Formant pattern and spectral shape ambiguity in vowel synthesis: The role of fundamental frequency and formant amplitude

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

4aSC3
Journal of the Acoustical Society
of America 143(3): 1919–1920

Zürcher Hochschule der Künste
Zurich University of the Arts

Universität Zürich[UZH]

Thayabaran Kathiresan[1], Dieter Maurer[2], Heidy Suter[2], Volker Dellwo[1]
[1] Department of Computational Linguistics, University of Zurich, Switzerland, [2] Institute for the Performing Arts and Film, Zurich University of the Arts, Switzerland
thayabaran.kathiresan@uzh.ch, dieter.maurer@zhdk.ch, heidy.suter@zhdk.ch, volker.dellwo@uzh.ch **www.phones-and-phonemes.org/asa/2018b**

Background – Previous study

When investigating formant pattern and spectral shape ambiguity in Klatt synthesis, an earlier study entitled "Formant pattern and spectral shape ambiguity of vowel sounds revisited in synthesis: Changing perceptual vowel quality by only changing the fundamental frequency" showed that the perceived vowel quality of Standard German vowel sounds can be changed by varying fundamental frequency only [1].

Follow-up study

In this follow-up study, the previous original synthesis experiment was repeated three times, with fundamental frequencies (f_o) of the corresponding sounds lowered by one octave, and with different ratios of the first and second formant amplitudes in terms of formant levels L_1' and L_2'.

Thus – for the investigation of formant pattern and spectral shape ambiguity in vowel synthesis – the role of the f_o range and of a limited variation of L_1' and L_2' was further examined.

Previous experiment

Synthesis (see Table 1, Klatt synthesis parameters): 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 levels of f_o (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 observed spectral envelopes of natural vowel sounds, 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 ([2], cascade mode, sampling frequency SF = 44.1kHz).

Replication with f_o and $L_{1'} – L_{2'}$ variation

Synthesis: Synthesis of the above sound sample was repeated three times creating three new samples, differing from the original sample of the previous experiment in terms of

Replication experiment 1 (RE1): f_o lowered by one octave

Replication experiment 2 (RE2): L_1' versus L_2' altered by -10dB / +10dB

Replication experiment 3 (RE3): L_1' versus L_2' altered by -20dB / +10dB

Listening tests

The same five phonetic expert listeners that participated in the previous experiment also identified all of the newly synthesised sounds (random order) in a multiple-choice identification task according to Standard German vowel qualities and (/ə/). Each sound was presented twice. For details, see [3].

Results

The general results are given in Table 1 in terms of comparing the listening test results of the original study (see Column "Orig.") and of the three replication experiments (see Columns "RE1, RE2, RE3"; vowel recognition in terms of the majority of the listener's assignments). Colour indications in Table 1 are as follows:

Blue bars indicate sound pairs or triples showing changes in the recognised vowel quality as an effect of changing f_o in the original study.

Vowel recognition results of the original experiment is highlighted in green (see Column "Orig."). If, for the sounds pairs or tripples in question, no substantial change in vowel recognition was found in the replication experiments compared with the the original experiment, the results are also highlighted in green (see Columns "RE2 ", "RE3").

However, if substantial changes occured, the results are highlighted in orange.

As is shown in Table 1:

- **For f_o variation in the range of 100–200Hz or 100–150–300Hz, no or only limited changes in the recognised vowel quality were found (compare Columns "Orig." and "RE1"), in strong contrast to the original experiment with f_o variation in the range of 200–400Hz or 200–300–600Hz;**
- **For L_1'–L_2' ratio variations, vowel recognition is clearly affected only for sounds of back vowels (compare Columns "Orig.", "RE2" and "RE3").**

The results of the listening tests in terms of detailed confusion matrices are presented online (see [3]).

References:
[1] Maurer, D., Dellwo, V., Suter, H., Kathiresan, T. (2017): Formant pattern and spectral shape ambiguity of vowel sounds revisited: Changing perceptual vowel quality by only changing the fundamental frequency. Journal of the Acoustical Society of America, 141(5), 3469–3470.
[2] Boersma, P., & Weenink, D. (2015). Praat: doing phonetics by computer. [Computer program]. Version 6.0.28. Accessible online: http://www.praat.org (retrieved on 30 March 2017).
[3] Kathiresan, T., Maurer, D., Suter, H., Dellwo, V. (2018): Formant pattern and spectral shape ambiguity of vowel sounds revisited in synthesis: The role of funramental frequency and formant amplitude. Materials. Accessible online: http://www.phones-and-phonemes.org/asa/2018b (retrieved on 6 May 2018).
[4] 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.

Table 1

	Original study											Replication 1		Replication 2		Replication 3	
	Klatt synthesis parameters (cascade mode)										↓ Orig.		↓ RE1		↓ RE2		↓ RE3
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	VR Maj.	f_o Hz	VR Maj.	$L1'–L2'$ db	VR Maj.	$L1'–L2'$ db	VR Maj.
Investigation of back vowels																	
o	200	400	100	100	800	105	100	2800	90	200	o	100	o	-10/+10	o	-20/+10	o
u	400										u	200	o	-10/+10	o	-20/+10	o
ɔ	200										ɔ	100	a	-10/+10	a	-20/+10	a
o	300	600	100	100	1200	95	100	3000	85	200	o	150	a	-10/+10	o	-20/+10	o
u	600										u	300	o	-10/+10	–	-20/+10	o
Investigation of front vowels																	
e	200	400	100	100	2400	100	200	2800	100	200	e	100	e	-10/+10	e	-20/+10	e
i	400										i	200	e	-10/+10	i	-20/+10	i
e	200	400	100	100	2800	100	200	3200	100	200	e	100	e	-10/+10	e	-20/+10	e
i	400										i	200	e	-10/+10	i	-20/+10	i
ø	200	400	100	100	2000	100	150	2800	100	200	ø	100	ø	-10/+10	ø	-20/+10	ø
y	400										y	200	ø	-10/+10	y	-20/+10	y
ɛ	200	600	100	100	2400	100	200	3000	100	200	ɛ	100	ɛ	-10/+10	ɛ	-20/+10	ɛ
e	300										e	150	ɛ	-10/+10	e	-20/+10	e
i	600										–	300	e	-10/+10	i	-20/+10	–
ɛ	200	600	100	100	1800	100	150	3000	100	200	ə-ɛ	100	ɛ	-10/+10	ɛ	-20/+10	ɛ
ø	300										ø	150	ɛ	-10/+10	ø	-20/+10	ø
y	600										y	300	ø	-10/+10	y	-20/+10	–

Table 1: Synthesis parameters, and comparison of the results of vowel recognition for the original study (Orig.) and the three replication experiments (RE1, RE2, RE3). VR = vowel recognition. Maj. = vowel quality with a majority > 50% of recognition ("–" = no majority).

Discussion

The study shows that:

- **Formant pattern and spectral shape ambiguity as a consequence of f_o variation depends strongly on the frequency range of f_o investigated; the effect of f_o on vowel recognition is thus not uniform.**
- **Above all for sounds of back vowels, vowel recognition is highly sensitive to the ratio of the formant levels L_1' and L_2'.**

These results confirm earlier indications of a non-systematic relation between f_o or pitch and formant patterns (including the role of formant amplitude) or spectral envelopes for vowel recognition ([3–6], see also [7]). The results are of importance for the design of future studies addressing the ambiguity of formant patterns and spectral envelopes in general, and corresponding synthesis experiments in particular.

References:
[5] Maurer, D., d'Heureuse, C., Landis, T. (2000): Formant pattern ambiguity of vowel sounds. International Journal of Neuroscience, vol. 100, 39–76.
[6] Maurer, D. (2016): Acoustics of the Vowel – Preliminaries. Bern/Frankfurt: Peter Lang.
[7] Kiefte, M., Enright, T., Marshall, L. (2010): The role of formant amplitude in the perception of /i/ and /u/. The Journal of the Acoustical Society of America, 127(4), 2611–21.