hdk
Zürcher Hochschule der Künste
Zurich University of the Arts

Universität Zürich[UZH]

Formant pattern and spectral shape ambiguity of vowel sounds revisited in synthesis: changing perceptual vowel quality by only changing the fundamental frequency

Acoustics '17, Boston
173rd Meeting of the Acoustical Society of America and the 8th Forum Acusticum

Dieter Maurer[1], Volker Dellwo[2], Heidy Suter[1], Thayabaran Kathiresan[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, volker.dellwo@uzh.ch, heidy.suter@zhdk.ch, thayabaran.kathiresan@uzh.ch **www.phones-and-phonemes.org/asa**

Journal of the Acoustical Society of America
141(5): 3469–3470

Background

Relationship between vowel-specific lower formants and fundamental frequency: Vowel-specific lower formants (if measurable) relate to fundamental frequency (f_o), above all for f_o > 200Hz [1–6, see also 7, pp. 59–63 and pp. 158–186]. (Note that this relation is indicated to be unsystematic concerning both the frequency ranges of f_o variation and the perceived vowel qualities.)

Formant pattern and spectral envelope ambiguity in natural vowel sounds: Because of the above mentioned relation, for natural vowel sounds produced at very different levels of f_o, formant patterns often prove to be ambiguous: sounds manifesting a quasi-identical formant pattern – and in some cases even the entire spectral envelope – can represent different vowel qualities, the main difference being their level of f_o [8; see also 7, pp. 64–65 and pp. 187–216].

Remaining problem: Since formant pattern as well as spectral envelope estimation for natural vowel sounds at very different f_o are subject to methodological criticism, the ambiguity needs to be confirmed in vowel synthesis experiments, in which formant pattern and spectral shape determination can be fully controlled.

Question

Is it possible to synthesise vowel sounds of varying perceived vowel quality by varying f_o only (i.e. use identical source signal quality and identical filter transfer function)?

Experiment 1

Synthesis: Based on investigations of natural Standard German vowel sounds, various model formant patterns *F*1'–*F*2'–*F*3' were created and, for each single pattern, sounds were synthesised on two or three fundamental frequencies (200–400Hz, and 200–300–600Hz, respectively). Thereby, the frequencies of *F*1'–*F*2'–*F*3' were set to always coincide with a harmonic frequency of the sound spectrum. The levels and bandwidths of the formants were set to create filter curves matching the spectral envelopes of the natural vowels imitated in synthesis. – *F*4' and *F*5' with 200Hz bandwidths and low levels were added to smoothen the higher frequencies > 3.5kHz. – Monotonous sounds of 1 sec. were synthesised using the Klatt synthesiser in PRAAT ([9], cascade mode, sampling frequency SF = 44.1kHz; synthesis parameters see Tables 1).

Experiment 2

Synthesis: Corresponding to open-tube resonance characteristics for men, women and children, respectively, sounds were synthesised with formant patterns *F*1' to *F*5', formant frequencies being odd multiples of 500, 600, or 700Hz, and fundamental frequencies being 1/3, 1/2 or 1/1 of the first formant frequency. In consequence, formant frequencies always match with frequencies of harmonics in the sound spectrum. – All formant bandwidths were set = 100Hz. – Monotonous sounds of 1 sec. were synthesised using the Klatt synthesiser in PRAAT ([9], parallel mode, SF = 44.1kHz; synthesis parameters see Tables 2).

Experiments 1 and 2 – illustration

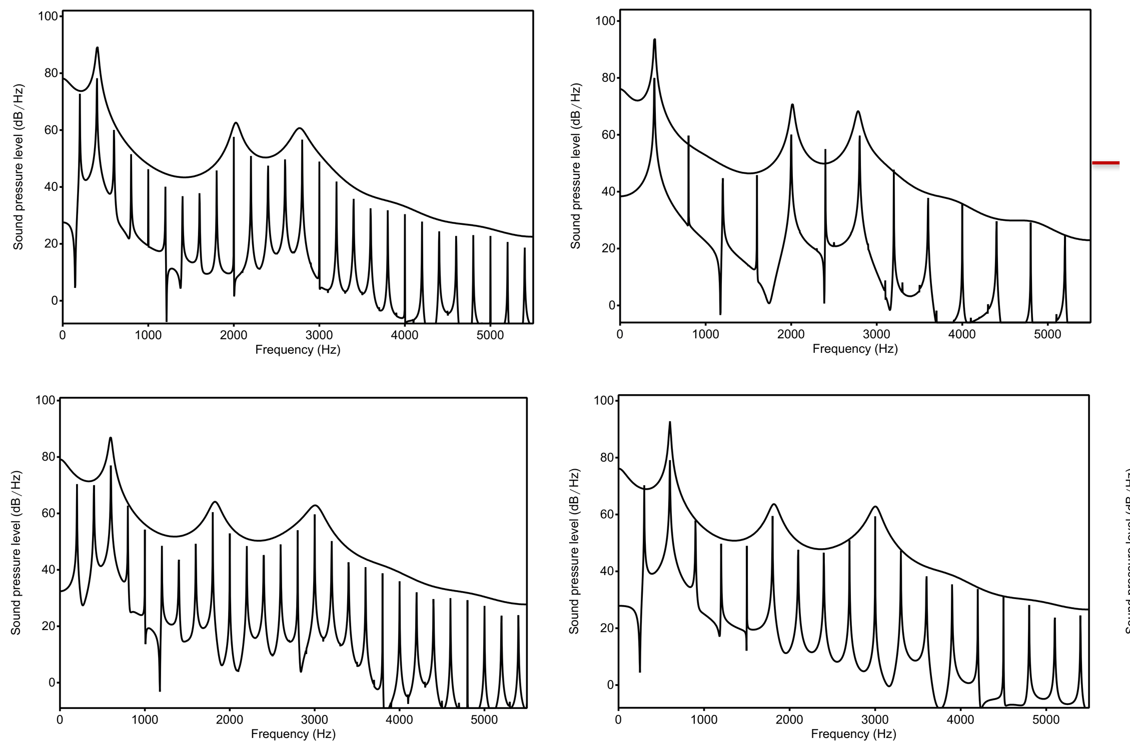
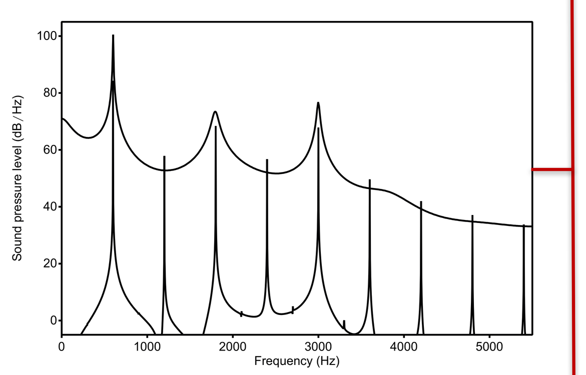
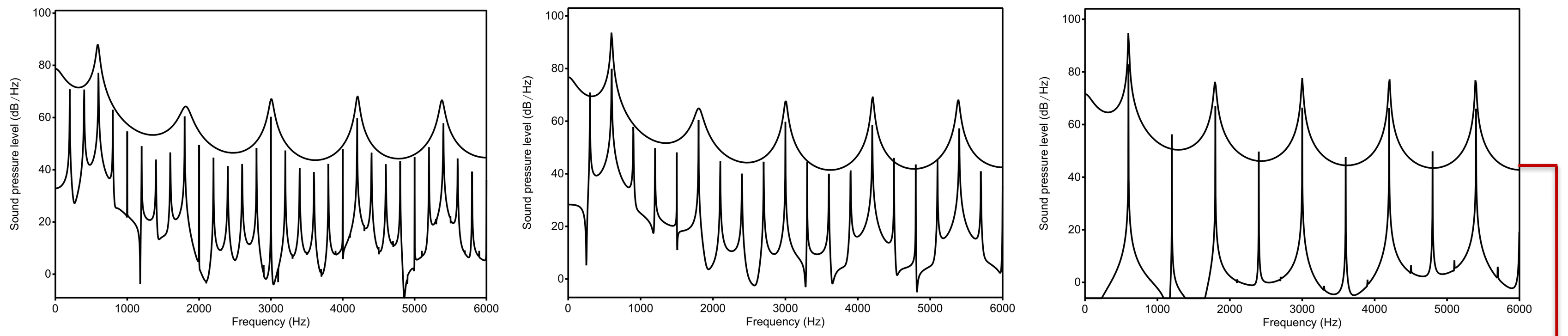



Figure 1: Illustration of the first experiment. **Left** = equally filtered sound pair with two different f_o of 200 and 400Hz (see Table 1, Series E). **Below** = three equally filtered sounds with different f_o of 200–300–600Hz (see Table 1, Series F).

Figure 2 (below): Illustration of the second experiment. Three sounds synthesised on the basis of a single open-tube resonance pattern = 600–1800–3000–4200–5400Hz but different f_o = 200–300–600Hz (see Table 2, Series B). – **Note that all illustrations show spectra and LPC curves measured with PRAAT [9] for the sounds after synthesis.**



Experiment 1 and 2 – listening tests

For each experiment, 5 phonetic expert listeners identified the synthesised sounds in a multiple-choice identification task according to Standard German vowel qualities and Schwa (/ə/). Each sound was presented twice. All sounds were presented in random order.

Results

The vowel recognition results for both experiments show (see Tables 1 and 2):

- **Consistent perceptual open-closed shifts in vowel quality for all sound pairs and sound triples tested**
- **Perceptual shifts to an adjacent vowel quality for f_o variations of 1 octave**
- **Perceptual shifts to a non-adjacent vowel quality for f_o variations ≥ 1.5 octaves**

References:
[1] Traunmüller, H. (1981): Perceptual dimensions of openness in vowels. Journal of the Acoustical Society of America, 69, 1465–1475.
[2] Traunmüller, H. (1985): The role of the fundamental and the higher formants in the perception of speaker size, vocal effort, and vowel openness. PERILUS (Institute of Linguistics, University of Stockholm), IV, pp.92–102.
[3] Maurer, D., Landis, T. (1995): F0-dependence, number alteration, and non-systematic behaviour of the formants in German vowels. International Journal of Neuroscience, 83, 25–44.
[4] Maurer, D., Landis, T. (1996): Intelligibility and spectral differences in high pitched vowels. Folia phoniatrica et logopaedica, 48, 1–10.
[5] 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. Proceedings of the 15th Conference of the International Speech Communication Association, Interspeech 2014, 2132–2133.
[6] Friedrichs, D., Maurer, D., Dellwo, V. (2015): The phonological function of vowels is maintained at fundamental frequencies up to 880 Hz. Journal of the Acoustical Society of America, 138, EL36.
[7] Maurer, D. (2016): Acoustics of the Vowel – Preliminaries. Bern/Frankfurt: Peter Lang.
[8] Maurer, D., d'Heureuse, C., Landis, T. (2000): Formant pattern ambiguity of vowel sounds. International Journal of Neuroscience, vol. 100, pp. 39–76.
[9] Boersma, P., & Weenink, D. (2015). Praat: doing phonetics by computer. [Computer program]. Version 6.0.28. Retrieved March 30, 2017, from http://www.praat.org.
[10] Maurer, D., Suter, H., Friedrichs, D., Dellwo, V. (2015): Acoustic Characteristics of Voice in Music and Straight Theatre: Topics, Conceptions, Questions. In Leemann, A., M.-J. Kolly, M.-J., Schmid, S., Dellwo, V. (Eds.), Trends in Phonetics in German-speaking Europe. Bern/Frankfurt: Peter Lang, pp. 393–406.

Results (details)

Table 1: Results of the first experiment. Sound pairs and triples (Series A to G) synthesised on the basis of equal formant patterns/filter curves (Columns 4–12, formant frequencies, levels and bandwidths) but different f_o (Column 3), set according to observations on natural Standard German vowels (Column 2). The confusion matrix shows the results of the listening test (5 listeners, each sound presented twice, 10 identifications per sound; B = back vowel).

Model		Klatt synthesis parameters (cascade mode)										Vowel recognition (listening test)							
Sound series	Vowel model	f_o Hz	$F_{1'}$ Hz	$L_{1'}$ dB	$B_{1'}$ Hz	$F_{2'}$ Hz	$L_{2'}$ dB	$B_{1'}$ Hz	$F_{2'}$ Hz	$L_{3'}$ dB	$B_{3'}$ Hz	Confusion matrix							Majority
Investigation of back vowels												ə	a	ɔ	o	u			
A	o	200	400	100	100	800	105	100	2800	90	200	1			8	1			o
	u	400														10			u
B	ɔ	200	600	100	100	1200	95	100	3000	85	200		3	7					ɔ
	o	300													10				o
	u	600														10			u
Investigation of front vowels												ə	ɛ	ø	e	y	i	B	
C	e	200	400	100	100	2400	100	200	2800	100	200			1	9				e
	i	400													1	3	6		i
D	e	200	400	100	100	2800	100	200	3200	100	200				9			1	e
	i	400														2	8		i
E	ø	200	400	100	100	2000	100	150	2800	100	200			8	2				ø
	y	400														10			y
F	ɛ	200	600	100	100	2400	100	200	3000	100	200		8					2	ɛ
	e	300											4		6				e
	i	600													4	1	5		i
G	ɛ	200	600	100	100	1800	100	150	3000	100	200	5	5						ə-ɛ
	ø	300											2	8					ø
	y	600												2		8			y

Table 2: Results of the second experiment. Sound triples (Series A to C), each triple synthesised on the basis of a single open-tube resonance pattern usually attributed to either men, women, or children Column 4–8 and 10), but different f_o (Column 3). The confusion matrix shows the results of the listening test (5 listeners, each sound presented twice, 10 identifications per sound; B = back vowel).

Model		Klatt synthesis parameters (parallel mode)										Vowel recognition (listening test)							
Sound series	Speaker group	f_o Hz	$F_{1'}$ Hz	$F_{2'}$ Hz	$F_{3'}$ Hz	$F_{4'}$ Hz	$F_{5'}$ Hz		all *B*' Hz			Confusion matrix							Majority
												ə	ɛ	ø	e	y	i		
A	Men	125	500	1500	2500	3500	4500		100			10							ə
		250												10					ø
		500												1	1	8			y
B	Women	200	600	1800	3000	4200	5400		100			10							ə
		300												10					ø
		600										1				9			y
C	Children	233	700	2100	3500	4900	6300		100			4	6						ɛ
		350										1		8	1				ø
		700												2	1	7			y

Discussion

Vowel synthesis **confirms the ambiguity of formant patterns and spectral envelopes** as observed for natural vowel sounds. Further, vowel synthesis **provides evidence that open-tube resonance patterns are perceptually not "neutral", i.e. not exclusively related to the "neutral" Schwa sound, but that they are also ambiguous for vowel recognition.** – These findings cannot be regarded as solely an aspect of vowel perception: If filter curves as such are not neutral for vowel recognition, vowel production too cannot be described by simply defining a vowel-specific resonance pattern, and the same holds true for the acoustics of vowels. – There is an extensive and controversial debate in the literature about the relation between f_o and formant patterns. However, as to our knowledge, the findings reported here exceed existing assumptions on this relation, and the ambiguity shown, which in some cases involves three vowel qualities, challenges our understanding of vowel sounds in general. – We understand the arising question as indicating the need for further research on vowel sounds produced under extensive variation of production parameters [7, 10].