Zürcher Hochschule der Künste
Zurich University of the Arts

Universität Zürich[UZH]

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

5th Joint Meeting of the Acoustical Society of America and Acoustical Society of Japan
Honolulu, Hawaii, 2016

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

Dieter Maurer[1], Thayabaran Kathiresan[2], Heidy Suter[1], Volker Dellwo[2]
[1] Institute for the Performing Arts and Film, Zurich University of the Arts, Switzerland, [2] 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

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_o):

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_o 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_o, formants

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

- **a female speaker at f_o = c. 220Hz and 660Hz**
- **a male speaker at f_o = 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 *fo* (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 *.)
HP/LP 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_o. 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-ɛ-a-ɔ-o-u/ (forced choice). Note that /ɔ/ was added because of the large phonetic distance /o/–/a/ 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 /ɛ-a/ show marginal or no shifts.**

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

Table 1: Formant frequencies (in Hz) relevant for HP filtering

Vowels		Female speaker			Male speaker		
		F_1	F_2	F_3	F_1	F_2	F_3
closed to open	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
	ɛ	**849**	2655	3300*	**674**	2191	3308
	a	**1095**	1333	3416	**789**	1251	2733

Table 2: CF (in Hz) for HP filtering

Female speaker					Male speaker					Vowels	
f_o=220Hz				f_o=220/660Hz	f_o=130Hz				f_o=130/520Hz		
440	660	880	1100	1320	390	520	780	910	1040		
x	x	x	x	x	x	x	x	x	x	i	closed to open
x	x	x	x	x	x	x	x	x	x	y	
x	x	x	x	x	x	x	x	x	x	u	
x	x	x	x	x	x	x	x	x	x	e	
x	x	x	x	x	x	x	x	x	x	ø	
x	x	x	x	x	x	x	x	x	x	o	
x	x	x	x	x	x	x	x	x	x	ɛ	
x	x	x	x	x	x	x	x	x	x	a	

Table 3a: HP filter experiment, results of the listening test for the sounds of the female speaker

Intention		Original sounds: recognition		Filtered sounds: CF (Hz) and recognition					
		f_o=220Hz	f_o=660Hz	f_o=220Hz					f_o=660Hz
				440	660	880	1100	1320	1320
closed to open	i	i-i-i-i-i	i-i-i-i-i	o-e-e-e-e	ɛ-ɛ-e-e-i	ɛ-ɛ-ɛ-ɛ-i	ɛ-ɛ-i-i-i	e-e-e-i-i	ɛ-e-e-i-i
	y	y-y-y-y-y	y-y-y-y-y	ø-ø-ø-ø-ø	ɛ-ɛ-ø-ø-ø	ɛ-ɛ-ɛ-ø-ø	ɛ-ø-ø-y-y	a-ø-y-y-y	ɛ-ɛ-y-y-y
	u	u-u-u-u-u	u-u-u-u-u	o-o-o-o-o	o-ɔ-ɔ-ɔ-ɔ	o-ɔ-ɔ-ɔ-a	ɔ-a-a-a-a	u-a-ɛ-ɛ-y	a-ɛ-ɛ-ɛ-ɛ
	e	e-e-e-e-e	e-e-e-e-e	e-e-e-e-e	ɛ-ɛ-ɛ-ɛ-ɛ	ɛ-ɛ-ɛ-ɛ-e	ɛ-ɛ-ɛ-e-e	ɛ-e-e-i-i	ɛ-ɛ-ɛ-y-y
	ø	ø-ø-ø-ø-ø	ø-ø-ø-ø-ø	ø-ø-ø-ø-ø	ɛ-ɛ-ø-ø-ø	ɛ-ɛ-ɛ-ø-ø	ɛ-ɛ-ɛ-ø-y	a-ɛ-y-y-y	a-ɛ-y-y-y
	o	o-o-o-o-o	u-u-u-u-o	o-o-o-o-o	ɔ-ɔ-ɔ-ɔ-ɔ	ɔ-ɔ-ɔ-ɔ-a	a-a-a-a-ɛ	a-ø-ø-ø-y	a-a-a-a-ɛ
	ɛ	ɛ-ɛ-ɛ-ɛ-ɛ	ɛ-ɛ-e-e-e	a-ɛ-ɛ-ɛ-ɛ	a-ɛ-ɛ-ɛ-ɛ	ɛ-ɛ-ɛ-ɛ-ɛ	ɛ-ɛ-ɛ-ɛ-e	ɛ-ɛ-e-e-e	a-ɛ-ɛ-ɛ-y
	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

Table 3b: HP filter experiment, results of the listening test for the sounds of the male speaker

Intention		Original sounds: recognition		Filtered sounds: CF (Hz) and recognition					
		f_o=130Hz	f_o=520Hz	f_o= 130Hz					f_o=520Hz
				390	520	780	910	1040	1040
closed to open	i	i-i-i-i-i	i-i-i-i-i	e-e-e-e-e	ɛ-ɛ-e-e-i	e-i-i-i-i	e-i-i-i-i	i-i-i-i-i	ɛ-e-e-e-i
	y	y-y-y-y-y	y-y-y-y-y	ø-ø-ø-ø-ø	ø-ø-ø-y-y	ø-y-y-y-y	y-y-y-y-y	y-y-y-y-y	ɛ-y-y-y-y
	u	u-u-u-u-u	u-u-u-u-u	o-o-o-o-o	ɔ-ɔ-ɔ-ɔ-ɔ	o-ɔ-ɔ-a-ø	o-o-a-ɛ-ɛ	o-a-ɛ-y-i	u-o-a-a-ɛ
	e	e-e-e-e-e	e-e-e-e-e	ø-e-e-e-e	ɛ-ɛ-ɛ-e-e	ɛ-ɛ-e-y-y	ɛ-ɛ-y-i-i	ɛ-e-e-y-y	ɛ-ɛ-ɛ-e-e
	ø	ø-ø-ø-ø-ø	ø-ø-ø-ø-ø	ø-ø-ø-ø-ø	ø-ø-ø-ø-ø	ɛ-ɛ-ø-ø-ø	a-ɛ-ø-ø-ø	a-ø-ø-ø-y	a-a-a-a-ø
	o	o-o-o-o-o	o-o-ɔ-ɔ-ɔ	o-o-o-o-o	o-ɔ-ɔ-ɔ-ɔ	o-ɔ-ɔ-ɔ-ɔ	o-ɔ-ɔ-a-a	o-a-a-a-a	o-ɔ-a-a-ø
	ɛ	ɛ-ɛ-ɛ-ɛ-ɛ	ɛ-ɛ-ɛ-ɛ-ɛ	a-ɛ-ɛ-ɛ-ø	ɛ-ɛ-ɛ-ɛ-ɛ	ɛ-ɛ-ɛ-ɛ-ɛ	ɛ-ɛ-ɛ-ɛ-e	ɛ-ɛ-e-e-y	a-ɛ-ɛ-ɛ-e
	a	a-a-a-a-a	ɔ-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	ɔ-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 /ɛ-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_o and CFs, in addition to vowel qualities, and perceptual confusions of vowels also appear.

Table 4: Formant frequencies (in Hz) relevant for LP filtering

Vowels		Female speaker			Male speaker		
		F_1	F_2	F_3	F_1	F_2	F_3
unrounded to rounded / front to back	i	395	**3016**	3929	251	**2420**	3589
	e	429	**2615**	3132	368	**2316**	3000*
	ɛ	849	**2655**	3300*	674	**2189**	2900*
	y	346	**1923**	2646	253	**1885**	2122
	ø	419	**1966**	2400*	393	**1678**	2326
	a	1095	**1333**	3416	787	**1250**	2756
	o	416	**810**	3100*	359	**769**	2264
	u	289	**666**	3232	289	**703**	2402

Table 5: CF (in Hz) for LP filtering

Female speaker						Male speaker						Vowels	
All sounds					Add.	All sounds					Add.		
3300	2640	1980	1320	660		2600	2080	1560	1040	520			
x	x	x	x			x	x	x	x			unrounded to rounded / front to back	i
x	x	x	x			x	x	x	x				e
x	x	x	x			x	x	x	x				ɛ
		x	x				x	x	x				y
		x	x				x	x	x				ø
		x	x	x	1100			x	x	x	780		a
			x	x					x	x			o
			x	x	440				x	x	440		u

Table 6a: LP filter experiment, results of the listening test for the sounds of the female speaker

Intention		Original sounds: recognition		Filtered sounds: CF (Hz) and recognition							
		f_o=220Hz	f_o=660Hz	f_o=220Hz				f_o=660Hz			
				3300	2640	1980	1320	3300	2640	1980	1320
unrounded to rounded / front to back	i	i-i-i-i-i	i-i-i-i-i	i-i-i-i-i	e-y-y-y-y	y-y-y-y-y	u-u-u-u-u	e-e-i-i-i	y-y-y-y-y	u-u-u-u-y	u-u-u-u-i
	e	e-e-e-e-e	e-e-e-e-e	e-e-e-e-e	e-e-e-e-e	o-ø-ø-ø-ø	o-o-o-o-ø	e-e-e-e-i	e-e-e-e-y	ø-ø-ø-ø-y	u-u-o-o-o
	ɛ	ɛ-ɛ-ɛ-ɛ-ɛ	ɛ-ɛ-e-e-e	ɛ-ɛ-ɛ-ɛ-ɛ	a-ɛ-ɛ-ɛ-ɛ	a-ɛ-ɛ-ɛ-ɛ	ɔ-ɔ-a-a-a	ɛ-e-e-e-e	ɛ-e-e-e-e	ø-ø-ø-ø-ø	u-o-o-ɔ-a
	y	y-y-y-y-y	y-y-y-y-y	–	–	y-y-y-y-y	u-u-u-u-u	–	–	y-y-y-y-y	u-u-u-o-y
	ø	ø-ø-ø-ø-ø	ø-ø-ø-ø-ø	–	–	ø-ø-ø-ø-ø	o-o-o-o-ø	–	–	ø-ø-ø-ø-ø	u-u-o-o-o
				1980	1320	660	Add.	1980	1320	660	
	a	a-a-a-a-a	a-a-a-a-a	a-a-a-a-a	a-a-a-a-a	u-u-o-o-o	a-a-a-a-a	a-a-a-a-a	a-a-a-a-a	u-u-u-o-y	–
	o	o-o-o-o-o	u-u-u-u-o	–	o-o-o-o-o	o-o-o-o-o	–	–	u-u-u-o-y	u-u-u-o-ø	–
	u	u-u-u-u-u	u-u-u-u-u	–	u-u-u-u-o	u-u-u-u-y	u-u-u-u-u	–	u-u-u-u-y	u-u-u-u-y	–

Table 6b: LP filter experiment, results of the listening test for the sounds of the male speaker

Intention		Original sounds: recognition		Filtered sounds: CF (Hz) and recognition							
		f_o=130Hz	f_o=520Hz	f_o=130Hz				f_o=520Hz			
				2600	2080	1560	1040	2600	2080	1560	1040
unrounded to rounded / front to back	i	i-i-i-i-i	i-i-i-i-i	e-i-i-i-i	u-y-y-y-y	u-u-u-y-y	u-u-u-u-u	e-y-y-y-y	u-ø-ø-y-y	u-o-y-y-y	u-u-u-u-u
	e	e-e-e-e-e	e-e-e-e-e	e-e-e-e-e	ø-ø-e-e-e	o-o-ø-ø-ø	o-o-o-o-ø	e-e-e-e-y	ø-ø-ø-ø-ø	u-u-u-o-o	u-u-u-u-o
	ɛ	ɛ-ɛ-ɛ-ɛ-ɛ	ɛ-ɛ-ɛ-ɛ-ɛ	ɛ-ɛ-ɛ-ɛ-ɛ	ɛ-ɛ-ɛ-ɛ-ɛ	a-ɛ-ɛ-ɛ-ɛ	a-a-ɛ-ɛ-ɛ	ɛ-ɛ-ɛ-ɛ-ɛ	a-a-a-a-a	ɔ-a-a-a-a	u-o-o-o-ɔ
	y	y-y-y-y-y	y-y-y-y-y	–	y-y-y-y-y	u-u-y-y-y	u-u-u-u-u	–	ø-y-y-y-y	u-u-u-o-ø	u-u-u-u-u
	ø	ø-ø-ø-ø-ø	ø-ø-ø-ø-ø	–	ø-ø-ø-ø-ø	ø-ø-ø-ø-ø	o-o-o-o-ø	–	ø-ø-ø-ø-ø	ø-ø-ø-ø-y	u-u-u-o-o
				1560	1040	520	Add.	1560	1040	520	Add.
	a	a-a-a-a-a	ɔ-a-a-a-a	a-a-a-a-a	a-a-a-a-ɛ	o-o-o-ɔ-ɔ	ɔ-a-a-ɛ-ɛ	a-a-a-a-a	u-u-u-o-o	u-u-u-y-y	–
	o	o-o-o-o-o	o-o-ɔ-ɔ-ɔ	–	o-o-o-o-o	o-o-o-o-ø	–	–	u-u-o-o-ɔ	u-u-u-u-y	–
	u	u-u-u-u-u	u-u-u-u-u	–	u-u-u-u-u	u-u-u-u-i	u-u-u-u-u	–	u-u-u-y-y	u-u-u-u-i	–

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_o.**
- **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).**

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