of America*, *142*(2), 1025–1033.

Fujisaki, H., & Kawashima, T. (1968). The roles of pitch and higher formants in the perception of vowels. *IEEE transactions on audio and electroacoustics*, *16*(1), 73–77.

Gooding, F. (1986). Formantless vowels and theories of vowel perception. *The Journal of the Acoustical Society of America*, *80*(S1), S126–S126.

Harrington, J. (2012). Acoustic phonetics. In W. J. Hardcastle, J. Laver, & F. E.Gibbon (Eds.), *The handbook of phonetic sciences* (pp. 81–129). (2nd ed.). Malden MA: Wiley-Blackwell

Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *The Journal of the Acoustical Society of America*, *108*(6), 3013–3022.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical society of America*, *97*(5), 3099–3111.

Hillenbrand, J. M., Houde, R. A., Nearey, T. M., & Clark, M. J. (1999). Vowel recognition from harmonic spectra. *The Journal of the Acoustical society of America*, 105(2), 1396.

Hillenbrand, J. M., & Houde, R. A. (2003). A narrow band pattern-matching model of vowel perception. *The Journal of the Acoustical Society of America*, *113*(2), 1044–1055.

Hirahara, T., & Kato, H. (1992). The effect of F0 on vowel identification. In Y.Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka, Y. (Eds.), *Speech perception, production and linguistic structure* (pp. 89–112). Tokyo: Ohmsha.

Hollien, H. (1974). On vocal registers. Communication Sciences Laboratory Quarterly Report, 10(1), 1–33.

Ito, M., Tsuchida, J., & Yano, M. (2001). On the effectiveness of whole spectral shape for vowel perception. *The Journal of the Acoustical Society of America*, *110*(2), 1141–1149.

Johnson, K. (2008): Speaker normalization in speech perception. In D. B. Pisoni, R. E. Remez, *The handbook of speech perception* (pp. 363–389). Malden MA, Wiley-Blackwell.

Kasturi, K., Loizou, P. C., Dorman, M., & Spahr, T. (2002). The intelligibility of speech with "holes" in the spectrum. *The Journal of the Acoustical Society of America*, *112*(3), 1102–1111.

Kiefte, M., Enright, T., & Marshall, L. (2010). The role of formant amplitude in the perception of /i/ and /u/. *The Journal of the Acoustical Society of America*, *127*(4), 2611–2621.

Kiefte, M., Nearey, T. M., & Assmann, P. F. (2013). Vowel perception in normal speakers. *Handbook of vowels and vowel disorders* (pp. 160–185). New York NY: Psychology Press.

Ladefoged, P. (2003): Phonetic Data Analysis–An Introduction to Fieldwork and Instrumental Techniques. Malden MA, Wiley-Lackwell.

Ladefoged, P. (1967): Three Areas of Experimental Phonetics. London, Oxford University Press.

Liénard, J. S., & Di Benedetto, M. G. (1999). Effect of vocal effort on spectral properties of vowels. The Journal of the Acoustical Society of America, 106(1), 411–422.

Liu, C., & Eddins, D. A. (2008). Effects of spectral modulation filtering on vowel identification. *The Journal of the Acoustical Society of America*, *124*(3), 1704–1715.

Maurer, D. (2016). *Acoustics of the Vowel –Preliminaries.* Bern: Peter Lang.

Maurer, D., Dellwo, V., Suter, H., & Kathiresan, T. (2017): Formant pattern and spectral shape ambiguity of vowel sounds revisited in synthesis: changing perceptual vowel quality by only changing the fundamental frequency. *Journal of the Acoustical Society of America*, *141*(5), 3469–3470.

Maurer, D., d'Heureuse, C., & Landis, T. (2000). Formant pattern ambiguity of vowel sounds. *International Journal of Neuroscience*, *100*(1–4), 39–76.

Maurer, D., & Landis, T. (1995). FO-dependence, number alteration, and non-systematic behaviour of the formants in German vowels. *International Journal of Neuroscience*, *83*(1–2), 25–44.

Maurer, D., & Landis, T. (1996). Intelligibility and spectral differences in high-pitched vowels. *Folia phoniatrica et logopaedica*, *48*(1), 1–10.

Maurer, D., Mok, P., Friedrichs, D., & Dellwo, V. (2014). Intelligibility of high-pitched vowel sounds in the singing and speaking of a female Cantonese Opera singer. In *Fifteenth Annual Conference of the International Speech Communication Association*, 2132–2133.

Maurer, D., & Suter, H. (2017a, abstract). "Flat" vowel spectra revisited in vowel synthesis. *The Journal of the Acoustical Society of America*, *141*(5), 3469–3469. For details, see http://www.phones-and-phonemes.org/asa/Poster-ASA-2017b-170611.pdf (retrieved February 10th, 2018).

Maurer, D., & Suter, H. (2017b, abstract). Vowel synthesis related to equal-amplitude harmonic series in frequency ranges > 1 kHz combined with single harmonics < 1kHz, and including variation of fundamental frequency. *The Journal of the Acoustical Society of America*, *141*(5), 3469–3469. For details, see

http://www.phones-and-phonemes.org/asa/Poster-ASA-2017c-170611.pdf (retrieved February 10th, 2018).

Maurer, D., Dellwo, V., Suter, H., Kathiresan, T. (2017, abstract): Formant pattern and spectral shape ambiguity of vowel sounds revisited in synthesis: changing perceptual vowel quality by only changing the fundamental frequency. Journal of the Acoustical Society of America, 141(5):3469–3470. For details, see http://www.phones-and-phonemes.org/asa/Poster-ASA-2017a-170611.pdf (retrieved February 10th, 2018).

Maxfield, L., Palaparthi, A., & Titze, I. (2017). New evidence that nonlinear source-filter coupling affects harmonic intensity and Fo stability during instances of harmonics crossing formants. *Journal of Voice*, *31*(2), 149–156.

Ménard, L., Schwartz, J.-L., Boë, L.-J., Kandel, S., Vallée, N. (2002). Auditory normalization of French vowels synthesized by an articulatory model simulating growth from birth to adulthood. *Journal of the Acoustical Society of America 111*(4), 1892–1905.

Miller, R. L. (1953). Auditory tests with synthetic vowels. *The Journal of the Acoustical society of America*, *25*(1), 114–121.

Molis, M. R. (2005). Evaluating models of vowel perception. *The Journal of the Acoustical Society of America*, *111*(2), 2433–2434.

Monsen, R. B., & Engebretson, A. M. (1983). The accuracy of formant frequency measurements: A comparison of spectrographic analysis and linear prediction. *Journal of Speech and Hearing Research*, *26*(3), 89–97.

Morrison, G. S., & Assmann, P. F. (Eds., 2013). *Vowel inherent spectral change*. Berlin: Springer.

Pätzold, M., & Simpson, A. P. (1997). Acoustic analysis of German vowels in the Kiel Corpus of Read Speech. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung Universität Kiel*, *32*, 215–247.

Potter, R. K., & Steinberg, J. C. (1950). Towards the specification of speech. *The Journal of the Acoustical society of America*, *22*(6), 807–820.

Ryalls, J. H., & Lieberman, P. (1982). Fundamental frequency and vowel perception. *The Journal of the Acoustical Society of America*, *72*(5), 1631–1634.

Smith, L. A., & Scott, B. L. (1980). Increasing the intelligibility of sung vowels. *The Journal of the Acoustical Society of America*, *67*(5), 1795–1797.

Sundberg, J. (1987). *The science of the singing voice*. DeKalb IL: Northern Illinois University Press.

Sundberg, J. (2013). Perception of singing. In D. Deutsch (Ed.), *The Psychology of music*. (3rd ed.). London: Academic Press.

Swanepoel, R., Oosthuizen, D. J., & Hanekom, J. J. (2012). The relative importance of spectral cues for vowel recognition in severe noise. *The Journal of the Acoustical Society of America, 132*(4), 2652–2662.

Swerdlin, Y., Smith, J., & Wolfe, J. (2010). The effect of whisper and creak vocal mechanisms on vocal tract resonances. *The Journal of the Acoustical Society of America, 127*(4), 2590–2598.

Titze, I.R. (1992). Acoustic interpretation of the voice range profile (phonetogram). *Journal of Speech and Hearing Research, 35*(1), 21–34,

Titze, I. R., Baken, R. J., Bozeman, K. W., Granqvist, S., Henrich, N., Herbst, C. T.,... & Kreiman, J. (2015). Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization. *The Journal of the Acoustical Society of America*, *137*(5), 3005–3007.

Traunmüller, H. (1981). Perceptual dimension of openness in vowels. *The Journal of the Acoustical Society of America*, *69*(5), 1465–1475.

Traunmüller, H. (1988). Paralinguistic variation and invariance in the characteristic frequencies of vowels. *Phonetica*, *45*(1), 1–29.

Wolfe J. (n.d.). Formant: what *is* a formant? http://newt.phys.unsw.edu.au/jw/formant.htmlAccessed August 28, 2017.

Zahorian, S. A., & Jagharghi, A. J. (1993). Spectral-shape features versus formants as acoustic correlates for vowels. *The Journal of the Acoustical Society of America*, *94*(4), 1966–1982.