Mapping Timbral Spaces of Synthetic Vowels

an experimental study

Danila GOMULKIN
St. Petersburg, Russia
`www.teachers-lab.ru`

Abstract

Digitally synthesized vowel stimuli were aurally matched with reference vowel tokens, tagged with corresponding signs of IPA alphabet and plotted in F1/F2 planes – for different types and configurations of formant filters. The plots show how filtering parameters affect phonetic categorization of aural stimuli and, thus, the structure of vowel spaces. The results support the devision of vowel timbers into pure and complex (cf. one- and two-formant vowels). The maps clearly visualize the effects of masking on the shape of the borders between the perceived timbres. They also show that, within a critical band, timbres of two separate formants merge into a pure timbre of intermediate quality; when the distance between the formants increases, an interaction of their timbres takes place in the form of either *(a) mutual deduction* or *(b) overlaying* of two pure timbres, producing a timbre of a new, complex quality. Vowels synthesized with filters of constant resonance amplitude (ResonZ) showed more complex patterns of formant interaction than those synthesized with regular bandpass filters.

Introduction

Acoustic data is widely used in phonetic science nowadays, and there is no paper on the on-going language change that would not display plots of vowel formants. It is often forgotten, however, that formant frequencies do not define the timbre of a vowel alone: *(a) vowels with equal formant frequencies may sound different*, and *(b) vowels with different formant frequencies may sound the same.*

If we want to really understand how formant frequencies define vowel timbre and be able to parametrize it, we have to resort to synthetic vowels where we can keep amplitudes and qualities of 'vocal tract' resonances under control.

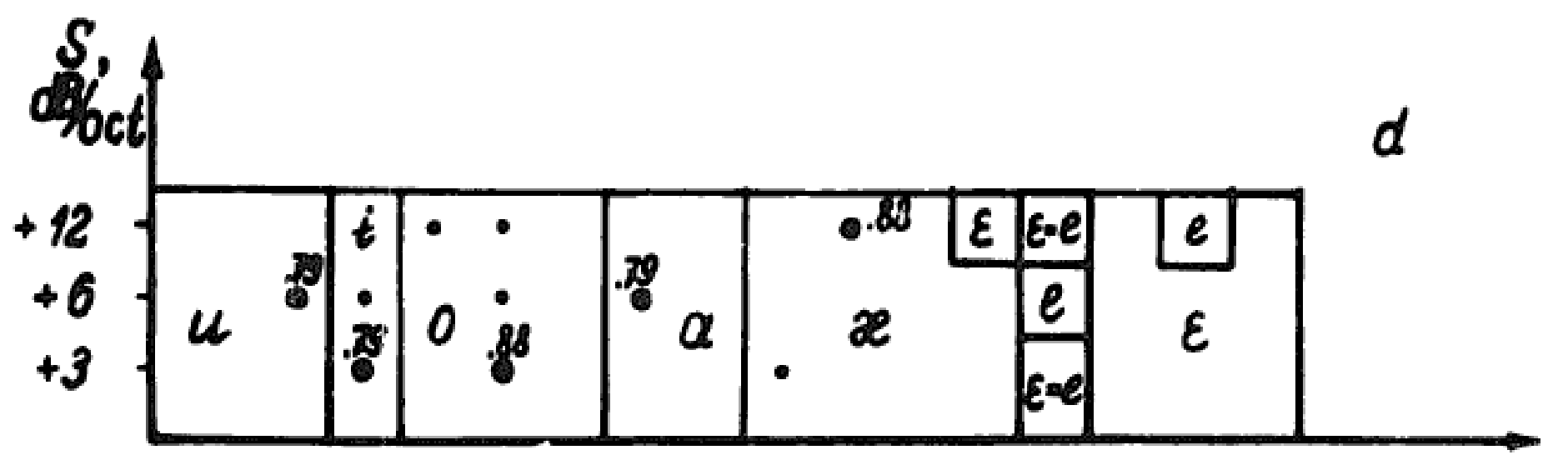


Figure 1: Perceived timber of a one-formant vowel (listener *d*) as a function of amplitude and frequency of the resonance [Chistovich, 1985]

History

Herman von Helmholz (1821–1894) was the first to *synthesize* a series of vowel timbres by spectral shaping of a harmonic series and to specify their spectral parameters, which experience he described in his "Die Lehre von den Tonempfindungen..." [Helmholz, 1963]

Sir Richard Paget (1969–1955) conducted a series of perceptual experiments with physical vowel models that he build in 1920s. He also succeeded in taking more precise measurements of formant frequencies in real vowels [Paget, 1930].

Tsutomi Chiba and Masato Kajiyama conducted a series of experiments exploring how transformation of vowel spectrum (through filtering, speeding up and slowing down recorded sounds) affects the perceived vowel timbre [Chiba and Kajiyama, 1942].

Further interesting experimental data on perception of one- and two-formant 'optically' synthesized vowels presented a group of scientists from Haskins Laboratory in 1952 [Delattre et al., 1952].

Detailed maps of the timbres of synthetic vowels (Picture 2) were published by Ralph Miller from Bell Labs in 1953 [Miller, 1953], who also described how changing the amplitude of the of resonances affects the perception of vowel timbre.

A daring aesthetic interpretation of vowel space was suggested Wayne Slawson in his later book "Vowel Color" [Slawson, 1985].

Numerous experiments on how changing the frequency and the amplitude of resonances affects perception of synthetic vowels (Picture 1) were conducted in the Pavlov Institute of Physiology during the 60s through the 80s, [e.g. Chistovich and Lublinskaya, 1979; Chistovich, 1985; Chistovich and Chernova 1986].

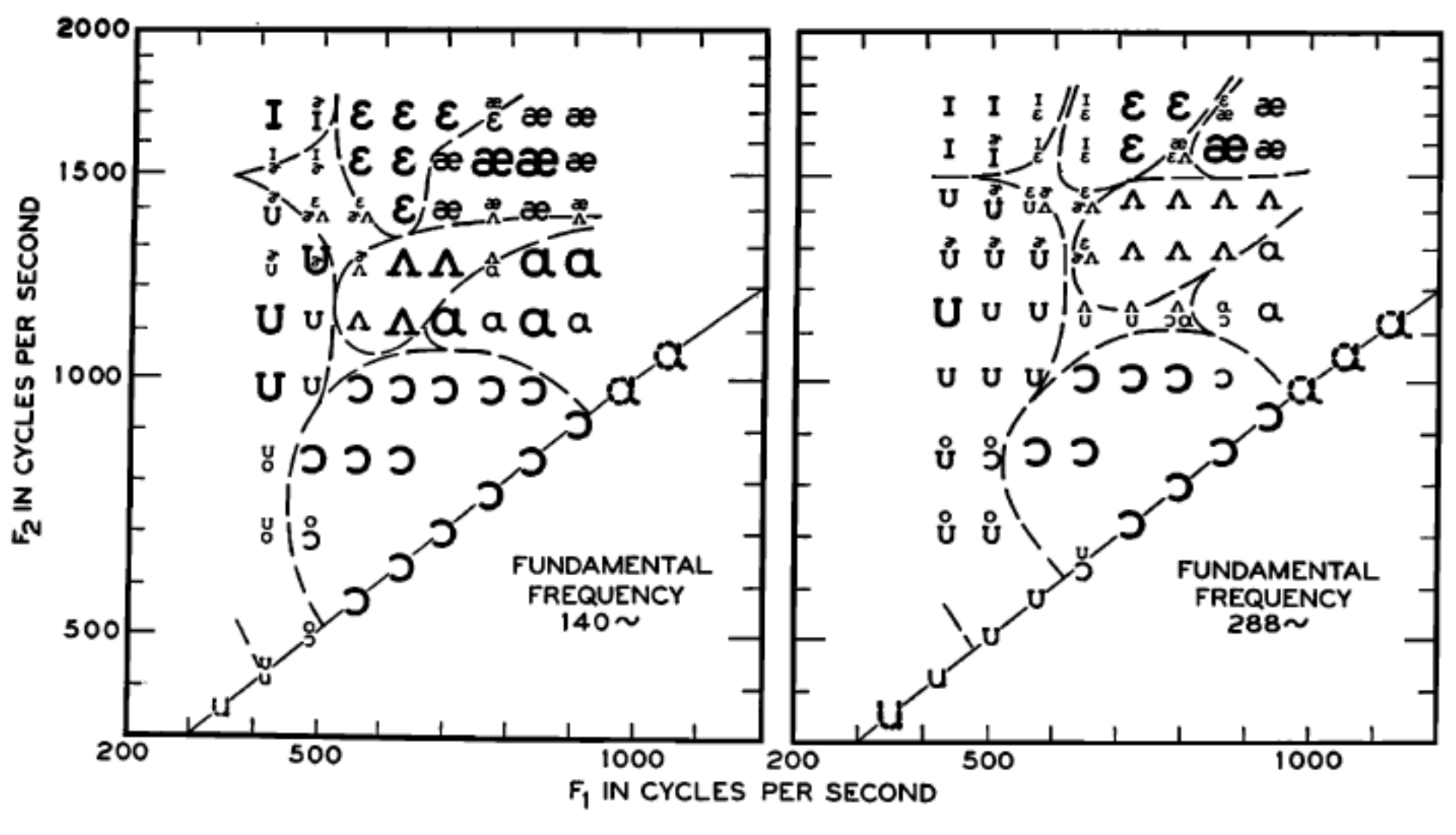


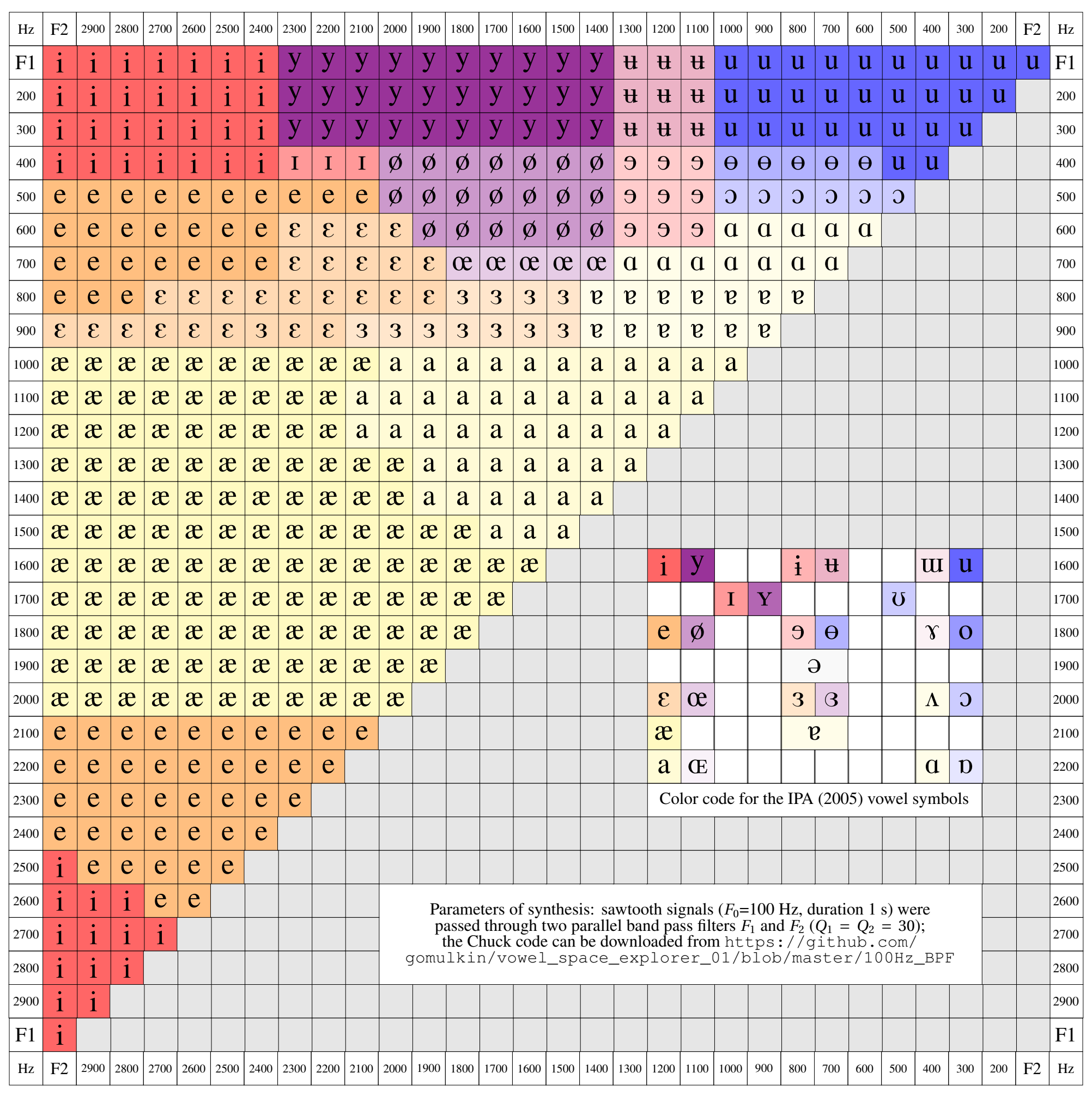
Figure 2: Vowel spaces plotted by Ralph Miller [Miller, 1953].

Method

Instrumentation

- Reference timbres for aural comparison: IPA samples [Wells and House, 2003; Whitley, 2007].
- Software: **Vowel Space Explorer 1.0** – a digital synthesizer written in Chuck programming language [Kapur et al., 2013].
- Computer interface: a 12-key numeric pad (NumPad).

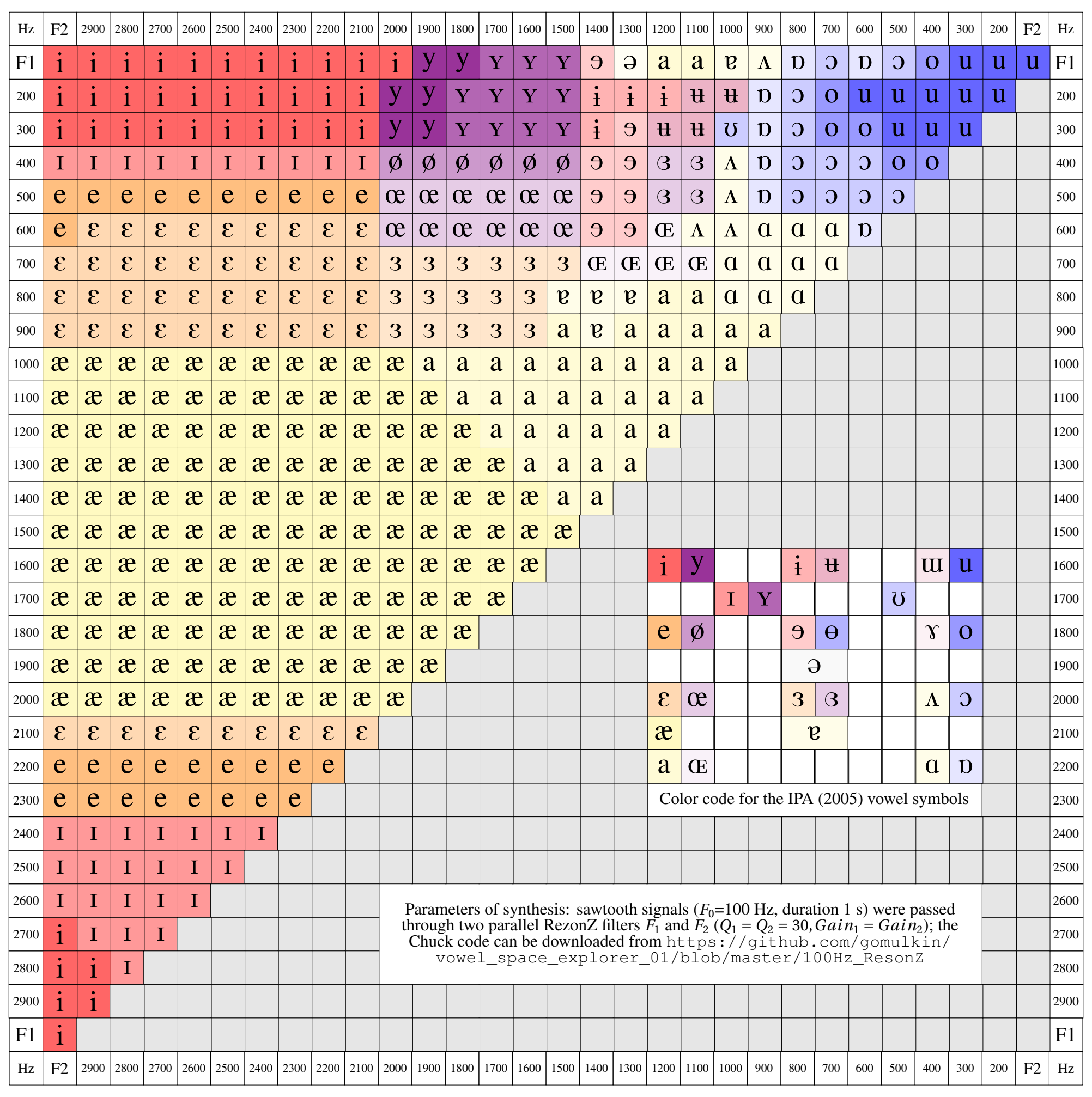
Procedure



Parameters of synthesis: sawtooth signals (F_0=100 Hz, duration 1 s) were passed through two parallel band pass filters F_1 and F_2 ($Q_1 = Q_2 = 30$); the Chuck code can be downloaded from `https://github.com/gomulkin/vowel_space_explorer_01/blob/master/100Hz_BPF`

Figure 3: The perceived timbres of 2-formant synthetic vowels synthesized with regular (not amplitude-normalized) bandpass filters. The masking effect of F1 over F2 is pronounced.

Vowel Space Explorer 1.0 – a basic vowel synthesizer – was written in Chuck programming language [Kapur et al., 2013]. Sound tokens of a constant duration and pitch were synthesized by passing a sawtooth signal through 2 parallel filters. Resonance frequencies of filters varied from 100 to 3000 Hertz in 100-Hz increments in different combinations. The timbre of each token was aurally matched with one of the 28 reference tokens [Wells and House, 2003; Whitley, 2007] and tagged with a correspondent sign of the IPA alphabet. Several maps of synthetic vowel timbres in F1/F2 plane were plotted for 2 different types/configurations of filters (400 tokens per each map).



Parameters of synthesis: sawtooth signals (F_0=100 Hz, duration 1 s) were passed through two parallel RezonZ filters F_1 and F_2 ($Q_1 = Q_2 = 30, Gain_1 = Gain_2$); the Chuck code can be downloaded from `https://github.com/gomulkin/vowel_space_explorer_01/blob/master/100Hz_ResonZ`

Figure 4: The perceived timbres of 2-formant synthetic vowels synthesized with amplitude-normalized resonant filters (ResonZ). Notice a more intricate pattern of F1/F2 interaction.

Future Plans

Chuck uses classical pre-compiled 'unit generators' from STK library [Cook, 1986], but allows for integration of your own algorithms. Writing your own filters 'from scratch' would provide more freedom in accessing and controlling filter parameters – and better understanding of the underlying processes.

Building a new HID (human interface device) to Chuck programming environment felt like an urgent issue. The newer physical – Arduino-based and C-programmed – interface is being tested.

Conclusions

The plotted maps visualize how filtering parameters influence phonetic categorization of aural stimuli and, thus, the structure of vowel spaces. The results support the devision of vowel timbers into pure and complex, which roughly corresponds to the older devision of vowels into one- and two-formant ones [Paget, 1930; Chiba and Kajiyama, 1942; et al.]. The plots visualize the effects of masking on the shape of the borders between the 'timbre zones' [Chistovich and Chernova, 1986].

It was also shown that within a critical band, timbres of two formants merge into a pure timbre of intermediate quality [Chistovich and Lublinskaya, 1979]; when the distance between the formants increases, an interaction of two timbres (rather than their averaging) takes place in the form of either *(a) mutual deduction* (timbres of back unrounded vowels), or *(b) overlaying* (timbres of front rounded vowels) of the two pure timbres, producing a timbre of complex quality [cf. Schane, 1996].

Vowels synthesized with filters of constant resonance amplitude, ResonZ (Smith, 2007), show more intricate patterns of formant interaction than produced with regular bandpass filters.

Discussion

A one-formant vowel synthesized with a single filter, function of changing parameters of the filter's amplitude and bandwidth, may produce complex timbres characteristic of two-formant vowels: this phenomena was reported in [Chistovich, 1985]. To my ear, the increase of resonance amplitude may change [u] and [o] into [y] and [œ] accordingly, sharpening of the resonances may change the former into [ɜ] and [i] accordingly, which finds some parallels in the data presented here.

Another interesting aspect neglected in the present study is the influence of the spectral shape of the voice source on the timbre of the final vowel, which adds another variable to the set of parameters that eventually define the timbre of a vowel.

References

[1] Chiba, T., and Kajiama, M. *The vowel, its nature and structure*. Tokyo-Kaiseikan, 1941.

[2] Chistovich, I. A., and Chernova, E. I. Identification of one- and two-formant steady state vowels: a model and experiment. *Speech Communication 5* (1986), 3–16.

[3] Chistovich, L. A. Central auditory processing of peripheral vowel spectra. *The Journal of the Acoustical Society of America 77* (1985), 789–805.

[4] Chistovich, L. A., and Lublinskaja, V. V. The 'center of gravity' effect in vowel spectra and critical distance between the formants: psycho-acoustical study of the perception of vowel-like stimuli. *Hearing Research 1* (1979), 185–195.

[5] Cook, P. R. *Real Sound Synthesis for Interactive Applications*. A. K. Peters, 2002.

[6] Delattre, P., Liberman, A. M., Cooper, F. S., and Gerstman, L. J. An experimental study of the acoustic determinants of vowel color; observations on one- and two-formant vowels synthesized from spectrographic patterns. *Word 8*, 3 (1952), 195–210.

[7] Helmholtz, H. L. F. v. *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*. Vieweg, 1863.

[8] Kapur, A., Cook, P., Salazar, S., and Wang, G. *Programming for musicians and digital artists: creating music with Chuck*. Manning Publications, 2013.

[9] Miller, R. L. Auditory tests with synthetic vowels. *The Journal of the Acoustical Society of America 25*, 1 (1953), 114–121.

[10] Paget, R. *Human speech; some observations, experiments, and conclusions as to the nature, origin, purpose and possible improvement of human speech*. Routledge & Kegan Paul, 1930.

[11] Schane, S. A. Diphthongization in particle phonology. In *The Handbook of Phonological Theory* (1996), Blackwell Publishing.

[12] Slawson, W. *Color Space*. University of California Press, 1985.

[13] Smith, J. O. *Introduction to digital filters: with audio applications*. W3K Publishing, 2007.

[14] Wells, J., and House, J. *The sounds of the International Phonetic Alphabet*. Phonetics and Linguistics UCL, 2003.

[15] Whitley, S. Vowels. In *SIL IPA Help 2.1* (2007), SIL International.