Zürcher Hochschule der Künste
Zurich University of the Arts

Universität Zürich UZH

Vowel synthesis related to equal-amplitude harmonic series in frequency ranges > 1 kHz combined with single harmonics < 1 kHz, and including variation of fundamental frequency

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

Dieter Maurer, Heidy Suter
Institute for the Performing Arts and Film, Zurich University of the Arts, Switzerland
dieter.maurer@zhdk.ch, heidy.suter@zhdk.ch **www.phones-and-phonemes.org/asa**

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

Background

"Flat" vowel spectra: Front vowels can be synthesised on the basis of series of harmonics equal in amplitude, with frequencies only above 1 kHz ("flat" higher spectral bands). In these cases, spectral energy usually attributed to the frequency region of the first formant is lacking [1].

Perceptual ambiguity of formant patterns and spectral envelopes: Front (and also back) vowels can manifest ambiguous formant patterns as well as spectral envelopes, i.e., quasi-identical patterns/envelopes of sounds representing different perceived vowel qualities, the main difference being their level of f_o. (For first indications, see [2, 3]; for extensive demonstration, see [4–6].)

Question of investigation

The present investigation combines both aspects in a vowel synthesis experiment, with the following question: **If two synthesised sounds relate to harmonic spectra equal in both the frequency of the lower harmonic and the frequency ranges of the higher harmonics, all harmonics also equal in amplitude, and only the frequency spacing of the higher harmonics differs – is the perceived vowel quality maintained or does it alter?**

Note: Key aspects in this poster are highlighted in red.

Vowel synthesis, f_o variation

Synthesis: Series of monotonous sounds (1.2 sec., including 0.1 sec. fade in/out) were produced using a harmonic synthesiser, the harmonic spectra consisting of:

- Series of higher harmonics covering equal higher frequency ranges > 1kHz
- Combined with a single lower harmonic < 1 kHz
- All harmonics multiples of a common reference frequency and equal in amplitude
- Phases = 0

Sound series were created including **two types of sound pairs/comparisons:**

- **Comparison type T1:** Sounds with equal frequency ranges and equal frequency spacing of higher harmonics **but different single lower harmonic frequencies**
- **Comparison type T2: Sounds with equal frequency ranges of higher harmonics and equal frequencies of the single lower harmonic but different frequency spacing between the higher harmonics**

Note: Sound pairs of T1 manifest equal periodicity and are perceived on the same pitch, sound pairs of T2 manifest different periodicity and are perceived on different pitches.

On the basis of extensive acoustic analysis of natural front vowel sounds [5], seven series of three and one series of six sounds were created (see Table 1) with pitch variations of 220–440Hz (Series A to C), 150–450Hz (Series D to F), 150–300Hz (series G) and 150–300–600Hz (Series H). Each single series consists of both sound comparison types T1 and T2.

References
[1] Maurer, D, Suter, H. (2017): "Flat" vowel spectra revisited in vowel synthesis. Parallel Poster, Acoustics '17, Boston.
[2] Potter, R.K., Steinberg, J.C. (1950): Toward the Specification of Speech. Journal of the Acoustical Society of America, 22, 807–820.
[3] Miller, R.L., 1953. Auditory Tests with Synthetic Vowels. Journal of the Acoustical Society of America, 25, 114–121.
[4] Maurer, D., and Landis, T. (2000): Formant pattern ambiguity of vowel sounds International Journal of Neuroscience, 100, 39-76.

Illustration of sound synthesis

Figure 1: Comparison of sound triples of front vowels (pitch variation of 1 or 1.5 octaves; example see Table 1, Series A).
Comparison type T1: difference in the harmonic spectra = frequency of the lower harmonic < 1kHz.
Comparison type T2: difference in the harmonic spectra = frequency spacing of the higher harmonics > 1kHz, and perceived pitch.

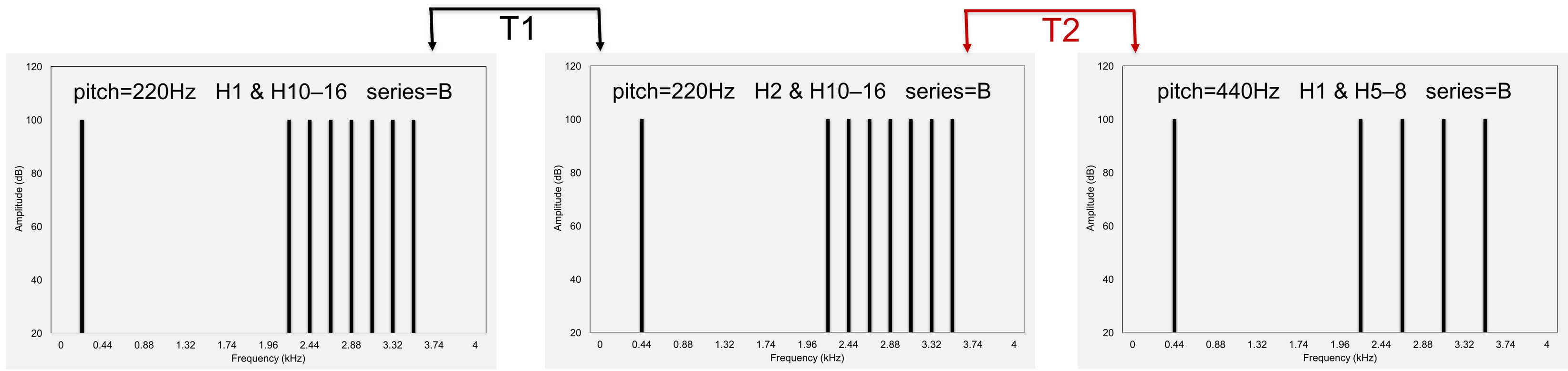
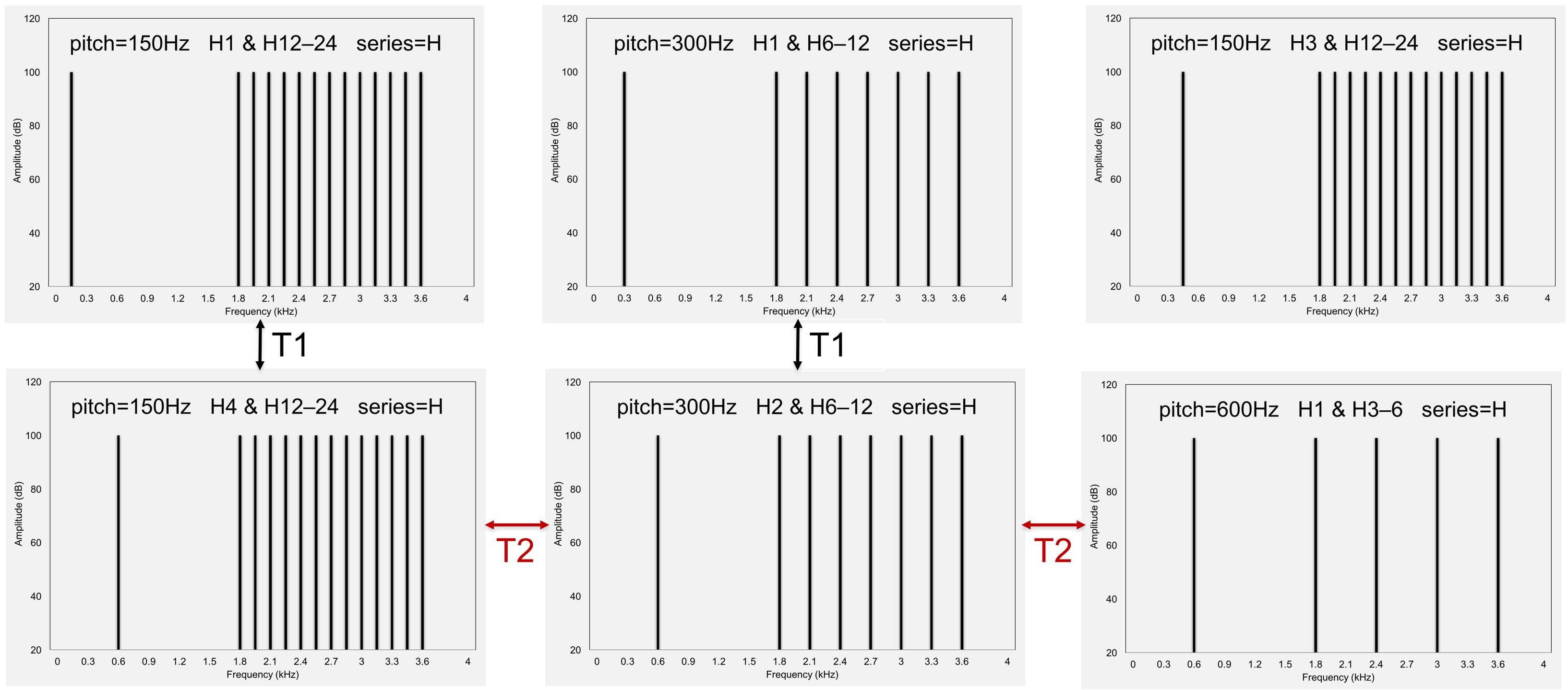


Figure 2: Comparison of six sounds of front vowels related to a pitch variation of 2 octaves (see Table 1, Series H).



Listening test

Vowel recognition was investigated by means of two listening tests, involving five phonetic expert listeners who assigned the sounds (Test A) or sound pairs (Test B) to Standard German vowel qualities. Test A = all single sounds were presented in random order (maximum of votes/sound = 5). Test B = for each single series separately, all possible pair combinations of sounds (including their order in a pair) were presented in a subtests, with a pause in between the substests (Series A–G = 20 votes/sound in total, Series H = 50 votes/sound in total).

Results

The results of the listening tests in Table 1 show

- **consistent perceptual open-closed shifts in vowel quality for all sound pairs and sound triples tested which relate to the comparison type T2, with**
- **shifts to an adjacent vowel quality for pitch variations of 1–1.5 octave, and**
- **a shift to a non-adjacent vowel quality for pitch variations of 2 octaves.**

References (continuation)
[5] Maurer, D. (2016): Acoustics of the Vowel – Preliminaries. Bern/Frankfurt: Peter Lang.
[6] Maurer, D., Dellwo, V., Suter, H., Kathiresan, T. (2017): Formant pattern ambiguity of vowel sounds revisited in synthesis: Changing perceptual vowel quality by only changing fundamental frequency. Parallel Poster, Acoustics '17, Boston.
[7] Maurer, D., and Landis, T. (1995): F0-dependence, number alteration, and non-systematic behaviour of the formants in German vowels. International Journal of Neuroscience, 83, 25-44.

Results (continuation)

Table 1: Series of sound comparisons. Comparisons relating to T2 are marked in red. The perceived pitch (Column 2) relates to the frequency spacing ΔH of the consecutive harmonics > 1kHz. Columns 3 to 6 indicate the harmonic structure and the related frequencies/frequency ranges of the harmonics in the sound spectrum. Columns 7 to 18 show the results (confusion matrices) of the first and second listening test, with maj. = the majority of the listeners votes.

Sound series	Pitch	Sound spectrum				Listening test A						Listening test B					
	ΔH	Harmonics H(i)		Frequencies (Hz)		Confusion matrix					maj.	Confusion matrix					maj.
	Hz	< 1kHz	> 1kHz	< 1kHz	> 1kHz	ε	ø	e	y	i		ε	ø	e	y	i	
A	220	H1	H10–H16	220	2200–3520				1	4	i			1	4	15	i
	220	H2	H10–H16	**440**	**2200–3520**			**5**			**e**			**20**			**e**
	440	H1	H5–H8	**440**	**2200–3520**				1	**4**	**i**				1	**19**	**i**
B	220	H1	H6–H10	220	1320–2200				5		y				20		y
	220	H2	H6–H10	**440**	**1320–2200**		**5**				**ø**		**20**				**ø**
	440	H1	H3–H5	**440**	**1320–2200**		2		**3**		**y**		1		**19**		**y**
C	220	H2	H12–H20	440	2200–3520			5			e			20			e
	220	H4	H12–H20	**880**	**2200–3520**	**5**					**ε**	**20**					**ε**
	440	H2	H5–H8	**880**	**2200–3520**	2		**3**			**e**	2		**18**			**e**
D	150	H2	H15–H24	300	2250–3600					5	i			2		18	i
	150	H3	H15–H24	**450**	**2250–3600**			5			**e**	2		**18**			**e**
	450	H1	H5–H8	**450**	**2250–3600**					5	**i**					**20**	**i**
E	150	H2	H15–H21	300	2250–3150					5	i				8	12	i
	150	H3	H15–H21	**450**	**2250–3150**			**5**			**e**			**20**			**e**
	450	H1	H5–H7	**450**	**2250–3150**					**5**	**i**				5	**15**	**i**
F	150	H2	H9–H15	300	1350–2250				5		y				20		y
	150	H3	H9–H15	**450**	**1350–2250**		**5**				**ø**		**18**	2			**ø**
	450	H1	H3–H5	**450**	**1350–2250**		1		**4**		**y**				**20**		**y**
G	150	H3	H12–H20	450	1800–3000		2	3			e		5	15			e
	150	H4	H12–H20	**600**	**1800–3000**	**5**					**ε**	**20**					**ε**
	300	H2	H6–H10	**600**	**1800–3000**		2	**3**			**e**		1	**19**			**e**
H	150	H2	H12–H24	300	1800–3600				4	1	y				38	12	y
	300	H1	H6–H12	300	1800–3600				4	1	y				31	19	y
	150	H3	H12–H24	450	1800–3600			5			e	3	10	37			e
	150	H4	H12–H24	**600**	**1800–3600**	**5**					**ε**	**50**					**ε**
	300	H2	H6–H12	**600**	**1800–3600**			**5**			**e**			**50**			**e**
	600	H1	H3–H6	**600**	**1800–3600**				**4**	1	**y**				**25**	**25**	**y-i**

Discussion

Given two synthesised sounds, for which the frequency range of equal amplitude harmonics > 1kHz corresponds, **the perceived vowel quality can be changed by** either altering the frequency of a single lower equal amplitude harmonic or **only changing the frequency spacing of the higher harmonics.** In the first case, the perceptual change may be understood as resembling a change in the first formant frequency. However, in the second case, the perceptual change seems to indicate vowel-related spectral characteristics being pitch-related. (Note, that pitch-related lower spectral characteristics were observed for natural vowel sounds [5, 7], and that the perceived vowel quality can be changed by only changing the fundamental frequency [4–6].) – The investgation of vowel sounds related to "flat" spectra offers a promising approach for the search of a relation between the perception of vowel qualities and vowel-specific acoustic characteristics. Thereby, sounds manifesting "flat" spectra can only be considered in a perspective of the whole spectrum, but not in a perspective of spectral peaks in terms of formant patterns.