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

Materials:
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

Synthesis:
Christian d'Heureuse
http://www.source-code.biz/

Corrections included:
Version 2019-11-04

Content

1 Listening tests (details)

The five professionally trained speakers or singers performing the standard listening test of the Zurich Corpus also participated in the recognition experiment of this study. For the sample of the six synthesised sounds, two separate recognition tests were conducted.

For vowel recognition (first test performed), the sounds were presented in random order, with each sound presented twice in the test. The listeners were asked to label one of the long Standard German vowels /i–y–e–ø–ɛ–a–ɔ–o–u/ or /ə/ (Schwa) or “no vowel recognised”. Before labeling, the listeners listened to the sounds of the entire test set (in random order) so as to get familiar with the sound quality of the synthesised sounds. – Sounds recognised as lying in the region of /ɑ–a/ were labeled as /a/. The vowel /ɔ/ was included as an option in the recognition task because the phonetic distance /ɑ–a/–/o/ exceeds the distance of the other neighbouring long Standard German vowels. Schwa was included to allow for the assignment of a sound to be recognised as a central vowel.

For pitch recognition (second test performed), the sounds were presented in random order, with each sound presented only once in the test. The listeners were asked to label the recognised pitch using a virtual electronic piano (assignment of the nearest musical note). Piano key frequencies recognised by the listeners and shown in the tables were as follows:

- According to musical scale = E3–F3–G3–G♯3–A3 and E4–F4–G4–G♯4–A4
- According to frequency in Hz (rounded) = 165–175–196–208–220 and 330–349–392–415–440

In both tests, the listeners were allowed to play back a sound one or multiple times before labeling a vowel quality or a pitch frequency.

2 Results (details)

2.1 Sinewave synthesis 1, replicas of formant patterns (Table 1, and additions)

Indications in the Tables 1–3

In Tables 1–3, the synthesis parameters and the results of the listening tests are given.
Column indications are as follows:

1	Vowel quality intended as given in the formant statistics mentioned (see [1]).
2–4	Sinewave frequencies S_1–S_2–S_3
5	Harmonicity (highest common factor HCF; lacking for the direct formant pattern replication in first synthesis experiment)
6–7	Vowel quality recognised in the listening test in terms of the majority of vowel recognition (maximum = 10; majority equal > 5)
8–11	Pitch recognised (maximum = 5) in terms of the assignment of the nearest musical pitch according to the C-major including all semitones

[1] 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.

Table 1

Table 1 (as shown in the poster) presents the recognition results of vowel quality and pitch for S_1–S_2–S_3 sounds replicating statistical F_1–F_2–F_3 of Standard German closed and mid-closed vowels /i–y–e–ø–o–u/ of women (see [1]).
In the Addition to Table 1, the entire confusion matrix of the vowel recognition test is given.

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.		175	330/349	220	440	
i	329	2316	2796	–	i	7		5			closed
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

Addition to Table 1

Confusion matrix

int.	a	ɔ	o	u	n.s.	ə	ɛ	ø	e	y	i
i									1	2	7
y										10	
u			1	9							
e									2	1	7
ø								4		6	
o			2	8							

Concerning vowel recognition, the results show 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.
Concerning pitch recognition, noteworthy, sinewave replicas of formant patterns lacking harmonicity (HCF) were perceived as having a pitch. However, pitch recognition is difficult to interpret in details and it may relate to S_1 and/or to HCF and/or quasi-periodicity of the signal and/or perceptual octave confusion. Nevertheless, pitch recognition is indicated as possibly related to S_1 or $0.5\times S_1$ (replicating F_1 or $0.5\times F_1$), with a lower pitch range for sounds of closed than for mid-closed vowels.

Related sounds: Below, for each single sinewave vowel, the corresponding synthesis parameters and a link to the synthesised sound is given. For the synthesis tool, see
http://www.source-code.biz/sinSyn/ (Christian d'Heureuse; retrieved April 29, 2018)

Note: Please use high quality headphones and lower volume to listen to the sounds. Don't rely on laptop integrated loudspeakers.

Closed vowels (correspondence of intention and recognition of vowel quality)

Sound 1, natural sound intended as /i/, sinewave replica of F1–F2–F3 correspondingly recognised by the majority of listeners as /i/.
S_1–S_2–S_3 = 329–2316–2796 Hz A_1–A_2–A_3 = 100–90–90 dB
Link to synthesised sound

Sound 2, natural sound intended as /y/, sinewave replica correspondingly recognised as /y/.
S_1–S_2–S_3 = 342–1667–2585 Hz A_1–A_2–A_3 = 100–90–90 dB
Link to synthesised sound

Sound 3, natural sound intended as /u/, sinewave replica correspondingly recognised as /u/.
350 100 1048 90 2795 70
S_1–S_2–S_3 = 350–1048–2795 Hz A_1–A_2–A_3 = 100–90–70 dB
Link to synthesised sound

Mid-closed vowels (confusion between intention and recognition of vowel quality)

Sound 4, natural sound intended as /e/, but sinewave replica of F1–F2–F3 recognised as /i/.
S_1–S_2–S_3 = 431–2241–2871 Hz A_1–A_2–A_3 = 100–90–90 dB
Link to synthesised sound

Sound 5, natural sound intended as /ö/, but sinewave replica of F1–F2–F3 recognised as /y/.
S_1–S_2–S_3 = 434–1646–2573 A_1–A_2–A_3 = 100–90–90
Link to synthesised sound

Sound 6, natural sound intended as /o/, but sinewave replica of F1–F2–F3 recognised as /u/.
S_1–S_2–S_3 = 438–953–2835 A_1–A_2–A_3 = 100–90–90
Link to synthesised sound

2.2 Sinewave synthesis 2, replicas of “harmonically corrected” formant patterns (Table 2, and additions)

Table 2 (as shown in the poster) presents the recognition results of vowel quality and pitch for S_1–S_2–S_3 sounds replicating “harmonically corrected” statistical F_1–F_2–F_3 of Standard German closed and mid-closed vowels: The above S_1–S_2–S_3 patterns of synthesis experiment 1 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.
In the Addition to Table 2, the entire confusion matrix of the vowel recognition test is given.

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
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

Addition to Table 2

Confusion matrix

a	ɔ	o	u	n.s.	ə	ɛ	ø	e	y	i
									2	8
								1		9
									10	
									10	
									10	
			10							
			10							
		1	9							
								1	3	6
								5		5
									9	1
							2		8	
									9	1
		4	6							
		3	7							

The results generally correspond the findings in the first experiment. Thus, vowel quality correspondence for the sounds of closed vowels and vowel confusion for the sounds of mid-closed vowels do not (or not directly) relate to harmonicity.

Related sounds: Below, for each single sinewave vowel, the corresponding synthesis parameters and a link to the synthesised sound is given.

Closed vowels (correspondence of intention and recognition of vowel quality)

Sound 1, natural sound intended as /i/, sinewave replica of F1–F2–F3 correspondingly recognised by the majority of listeners as /i/.
S_1–S_2–S_3 = 330–2310–2640 Hz A_1–A_2–A_3 = 100–90–90 dB HFC = 330 Hz
Link to synthesised sound

Sound 2, natural sound intended as /i/, sinewave replica of F1–F2–F3 correspondingly recognised by the majority of listeners as /i/.
S_1–S_2–S_3 = 330–2475–2640 Hz A_1–A_2–A_3 = 100–90–90 dB HFC = 165 Hz
Link to synthesised sound

Sound 3, natural sound intended as /y/, sinewave replica of F1–F2–F3 correspondingly recognised by the majority of listeners as /y/.
S_1–S_2–S_3 = 330–1650–1980 Hz A_1–A_2–A_3 = 100–90–90 dB HFC = 330 Hz
Link to synthesised sound

Sound 4, natural sound intended as /y/, sinewave replica of F1–F2–F3 correspondingly recognised by the majority of listeners as /y/.
S_1–S_2–S_3 = 330–1650–1815 Hz A_1–A_2–A_3 = 100–90–90 dB HFC = 165 Hz
Link to synthesised sound

Sound 5, natural sound intended as /y/, sinewave replica of F1–F2–F3 correspondingly recognised by the majority of listeners as /y/.
S_1–S_2–S_3 = 330–1650–2310 Hz A_1–A_2–A_3 = 100–90–90 dB HFC = 330 Hz
Link to synthesised sound

Sound 6, natural sound intended as /u/, sinewave replica of F1–F2–F3 correspondingly recognised by the majority of listeners as /u/.
S_1–S_2–S_3 = 330–990–2640 Hz A_1– A_2– A_3 = 100–90–70 dB HFC = 330 Hz
Link to synthesised sound

Sound 7, natural sound intended as /u/, sinewave replica of F1–F2–F3 correspondingly recognised by the majority of listeners as /u/.
S_1–S_2–S_3 = 330–1155–2640 Hz A_1–A_2–A_3 = 100–90–70 dB HFC = 165 Hz
Link to synthesised sound

Sound 8, natural sound intended as /u/, sinewave replica of F1–F2–F3 correspondingly recognised by the majority of listeners as /u/.
S_1–S_2–S_3 = 330–825–2640 Hz A_1–A_2–A_3 = 100–90–70 dB HFC = 165 Hz
Link to synthesised sound

Mid-closed vowels (confusion between intention and recognition of vowel quality)

Sound 9, natural sound intended as /e/, but sinewave replica of F1–F2–F3 recognised as /i/.
S_1–S_2–S_3 = 440–2200–2640 Hz A_1–A_2–A_3 = 100–90–90 dB HFC = 440 Hz
Link to synthesised sound

Sound 10, natural sound intended as /e/, but sinewave replica of F1–F2–F3 recognised as /e/ or /i/.
S_1–S_2–S_3 = 440–2420–2640 Hz A_1–A_2–A_3 = 100–90–90 dB HFC = 220 Hz
Link to synthesised sound

Sound 11, natural sound intended as /ø/, but sinewave replica of F1–F2–F3 recognised as /y/.
S_1–S_2–S_3 = 440–1760–2200 Hz A_1–A_2–A_3 = 100–90–90 dB HFC = 440 Hz
Link to synthesised sound

Sound 12, natural sound intended as /ø/, but sinewave replica of F1–F2–F3 recognised as /y/.
S_1–S_2–S_3 = 440–1760–1980 Hz A_1–A_2–A_3 = 100–90–90 dB HFC = 220 Hz
Link to synthesised sound

Sound 13, natural sound intended as /ø/, but sinewave replica of F1–F2–F3 recognised as /y/.
S_1–S_2–S_3 = 440–1760–2640 Hz A_1–A_2–A_3 = 100–90–90 dB HFC = 440 Hz
Link to synthesised sound

Sound 14, natural sound intended as /o/, but sinewave replica of F1–F2–F3 recognised as /u/.
S_1–S_2–S_3 = 440–880–2640 Hz A_1–A_2–A_3 = 100–90–70 dB HFC = 440 Hz
Link to synthesised sound

Sound 15, natural sound intended as /o/, but sinewave replica of F1–F2–F3 recognised as /u/.
S_1–S_2–S_3 = 440–1100–2640 Hz A_1–A_2–A_3 = 100–90–70 dB HFC = 220 Hz
Link to synthesised sound

2.3 Sinewave synthesis 3, sinewave pairs with fixed S_1–S_3, varying S_2, maintaining harmonicity, but changing the HCF by one octave (selection; Table 3, and additions)

Table 3 (as shown in the poster) presents the recognition results of vowel quality and pitch for selected S_1–S_2–S_3 sound pairs with fixed S_1–S_3 but different S_2 effecting a change in HFC of one octave.
In the Addition to Table 3, the entire confusion matrix of the vowel recognition test is given.

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	210	420
–	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		
–	400	**1400**	1600	**200**	**ø**	**8**	3	2		
–	400	**1200**	1600	**400**	**y**	**7**	1	4		
–	400	**600**	2800	**200**	**o**	**10**	5			
–	400	**800**	2800	**400**	**u**	**7**	1	4		

Addition to Table 2

Confusion matrix

a	ɔ	o	u	n.s.	ə	ɛ	ø	e	y	i
								6		4
									1	9
								7	3	
									10	
			1				8		1	
			2				1		7	
		10								
		3	7							

Concerning vowel recognition, the results show that a parallel change in S_2 and HFC can effect a change in the recognised vowel quality. Noteworthy, this change is indicated to relate to HCF and not to S_2 frequency. Therefore, lowering S_2 can effect an open–closed shift in vowel quality, as is shown above all for the sound pairs recognised as /e–i/ and /ø–y/.

Related sounds: Below, for each single sinewave vowel, the corresponding synthesis parameters and a link to the synthesised sound is given.

Sound pair recognised as /e–i/

Sound 1, recognised as /e/.
S_1–S_2–S_3 = 420–2730–2940 Hz A_1–A_2–A_3 = 100–90–90 dB HFC = 220 Hz
Link to synthesised sound

Sound 2, recognised as /i/.
S_1–S_2–S_3 = 420–2520–2940 Hz A_1–A_2–A_3 = 100–90–90 dB HFC = 440 Hz
Link to synthesised sound

Sound pair recognised as /e–y/

Sound 3, recognised as /e/.
S_1–S_2–S_3 = 400–2200–2400 Hz A_1–A_2–A_3 = 100–90–90 dB HFC = 200 Hz
Link to synthesised sound

Sound 4, recognised as /i/.
S_1–S_2–S_3 = 400–2000–2400 Hz A_1–A_2–A_3 = 100–90–90 dB HFC = 400 Hz
Link to synthesised sound

Sound pair recognised as /ø–y/

Sound 3, recognised as /ø/.
S_1–S_2–S_3 = 400–1400–1600 Hz A_1–A_2–A_3 = 100–90–90 dB HFC = 200 Hz
Link to synthesised sound

Sound 4, recognised as /y/.
S_1–S_2–S_3 = 400–1200–1600 Hz A_1–A_2–A_3 = 100–90–90 dB HFC = 400 Hz
Link to synthesised sound

Sound pair recognised as /o–u/

Sound 3, recognised as /o/.
S_1–S_2–S_3 = 400–600–2800 Hz A_1–A_2–A_3 = 100–90–70 dB HFC = 200 Hz
Link to synthesised sound

Sound 4, recognised as /u/.
S_1–S_2–S_3 = 400–800–2800 Hz A_1–A_2–A_3 = 100–90–70 dB HFC = 400 Hz
Link to synthesised sound

2.4 Sinewave synthesis 3, sinewave pairs with fixed S_1–S_3, varying S_2, maintaining harmonicity, but changing the HCF by one octave (full sample; Tables 4a and 4b)

For back vowels, only one configuration of HCF variation = 200–400Hz was investigated.

For front vowels, three configurations of one-octave HCF variation = 200–400Hz, 210–420Hz and 220–440Hz were investigated, because a smaller frequency change for lower harmonics is related to a larger change in the higher harmonics with increasing harmonic number, which might affect vowel recognition.

In Table 3, a selection of S_1–S_2–S_3 sound pairs with fixed S_1–S_3 but different S_2 effecting a change in HFC of one octave is shown. Below, the entire sound sample of experiment 3 and the results of the listening tests are shown in Tables 4a and 4b.

For synthesis replication, please use the above link.

Note the amplitude levels used:
- A1–A2–A3 = 100–90–70 dB for sounds of back vowels
- A1–A2–A3 = 100–90–90 dB for sounds of front vowels

Note also the fading used = 0.1 sec.

Table 4a: Results of the listening tests for experiment 3, sounds of back vowels.

Vowel	Sinewave synthesis (Hz)				Vowel	Maj.	Pitch rec. (Hz)					
int.	S1	S2	S3	HCF	rec.	>5	196	392	208	415	220	440
–	400	**600**	2800	200	**o**	**10**	5					
–	400	**800**	2800	400	**u**	**7**	1	4				
–	400	**1000**	2800	200	**–**	**–**	3	2				
–	400	**1200**	2800	400	**u**	**8**	1	4				

Addition to Table 4a

Confusion matrix

a	ɔ	o	u	n.s.	ə	ɛ	ø	e	y	i
		10								
		3	7							
		4	5	1						
1			8				1			

(Table 4b see next page)

Table 4b: Results of the listening tests for experiment 3, sounds of front vowels.												
Vowel	**Sinewave synthesis (Hz)**				**Vowel**	**Maj.**	**Pitch rec. (Hz)**					
int.	S1	S2	S3	HCF	**rec.**	**>5**	196	392	208	415	220	440
–	400	**1400**	1600	200	**ø**	**8**	3	2				
–	400	**1200**	1600	400	**y**	**7**	1	4				
–	400	**1800**	2000	200	**y**	**7**	3	2				
–	400	**1600**	2000	400	**y**	**9**	1	4				
–	400	**2200**	2400	200	**e**	**7**	3	2				
–	400	**2000**	2400	400	**y**	**10**	1	4				
–	400	**2600**	2800	200	**e**	**6**	3	2				
–	400	**2400**	2800	400	**i**	**6**	1	4				
–	400	**3000**	3200	200	**i**	**8**	2	3				
–	400	**2800**	3200	400	**i**	**10**	1	4				
–	410	**1470**	1680	205	**ø**	**8**			3	2		
–	410	**1260**	1680	410	**–**				1	4		
–	410	**1890**	2100	205	**y**	**7**			3	2		
–	410	**1680**	2100	410	**y**	**9**			1	4		
–	410	**2310**	2520	205	**e**	**7**			3	2		
–	410	**2100**	2520	410	**y**	**7**			1	4		
–	410	**2730**	2940	205	**e**	**6**			2	3		
–	410	**2520**	2940	410	**i**	**9**			1	4		
–	420	**1540**	1760	210	**ø**	**7**					3	2
–	420	**1320**	1760	420	**–**						1	4
–	420	**1980**	2200	210	**–**						2	3
–	420	**1769**	2200	420	**y**	**9**					1	4
–	420	**2420**	2640	210	**e**	**6**					2	3
–	420	**2200**	2640	420	**y**	**7**					1	4
–	420	**2860**	3080	210	**–**						2	3
–	420	**2640**	3080	420	**i**	**9**					1	4

Addition to Table 4b											
Confusion matrix											
a	ɔ	o	u	n.s.	ə	ɛ	ø	e	y	i
			1				8		1	
			2				1		7	
							2	1	7	
							1		9	
								7	3	
									10	
								6	1	3
								1	3	6
								2		8
										10
							8		2	
			2				2		5	1
								3	7	
							1		9	
								7	3	
							1		7	2
								6		4
									1	9
							7		3	
			1				5		4	
							1	5	4	
							1		9	
								6	2	2
								1	7	2
									5	5
									1	9