Acoustic description of vowels – An introduction

1. Acoustic studies in phonetics

Modern phonetics – i.e. the branch of science that investigates the sounds of speech – is largely based on acoustic studies. There are a number of reasons for this. Firstly, we know that most information conveyed during linguistic interaction is contained in the sound wave that is generated by the speaker's speech organs and then perceived by the listener's ear. (It is true that in normal face-to-face conversation some information is conveyed non-verbally, e.g. through gestures or facial expressions, but it is worth recalling that interaction based exclusively on sound – e.g. on the telephone – is possible and only slightly less efficient.) Secondly, modern methods of sound recording and analysis are "convenient", since the technologies used are mature and relatively easily accessible. These days, it is relatively easy to record sound with a sufficient quality, and the analyses can be done using free computer software (such as Praat). Thirdly, the results obtained in this way are more objective and repeatable than „impressionistic" analyses done „by ear". This is so since they are based on measurements of physical properties. Finally, the results of this kind of reseach are increasingly useful in practice, going beyond a mere description of the phenomena (e.g. in automatic speech recognition, speech synthesis, in the design of hearing aids or systems that improve sound quality in telecommunications, etc.).

2. The acoustics of vowels

Three main aspects of the sound wave are usually studied within phonetics: intensity (loudness), duration (length) and spectral or frequency-related properties (quality/timbre of sound). For vowels, spectral properties are the most important aspect; in some situations, duration is also studied.

This is so because contrasts between vowels are mainly based on their quality; the general manner of articulation is similar in all vowels. In this respect, vowels are significantly different from consonants, since the latter can be differentiated by manner of articulation and voicing. For example, the sound [p] is a voiceless plosive: one of its elements is so-called closure, i.e. the brief moment when the articulators (the upper and lower lip in this case) come into contact. From an acoustic point of view, this is simply a brief moment of silence. In turn, [s] is a voiceless fricative where there is no full closure, and noise is the main acoustic component.

Different vowels, in turn, share general features. First of all, they are usually voiced, i.e. there is vibration of the vocal folds which are found in the larynx. Usually (in normal speech), the frequency of this vibration is in the range from about 80 Hz to about 400 Hz (or from 80 to 400 cycles per second). The resulting sound wave contains this "fundamental frequency" as well as harmonic frequency at integer multiples of the fundamental frequency. Normally, you would hear this signal as a buzz of sorts that would not resemble speech. However, the buzz is "filtered" by the subsequent (upper) parts of the vocal tract: the pharynx (throat) and the oral and nasal cavities. Here, some frequencies are dampened; those frequencies that are not dampened, are called "formants" (resonant frequencies). Thus modified, the sound wave starts sounding like a vowel.

We know that the position of the tongue in the oral cavity is the most important factor responsible for this "filtering" (even though there are also other factors – in particular, whether part of the air is expelled through the nose). We also know that correct recognition of a vowel is possible using just the first two formants. Because of that, measurements of the first two formants are at the basis of the acoustic description of vowels. For this reason, measurements of this kind were performed for four languages from the database of the present project: Hałcnovian, Latgalian, Wilamowicean and Yiddish.

Fig. 1 shows the vowel [i] in the analysis window of Praat. The top panel shows the waveform, or a graphical representation of the amplitude of the soundwave (on the Y axis) against time (in the X axis). This is a rather common way of visualizing a soundwave and it must be familiar to anyone who has edited sounds on a computer using software such as Audacity, GoldWave or CoolEdit.

The bottom panel contains a spectrogram, or a representation of the component frequencies of a sound signal against time. Here, too, the X axis shows time; but the Y axis represents frequency. Darker areas have more energy than lighter areas. (The vertical stripes correspond to individual vibrations of the vocal folds.)

It can be seen that there are a number of darker horizontal bands throughout the duration of the entire vowel; these are the formants. The first formant (counting from the bottom; F1) is at about 260 Hz; the second (F2) – about 2200 Hz.

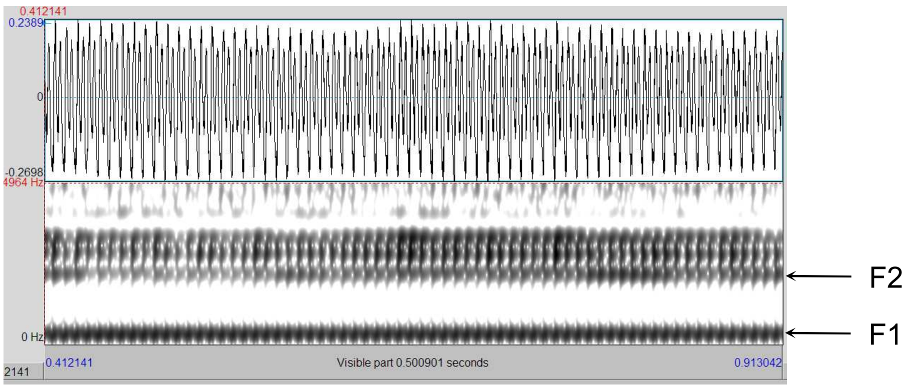

Fig. 1. The formants of an [i].

Fig. 2 shows the vowel [a]. It can be seen that the formants are located at different frequencies than in [i]: F1 = 750 Hz, and F2 = 1200 Hz. The remaining formants can also differ but they are far less important linguistically. Since the formants are quite close to each other, and they can be difficult to tell apart, Praat overlays the red dots which show estimated formants values calculated automatically.

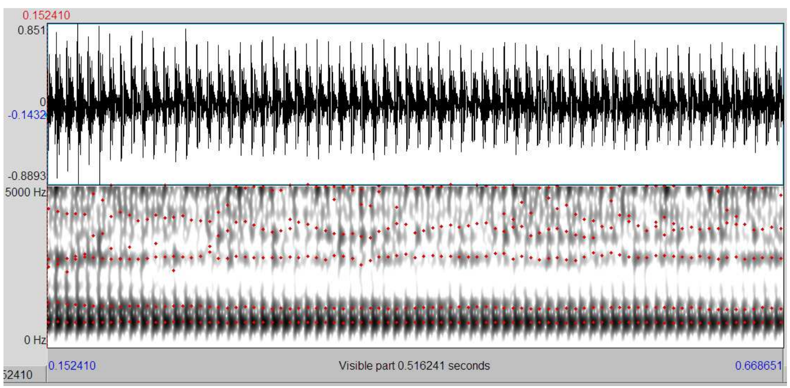

Fig. 2. The formants of an [a].

3. Acoustic vowel charts and the vowel quadrilateral

Results of formant measurements can be shown using a two-dimensional chart similar to that in Fig. 3. It has one rather surprising characteristic: the origin of the coordinate system is at the top right, with the first formant (F1) shown on the Y axis, and the second formant (F2) – on the X aaxis.

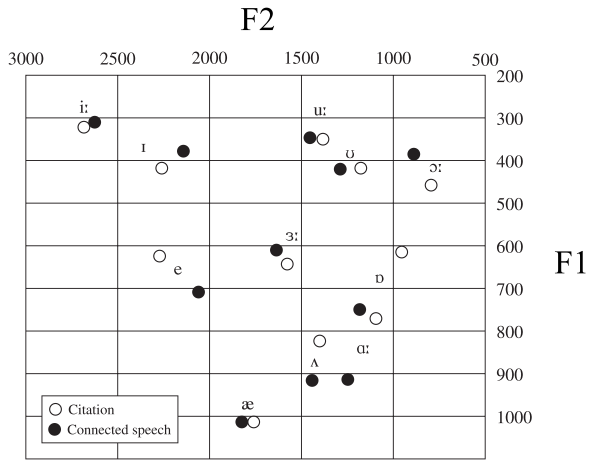

Fig. 3. A chart showing the relationship between F1 and F2 for Standard British English (Cruttenden 2008: 103).

This orientation is used because it reflects the orientation of the traditional "vowel quadrilateral" used in impressionistic descriptions. Vowels are placed on a quadrilateral of this kind on the basis of auditory impressions, although it is sometimes claimed that the chart show the higest point on the surface of the tongue during the articulation of a given vowel. The latter interpretation is the source of the terminology used to describe vowels. Thus, vowels producd with a wide open mouth, e.g. [a], are called „open" or „low", while those in which there is less mouth opening, e.g. [i] or [u], are called "close" or "high". Vowels articulated at the front of the mouth, e.g. [i] or [e], are called "front", while those made at the back (e.g. [u] or [o]) are called "back". Fig. 4 shows a traditional vowel chart for English.

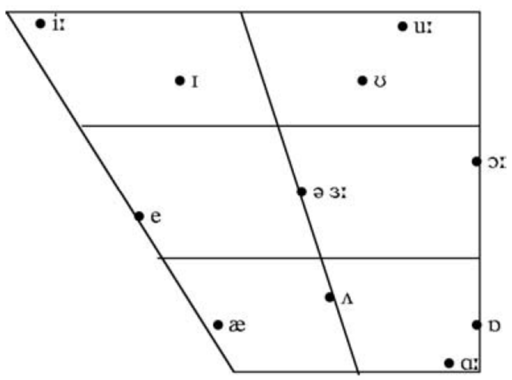

Fig. 4. A vowel chart (quadrilateral) for Standard British English (Roach 2004: 242).

The relationship between the locations of vowels on a traditional vowel quadrilateral and their locations on a formant chart was noticed already in the first acoustic studies of vowels. Today, we know that F1 is proportional to the degree of mouth opening in a vowel, while F2 is proportional to a vowel's "frontness". Even though in reality this relationship is not as simple, and there are complications, it is often said informally when discussing acoustic data that a vowel is open or close etc. (When you compare the quadrilateral in Fig. 4 with the chart in Fig. 3, you will see that not all the relationships between the vowels are graphically similar, but the general positions of the vowels are similar.)

The measurements made for Hałcnovian, Latgalian, Wilamowicean and Yiddish within the present project are illustrated using charts similar to that in Fig. 3. The charts provide an image of the actual acoustic characteristics of the sounds studied. Quadrilaterals similar to the one Fig. 4 are used to illustrate vowel systems on the basis of existing descriptions or proposed transcriptions. These were made by reference to the qualities of standard reference vowels in the so-called „cardinal vowel" system of the International Phonetic Association (IPA), unless stated otherwise.

4. Variability

Representations similar to those in Fig. 3 and 4 might suggest that there is no variability in the production of vowels: each vowel can be represented as one point on a plane. This could not be further from the truth. Due to the "analogue" character of the articulators (in particular the tongue), each time you produce a sound of speech, its physical properties will be different. Vowels have much more variability than consonants. (The reason is that during the production of a vowel the tongue is "suspended" in the oral cavity without much contact with the roof of the mouth or other articulators.) There are also differences between women and men (e.g. the chart in Fig. 3 shows measurements from women). There are also differences between individual speakers which can be both systematic and random.

For this reason, charts of the kind shown in Fig. 3 represent mean values – for one speaker or a group of speakers. This is illustrated in Fig. 5. On the left-hand side, individual measurements are shown of the first two formants of Polish vowels in one female speaker; the right-hand panel shows the mean values.

In the materials from the present project, both chart types are presented.

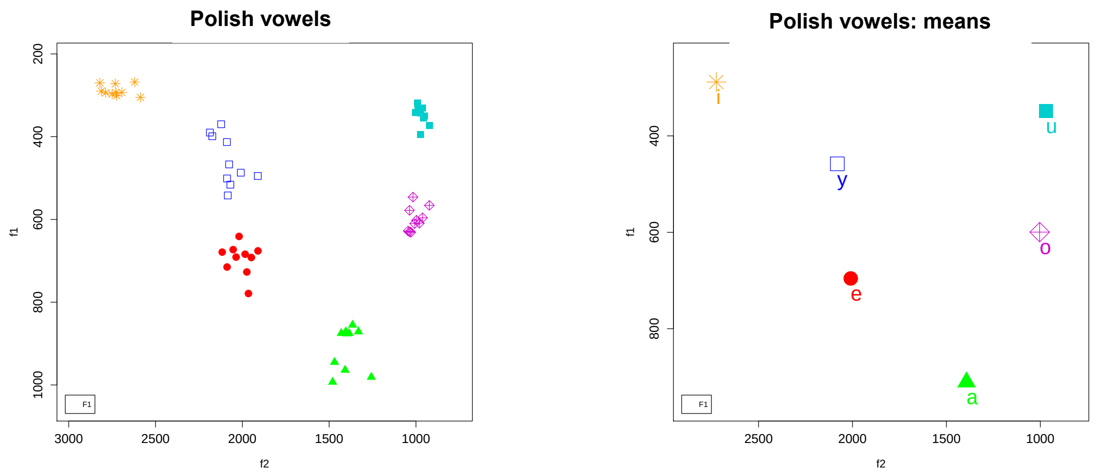

Fig. 5. Polish vowels in one female speaker. Left: individual measurements from 60 vowels. Right: mean values (Weckwerth 2013).

5. Sources

Cruttenden, Alan (2008). *Gimson's pronunciation of English*. London: Hodder Arnold.

Roach, Peter (2004). "British English: Received Pronunciation". *Journal of the International Phonetic Association* 34(2): 239–245.

Weckwerth, Jarosław (2013). "Formant values of English and Polish vowels in highly proficient Polish learners of English". Poster presented at the *New Sounds 2013* International Symposium on the Acquisition of Second Language Speech, Concordia University, Montreal.

Prepared by:
Jarosław Weckwerth
Faculty of English
Adam Mickiewicz University in Poznań
wjarek@wa.amu.edu.pl