hdk
Zürcher Hochschule der Künste
Zurich University of the Arts

Universität Zürich[UZH]

Sinewave vowel sounds: The role of vowel qualities, frequencies and harmonicity of sinusoids, and perceived pitch for vowel recognition

175th Meeting of the
Acoustical Society of America
Minneapolis, 7–11 May 2018

Dieter Maurer[1], Heidy Suter[1], Thayabaran Kathiresan[2] 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
dieter.maurer@zhdk.ch, heidy.suter@zhdk.ch, thayabaran.kathiresan@uzh.ch, volker.dellwo@uzh.ch **www.phones-and-phonemes.org/asa/2018a**

4aSC4
Journal of the Acoustical Society
of America 143(3): 1920

Questions and general approach

The present study questioned vowel quality and pitch recognition of sinewave vowels (three sinusoids S_1–S_2–S_3) in three experiments:

(1) S_1–S_2–S_3 directly related to statistical formant patterns F_1–F_2–F_3 of women, but harmonically unrelated (replicas of formant patterns)

(2) S_1–S_2–S_3 indirectly related to statistical F_1–F_2–F_3 of women, harmonically interrelated by manual "correction" (replicas of "harmonically corrected" formant patterns in terms of creating S_1–S_2–S_3 with a highest common factor HCF)

(3) S_1–S_2–S_3 pairs related to fixed S_1–S_3, only varying S_2, all sinewaves in harmonic relation but with different HCF (difference = 1 octave)

The main aspects of method and the major results are presented here. Details are given online, including the synthesised sounds (see [1]).

Vowels and sinewave synthesis

Vowel qualities investigated: Closed and mid-closed vowels were investigated because formant patterns of sounds of these vowels strongly relate to fundamental frequency (f_o) and, consequently, single formant patterns as well as a single spectral envelopes often represent different vowel qualities for sounds with different f_o. [2–4]

Synthesis type 1 (Table 1): Replicas of formant patterns. S_1–S_2–S_3 replicas of statistical F_1–F_2–F_3 for Standard German /i–y–e–ø–o–u/, of women (see [5]) were synthesised. Note that in such direct replication, no harmonicity exists.

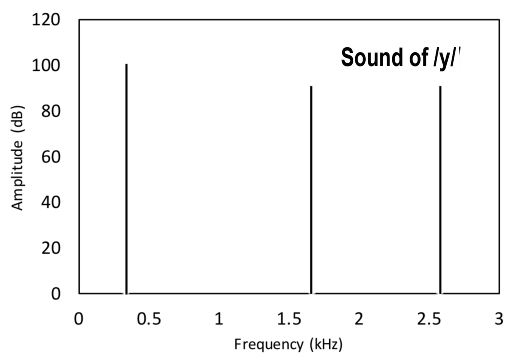
Synthesis type 2 (Table 2): Replicas of "harmonically corrected" formant patterns. The above S_1–S_2–S_3 patterns were manipulated in order to create harmonically related frequencies (creating S_1–S_2–S_3 with a HCF): S_1 was set = 330 Hz for the sounds of closed vowels and = 440 Hz for the sounds of mid-closed vowels. For each vowel, depending on the formant frequency configuration of the patterns of natural sounds, two or three S_2–S_3 versions (near to original F_2–F_3) were set so as to create two different harmonic relations, with HCF = 165Hz or 330Hz for closed vowels, and 220Hz or 440Hz for mid-closed vowels.

Synthesis type 3 (Table 3): Sinewave pairs with fixed S_1–S_3, varying S_2, maintaining harmonicity, but changing the HCF by one octave. Pairs of S_1–S_2–S_3 patterns were synthesised, with fixed S_1–S_3 but varying S_2, the sinewaves always in harmonic relation. Because a smaller frequency change for lower harmonics is related to a larger change in the higher harmonics, which might affect vowel recognition, three configurations of one-octave HCF variation = 200–400Hz, 210–420Hz and 220–440Hz were investigated for front vowels; however, only one configuration of HCF variation = 200–400Hz was investigated for back vowels.

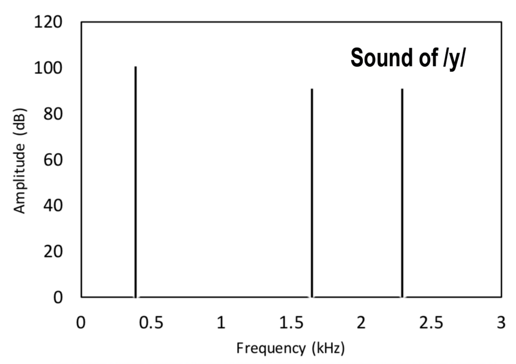
Amplitudes: Sinewave amplitudes A_1–A_2–A_3 were set = 100–90–90 dB for sounds of front vowels and 100–90–70 dB for sounds of back vowels.

References:
[1] Maurer, D., Suter, H., Kathiresan, T., Dellwo, V. (2018). Sinewave vowel sounds: The role of vowel qualities, frequencies and harmonicity of sinusoids, and perceived pitch for vowel recognition. Materials. Accessible online: http://www.phones-and-phonemes.org/asa/2018a (retreived on 6 May 2018).
[2] 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.
[3] Maurer, D., Landis, T. (2000): Formant pattern ambiguity of vowel sounds. International Journal of Neuroscience, 100 (1-4), 39-76.
[4] 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.
[5] Pätzold, M., Simpson, A. (1997). Acoustic analysis of German vowels in the Kiel Corpus of Read Speech. Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung Universität Kiel, 32(1978), 215–247.

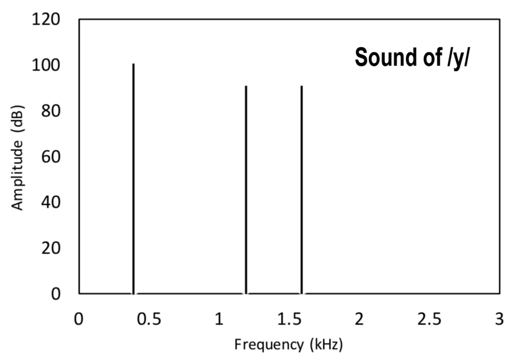
Sinewave syntheses – Illustration



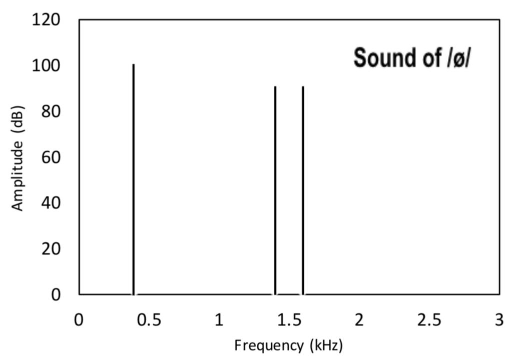
1: S_1–S_2–S_3=342–1667–2585Hz, no HCF



2: S_1–S_2–S_3=330–1650–2310Hz, HCF=330Hz



3a: S_1–S_2–S_3=400–1200–1600Hz, HCF=400Hz



3b: S_1–S_2–S_3=400–1400–1600Hz, HCF=200Hz

Figure 1: Example for synthesis type 1, S_1–S_2–S_3 replicating a statistical formant pattern F_1–F_2–F_3 for /y/ of women (Standard German) = 342–1667–2585 Hz. No HCF.

Figure 2: Example for synthesis type 2, replicating a "harmonically corrected" formant pattern F_1–F_2–F_3 for /y/ = 330–1650–2310Hz. HCF = 330Hz.

Figure 3: Example for synthesis type 3, sound pair with identical S_1–S_3 = 400–1600 Hz and S_2 variation = 1200Hz and 1400Hz, effecting HCF variation 400Hz to 200Hz; recognised as /y/ and /ø/.

Vowel quality and pitch recognition tests

For the sound samples of each experiment separately, 5 phonetic expert listeners (3 women, 2 men) performed two listening tests: **Vowel recognition task** according to Standard German vowel qualities and /ə/ (sounds presented twice); **pitch recognition task** using a virtual electronic piano. For details, see [1].

Results

(1) Table 1: Recognition of sinewave vowels replicating formant patterns as given for relaxed speech depend on vowel quality. Replicas of closed vowels were mainly recognised as closed vowels, but replicas of mid-closed vowels were perceptually confused with closed vowels.

(2) Table 2: The same holds true even if a harmonic relation for sinewave vowels replicating the above formant patterns is created.

(3) Table 3: Manipulating HFC by only changing the S_2 frequency can effect a change in the recognised vowel quality, even if lowering S_2 effects an upwards shift in vowel height (see the sound pairs of front vowels). Table 3 shows a selection of examples of corresponding sound pairs. Note the indication of a parallel shift of increasing HCF, increasing pitch and vowel height. Note also that the frequency distance of S_2–S_3 is < 3 Bark. For further details, see [1].

Additional findings for sinewave frequencies: (i) Substantial variations of S_2 and S_3 corresponded to sounds of a single vowel category, above all for sounds of front vowels (compare vowel related values in the tables). (ii) Vowel recognition for sounds of synthesis type 3 proved to be highly prone on S_1 and the effect on S_2 and S_3 as its multiples (see [1]).

Additional findings for pitch recognition: (i) Sinewave replicas of formant patterns lacking harmonicity (HCF) were perceived as having a pitch. (ii) Pitch recognition is difficult to interpret and may relate to S_1 and/or to HCF and/or quasi-periodicity of the signal and/or perceptual octave confusion.

[6] Hillenbrand, J. M., Clark, M. J., Baer, C. A. (2011). Perception of sinewave vowels. The Journal of the Acoustical Society of America, 129(6), 3991–4000.
[7] Maurer, D. (2016): Acoustics of the Vowel – Preliminaries. Bern/Frankfurt: Peter Lang.

Tables (synthesis and results of the listening tests)

Table 1: Results of the listening tests for experiment 1.

	Vowel	Sinewave synthesis (Hz)				Vowel	Maj.	Pitch rec. (Hz)				
	int.	S1	S2	S3	HCF	rec.		165	330	220	440	
	i	329	2316	2796	–	i	7		5			closed
1	y	342	1667	2585	–	y	10		5			closed
	u	350	825	2795	–	u	9	1	4			closed
	e	431	2241	2871	–	i	7			1	4	mid-closed
	ø	434	1646	2573	–	y	6			1	4	mid-closed
	o	438	953	2835	–	u	8			2	3	mid-closed

Table 2: Results of the listening tests for experiment 2.

	Vowel	Sinewave synthesis (Hz)				Vowel	Maj.	Pitch rec. (Hz)				
	int.	S1	S2	S3	HCF	rec.		165	330	220	440	
	i	330	2310	2640	**330**	i	8		5			closed
	i	330	2475	2640	**165**	i	9	2	3			closed
	y	330	1650	1980	**330**	y	10	1	4			closed
	y	330	1650	1815	**165**	y	10	2	3			closed
2	y	330	1650	2310	**330**	y	10	1	4			closed
	u	330	990	2640	**330**	u	10	1	4			closed
	u	330	1155	2640	**165**	u	10	1	4			closed
	u	330	825	2640	**165**	u	9	1	4			closed
	e	440	2200	2640	**440**	i	6			1	4	mid-closed
	e	440	2420	2640	**220**	e–i	5-5			2	3	mid-closed
	ø	440	1760	2200	**440**	y	9			1	4	mid-closed
	ø	440	1760	1980	**220**	y	8			2	3	mid-closed
	ø	440	1760	2640	**440**	y	9			1	4	mid-closed
	o	440	880	2640	**440**	u	6			1	4	mid-closed
	o	440	1100	2640	**220**	u	7			2	3	mid-closed

Table 3: Results of the listening tests for experiment 3.

	Vowel	Sinewave synthesis (Hz)				Vowel	Maj.	Pitch rec. (Hz)			
	int.	S1	S2	S3	HCF	rec.		200	400	220	440
	–	420	**2730**	2940	**210**	e	6			2	3
	–	420	**2520**	2940	**420**	i	9			1	4
	–	400	**2200**	2400	**200**	e	7	3	2		
	–	400	**2000**	2400	**400**	y	10	1	4		
3b	–	400	**1400**	1600	**200**	ø	8	3	2		
3a	–	400	**1200**	1600	**400**	y	7	1	4		
	–	400	**600**	2800	**200**	o	10	5			
	–	400	**800**	2800	**400**	u	7	1	4		

Tables 1 to 3: Results of the listening tests.

Columns
1 Vowel quality intended
2–4 Sinewave frequencies S_1–S_2–S_3
5 Harmonicity (HCF)
6 Vowel quality recognised
7 Majority of vowel recognition (maximmum = 10, majority >5)
8–11 Pitch recognised (maximum = 5; frequencies according to musical scale)

Colours in Tables 1, 2 (experiments 1, 2)
Blue Vowel quality intended, and vowel quality recognised correspondingly
Purple Vowel quality recognised differently compared with the intended vowel

Colour in Table 3 (experiment 3)
Orange S_2 variation effecting a change in the recognised vowel quality

Vowel recognition
Tables 1 shows correspondence of vowel intention and recognition for sinewave sounds replicating formant patterns of natural sounds of closed vowels, but vowel confusion for the replicas of mid-closed vowels.
Tables 2 shows similar vowel recognition of "harmonically corrected" sinewave replicas as found in experiment 1.
Tables 3 shows that a parallel change in HCF and S_2 can effect a change in the recognised vowel quality. Noteworthy, this change is indicated torelate to HCF and not to S_2 frequency. Therefore, lowering S_2 can effect an open–closed shift in vowel quality.

Pitch recognition
Sounds with no HCF in the spectrum were perceived as having a pitch. Besides, the effect of pitch on vowel recognition remains a matter of further investigation.

Discussion

Concerning sinewave replicas of statistical formant patterns, this study confirms earlier indications of **vowel confusion relating to vowel height for untrained vowel recognition** (see Table 1 in [6]), and it shows that this relation is not due to a lack of harmonicity. Further, the study presents **cases in which mid-closed to closed shifts in vowel recognition were effected by in parallel lowering S_2 and doubling HCF, in contrast to a common assumption that F_2 rises with increasing vowel height. Note that pitch recognition is indicated to play a role in this vowel shift.** However, the additional effect of different amplitude ratios needs further investigation.

The same holds true for the general relation between vowel recognition, HCF and pitch recognition. Thereby, the effect of pitch on vowel recognition for aperiodic sounds and for sounds with a "missing fundamental" are of specific interest.

We conclude that no direct relation exists between formant patterns, sinewave replicas and vowel recognition, i.e. **sinewave vowels do not support the assumption of formant patterns as per se vowel quality specific.** We interpret the findings as relating to the reported (unsystematic) f_o-dependence of formant patterns and spectral envelopes, the resulting ambiguity of these spectral features [2–4] and the assumption of vowel sounds as a phenomenon of foreground-background [7].