

# How listeners recognise vowel sounds under highpass or lowpass filtering of vowel-specific frequency ranges

Dieter Maurer<sup>1</sup>, Thayabaran Kathiresan<sup>2</sup>, Heidy Suter<sup>1</sup>, Volker Dellwo<sup>2</sup>

<sup>1</sup> Institute for the Performing Arts and Film, Zurich University of the Arts, Switzerland, <sup>2</sup> Phonetics Laboratory, University of Zurich, Switzerland  
dieter.maurer@zhdk.ch, thayabaran.kathiresan@uzh.ch, heidy.suter@zhdk.ch, volker.dellwo@uzh.ch www.phones-and-phonemes.org

Session 3pSC  
JASA, 2016, 137, p. 3217

## Background and primary questions

The first two formants in terms of the lower two relative spectral energy maxima are usually assumed as the primary acoustic cue for vowel recognition. Given this general background, the following questions are investigated concerning vowel sounds produced in isolation (V), including a variation in fundamental frequency ( $f_0$ ):

What is the perceptual effect on vowel recognition, if

- (1) the F1 frequency region of a sound is high-pass (HP) filtered?
- (2) the F2 frequency region of a sound is low-pass (LP) filtered?

## Additional question

(3) If, for sounds of single speakers, the F1 region of a low-pitched sound of a given vowel is HP filtered with CF below or equal to  $f_0$  of an (unfiltered) high-pitched sound of that vowel, do the sounds then differ in the perceived vowel quality?

Note: Of specific interest are sounds of single speakers of the vowels /i-y-u/, since their F1 frequency related to low-pitched sounds can by many speakers easily be surpassed producing high-pitched sounds. The same holds true for sounds of the vowels /e-ø-o/ produced by speakers with highly developed vocal abilities.

## Material: Vowel sounds, speakers, $f_0$ , formants

In the present paper, we report the results of an investigation of sounds of the long Standard German vowels /i-y-e-ø-ε-a-o-u/ produced in isolation (V) with a duration of 1.5–2 sec by

- a female speaker at  $f_0 = c.$  220Hz and 660Hz
- a male speaker at  $f_0 = c.$  130Hz and 520Hz

The middle part of the sound of 1 sec (steady state sound nucleus) was extracted for further investigation, including a 0.02 sec fade in/out.

Formant patterns were measured for the low-pitched sounds (see Tables 1 and 4).

**Recordings:** Speakers were asked to produce vowel sounds in isolation on a given  $f_0$  (presented as an electronic piano sound) and sustain the sounds for c. 1.5–2 sec. **Formant measurement** was conducted using PRAAT (Boersma and Weenink, 2016), LPC (robust) algorithm and gender-specific maximum number of formants. The middle 0.3 sec. of a sound of investigation was analysed. Formant tracks were visually crosschecked with sound spectrogram and spectrum. Mismatching values were replaced with estimations based on the spectrogram/spectrum (see the values in the Tables 1 and 4 marked with \*). **HPLP filter** (Hann filter with 6 dB decay per 100 Hz in cutoff frequency) was used for filtering the sounds. CFs correspond to multiples of  $f_0$ . Except are additional CFs which were included for the LP filtering experiment in order to account for F2 of the vowel in question (see Table 4, CF = \*Add.). Boersma, P., and Weenink, D. (2016). Praat: doing phonetics by computer. V6.0.21, retrieved 25 September 2016 from <http://www.praat.org/>.

## Experiment 1: HP filtering the F1 frequency region

- Sounds were HP filtered with stepwise increasing CF as shown in Table 2.
- As a consequence, the F1 frequency region of the sounds was filtered.
- Perceptual vowel quality of all sounds, unfiltered and filtered, was tested in a listening test.

**Listening tests** were performed as blocked speaker tests in the following order: HP filtered, LP filtered and unfiltered sounds. Each single subtest was preceded by a presentation of the test sounds (listening only). Sounds were presented in random order. Concerning the tests, five singers (with professional training) were asked to listen to each sound twice (interval = 1 sec) and to assign one of the nine vowel qualities /i-y-e-ø-ε-a-o-u/ (forced choice). Note that /ɔ/ was added because of the large phonetic distance /a/-/ɔ/ in Standard German vowels.

## Experiment 2: LP filtering the F2 frequency region

- Sounds were LP filtered with stepwise decreasing CF as shown in Table 5.
- As a consequence, the F2 frequency region of the sounds was filtered.
- Perceptual vowel quality of all sounds was tested in a listening test.

## Results 1: HP filtering the F1 frequency region

The results of the HP filtering are given in Tables 3a and 3b:

- Filtered sounds of /i-y-u/ show marked closed–open shifts in the perceived vowel quality for the low and middle CFs.
- Filtered sounds of /e-ø-o/ show a tendency for closed–open shifts in the perceived vowel quality for the middle CFs.
- Filtered Sounds of /ε-a/ show marginal or no shifts.

However, the shifts mentioned vary across speaker,  $f_0$  and CFs, in addition to vowel qualities, and perceptual confusions of vowels also appear.

Vowels	Female speaker			Male speaker		
	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>
i	395	3016	3929	251	2420	3589
y	346	1923	2646	253	1885	2122
u	289	666	3232	289	703	2402
e	429	2633	3120	370	2317	3000*
ø	419	1967	2400*	393	1678	2326
o	415	809	3100*	356	771	2264
ε	849	2655	3300*	674	2191	3308
a	1095	1333	3416	789	1251	2733

Intention	Original sounds: recognition		Filtered sounds: CF (Hz) and recognition					
	$f_c=220\text{Hz}$	$f_c=660\text{Hz}$	$f_0=220\text{Hz}$					
			440	660	880	1100	1320	1320
i	iiii	iiii	o-e-e-e	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε
y	y-y-y-y	y-y-y-y	ø-ø-ø-ø	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε
u	u-u-u-u	u-u-u-u	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o
e	e-e-e-e	e-e-e-e	ø-ø-ø-ø	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε
ø	ø-ø-ø-ø	ø-ø-ø-ø	ø-ø-ø-ø	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε
o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o
ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε
a	a-a-a-a	a-a-a-a	a-a-a-a	a-a-a-a	a-a-a-a	a-a-a-a	a-a-a-a	a-a-a-a

Intention	Original sounds: recognition		Filtered sounds: CF (Hz) and recognition					
	$f_c=220\text{Hz}$	$f_c=660\text{Hz}$	$f_0=130\text{Hz}$					
			390	520	780	910	1040	1040
i	iiii	iiii	ø-ø-ø-ø	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε
y	y-y-y-y	y-y-y-y	ø-ø-ø-ø	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε
u	u-u-u-u	u-u-u-u	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o
e	e-e-e-e	e-e-e-e	ø-ø-ø-ø	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε
ø	ø-ø-ø-ø	ø-ø-ø-ø	ø-ø-ø-ø	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε
o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o
ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε
a	a-a-a-a	a-a-a-a	a-a-a-a	a-a-a-a	a-a-a-a	a-a-a-a	a-a-a-a	a-a-a-a

**Table 1:** Formant frequencies  $F_1$ ,  $F_2$  and  $F_3$ , measured for the low-pitched sounds of the speakers. The order of the vowels is given in closed–open direction related to correspondingly increasing  $F_1$ .

**Table 2:** CF frequencies (Hz) for HP filtering. Grey: CFs clearly above measured  $F_1$  of a vowel.

**Table 3:** Results of the listening test for unfiltered and HP filtered sounds (3a = sounds of the female speaker; 3b = sounds of the male speaker). Blue: Differences between intended and recognised vowel quality of the unfiltered sound. Red: Indications of closed–open shifts in the perceived vowel quality (light red = 3/5 listeners, dark red = 4/5 or 5/5 listeners).

## Additional results

Addressing the additional question above, the results of the listening tests indicate:

- Sounds of /i-y-u/ show marked perceptual differences between filtered low- and unfiltered high-pitched sounds.
- Sounds of /e-ø-o/ show a tendency for a corresponding perceptual difference.
- Sounds of /ε-a/ show no corresponding perceptual effect.

## Results 2: LP filtering the F2 frequency region

The results of the LP filtering (see Tables 6a and 6b) show a general tendency of unrounded–rounded and front–back shifts in the perceived vowel quality. Again, the shifts vary across speaker,  $f_0$  and CFs, in addition to vowel qualities, and perceptual confusions of vowels also appear.

Vowels	Female speaker			Male speaker		
	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>
i	395	3016	3929	251	2420	3589
y	346	1923	2646	253	1885	2122
u	289	666	3232	289	703	2402
e	429	2633	3120	370	2317	3000*
ø	419	1967	2400*	393	1678	2326
o	415	809	3100*	356	771	2264
ε	849	2655	3300*	674	2191	3308
a	1095	1333	3416	789	1251	2733

Intention	Original sounds: recognition		Filtered sounds: CF (Hz) and recognition								
	$f_c=220\text{Hz}$	$f_c=660\text{Hz}$	$f_0=220\text{Hz}$								
			3300	2640	1980	1320	660	3300	2640	1980	1320
i	iiii	iiii	ø-ø-ø-ø	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε
e	e-e-e-e	e-e-e-e	ø-ø-ø-ø	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε
ø	ø-ø-ø-ø	ø-ø-ø-ø	ø-ø-ø-ø	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε
o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o
ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε
y	y-y-y-y	y-y-y-y	y-y-y-y	y-y-y-y	y-y-y-y	y-y-y-y	y-y-y-y	y-y-y-y	y-y-y-y	y-y-y-y	y-y-y-y
ø	ø-ø-ø-ø	ø-ø-ø-ø	ø-ø-ø-ø	ø-ø-ø-ø	ø-ø-ø-ø	ø-ø-ø-ø	ø-ø-ø-ø	ø-ø-ø-ø	ø-ø-ø-ø	ø-ø-ø-ø	ø-ø-ø-ø
a	a-a-a-a	a-a-a-a	a-a-a-a	a-a-a-a	a-a-a-a	a-a-a-a	a-a-a-a	a-a-a-a	a-a-a-a	a-a-a-a	a-a-a-a
o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o
u	u-u-u-u	u-u-u-u	u-u-u-u	u-u-u-u	u-u-u-u	u-u-u-u	u-u-u-u	u-u-u-u	u-u-u-u	u-u-u-u	u-u-u-u

Intention	Original sounds: recognition		Filtered sounds: CF (Hz) and recognition								
	$f_c=220\text{Hz}$	$f_c=660\text{Hz}$	$f_0=130\text{Hz}$								
			2600	2080	1560	1040	660	2600	2080	1560	1040
i	iiii	iiii	ø-ø-ø-ø	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε
e	e-e-e-e	e-e-e-e	ø-ø-ø-ø	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε
ø	ø-ø-ø-ø	ø-ø-ø-ø	ø-ø-ø-ø	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε
o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o
ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε	ε-ε-ε-ε
y	y-y-y-y	y-y-y-y	y-y-y-y	y-y-y-y	y-y-y-y	y-y-y-y	y-y-y-y	y-y-y-y	y-y-y-y	y-y-y-y	y-y-y-y
ø	ø-ø-ø-ø	ø-ø-ø-ø	ø-ø-ø-ø	ø-ø-ø-ø	ø-ø-ø-ø	ø-ø-ø-ø	ø-ø-ø-ø	ø-ø-ø-ø	ø-ø-ø-ø	ø-ø-ø-ø	ø-ø-ø-ø
a	a-a-a-a	a-a-a-a	a-a-a-a	a-a-a-a	a-a-a-a	a-a-a-a	a-a-a-a	a-a-a-a	a-a-a-a	a-a-a-a	a-a-a-a
o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o	o-o-o-o
u	u-u-u-u	u-u-u-u	u-u-u-u	u-u-u-u	u-u-u-u	u-u-u-u	u-u-u-u	u-u-u-u	u-u-u-u	u-u-u-u	u-u-u-u

**Table 4:** The same formant frequencies  $F_1$ ,  $F_2$  and  $F_3$  as shown in Table 1, but presented in the order unrounded–rounded and front–back, related to correspondingly decreasing  $F_2$ .

**Table 5:** CF frequencies (Hz) for LP filtering. Grey: CFs clearly below measured  $F_2$  of a vowel.

**Table 6:** Results of the listening test for unfiltered and LP filtered sounds (6a = sounds of the female speaker; 6b = sounds of the male speaker). Blue: Differences between intended and recognised vowel quality of the unfiltered sound. Red: Indications of unrounded–rounded and front–back shifts in the perceived vowel quality (light red = 3/5 listeners, dark red = 4/5 or 5/5 listeners).

## Discussion

Although the present investigation relates to a highly limited number of sounds, speakers, fundamental frequencies and listeners, we consider three indications to be of primary importance for the acoustics of vowels:

- The perceptual role of the F1 frequency region for vowel recognition is neither uniform across vowel qualities nor independent of  $f_0$ .
- Sounds with filtered F1 or F2 frequency region are very often recognised as a specific vowel (equal or different from the intended vowel). Thus, F1-F2 is not a prerequisite of vowel recognition.
- The present perceptual findings confirm an earlier claim that the vowel sound has to be considered as a phenomenon of a spectral foreground–background relation (Maurer, 2016, pp. 80–83).