



Lecture 2-3: Acoustics of Vowels

Overview

1. **Describing vowels in the speech chain.** From speaker to hearer we can identify the following stages in the transmission of a vowel: a **phonological** stage where the lexicon specifies a particular vowel to take the role in the nucleus of a syllable; a **phonetic** stage where the articulators take on a particular configuration; an **acoustic** stage where the frequency response characteristics of the vocal tract affect the spectral shape of the sound produced; a **perceptual** stage where the timbre or quality of the vowel is analysed by the hearing mechanism; and a second **phonological** stage where the listener uses his/her knowledge of lexical contrasts to identify which vowel was selected by the speaker.
2. **Phonology → Phonetics.** The phonological specification for a vowel identifies a set of **target** values for the realisation of the vowel; some theories suggest these targets are **articulatory**, others that they are **acoustic**. A motor control plan is then designed and executed in an attempt to realise these targets. However the actual articulation may not perfectly meet the targets: possibly because of other overlapping and contradictory targets of adjacent segments, or for lack of time. Vowel articulation can thus be affected by the **consonantal context** in which they are produced and/or by **undershoot**.
3. **Phonetics → Acoustic-Phonetics.** At any given time in the production of a vowel, the spectrum of the sound radiated from the lips can be found from the product of the excitation (source) spectrum generated by the larynx and the frequency response (filter) of the vocal tract configuration. Typically the spectrum of a vowel is made up of a large number of harmonics, with those harmonics occurring at frequencies close to the resonant frequencies of the tract (formants) having the greatest amplitude. The location of formant frequencies can be reasonably well predicted from the vocal tract shape by means of a **two-tube approximation**, see figure 2-3.3. This demonstrates the observed correlation between F1 and size of cavity behind the tongue hump, and the correlation between F2 and the size of the cavity in front of the tongue hump. The large changes in F1 and F2 across vowel articulations explain why these two features of vowels have the greatest perceptual effect, see figure 2-3.1.
4. **Acoustic-Phonetics → Perception.** The hearing mechanism delivers to the brain of the listener a neural pattern related to the **spectral envelope** of the vowel spectrum. This spectral envelope varies according to both the pitch and the timbre of the sound. Since the absolute values of the formant frequencies depend on the size of the speaker's vocal tract, the listener must first compensate for vocal tract size before attempting to identify the vowel, see figure 2-3.2. This process is called **normalisation**, and can be thought of as a transformation of the actual vowel quality to the 'equivalent' vowel quality of a 'standard' speaker. The normalised quality can then be mapped to a position on a perceptual map (a lot like a vowel quadrilateral) and its phonetic quality retrieved with respect to known qualities.
5. **Perception → Phonology.** Given a normalised phonetic vowel quality, the choice of phonological category can still be influenced by other factors. In particular, the accent of the speaker can affect which category a quality is assigned to (e.g. [beks] -> "becks", "backs"). Listeners also compensate for undershoot caused by the consonantal context and speaking rate.

Reading

At least one from:

- Kent & Read, *The Acoustic Analysis of Speech* (1st edition), Chapter 5: The acoustic characteristics of vowels and diphthongs. *Comprehensive*.
- Hewlett & Beck, *An introduction to the science of phonetics*, Chapter 17: The vocal tract as a resonator, pp239-255. *Mathematical treatment of tube resonances*.

Learning Activities

You can help yourself understand and remember this week's teaching by doing the following activities before next week:

1. Write a one page essay on the concept of "vowel" at different points in the speech chain. Be sure to contrast a vowel phoneme, a vowel articulation, the acoustic signal of a vowel sound and the perceived timbre of a vowel.
2. Search on the web for X-ray and MRI images of speakers making vowel sounds. Read about the experimental set up of the recording. Make a list of some problems in obtaining realistic articulatory measurements of naturally-spoken vowels.
3. Make a list of as many different sources of variability in the acoustic form of (phonological) vowels as you can think of. For each of these think about how listeners might cope with the variation to still be able to identify the vowel.

If you are unsure about any of these, make sure you ask questions in the lab or in tutorial.

Reflections

You can improve your learning by reflecting on your understanding. Here are some suggestions for questions related to this week's teaching.

1. What is the difference between /e/ and [e] ?
2. What are the most extreme examples of accent variation in vowel quality that you can think of?
3. Give some examples of vowels changing in quality because of the context in which they occur.
4. What is meant by "target undershoot" in vowel articulation? What would be the difference between "articulatory" targets and "acoustic" targets?
5. What difficulties does a child face when learning to copy adult vowel sounds?

Figure 2-3.1 Table of formant frequencies

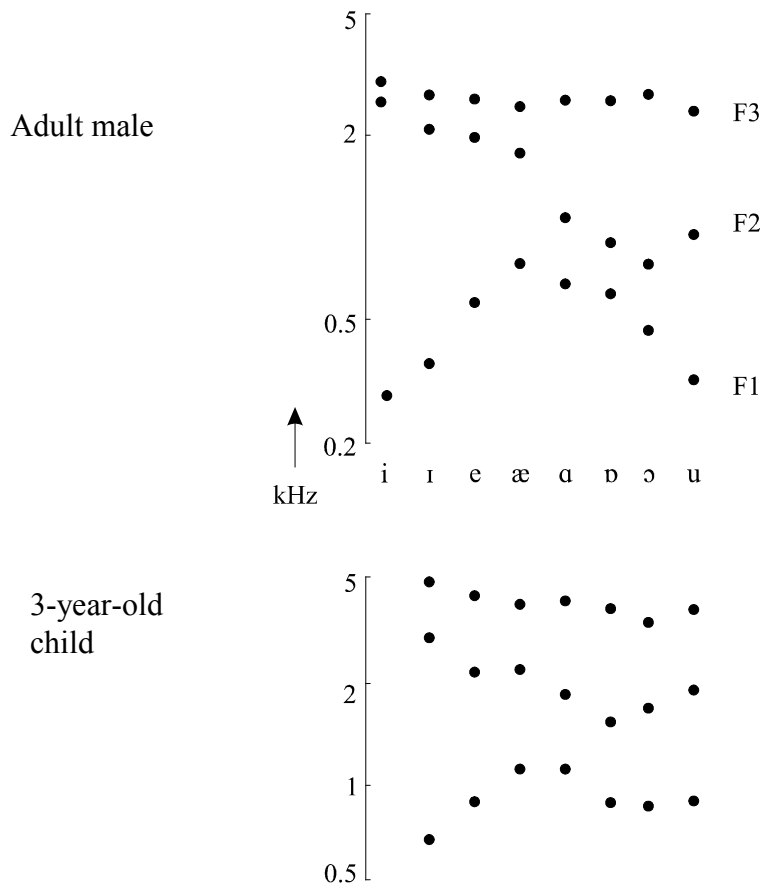
Vowel	F1(Hz)	F2(Hz)	F3(Hz)
i:	280	2620	3380
ɪ	360	2220	2960
e	600	2060	2840
æ	800	1760	2500
ʌ	760	1320	2500
ɑ:	740	1180	2640
ɒ	560	920	2560
ɔ:	480	760	2620
ʊ	380	940	2300
u:	320	920	2200
ɜ:	560	1480	2520

Adult male formant frequencies in Hertz collected by J.C. Wells around 1960.

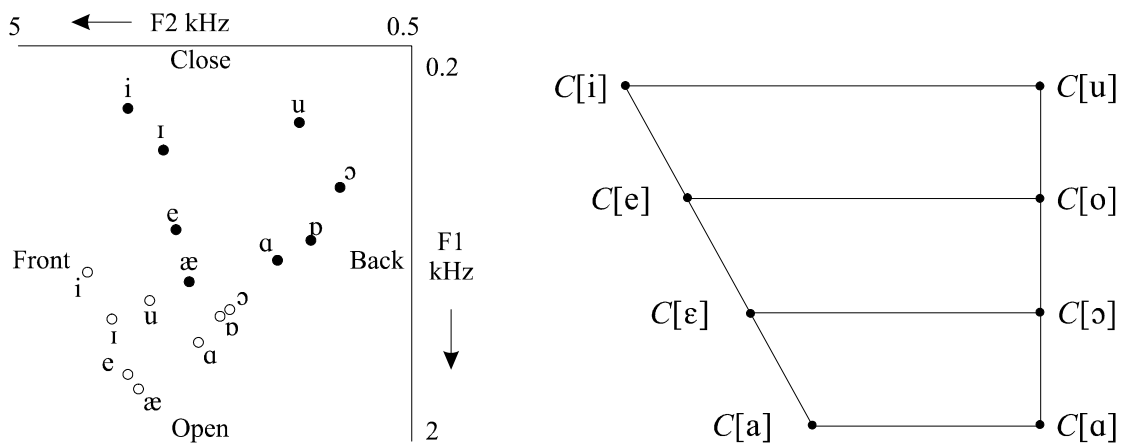
Note how F1 and F2 vary more than F3.

Figure 2-3.2 Formant Frequency Patterning in British English Vowels

Raw formant frequency data from adult males and 3-year-old children

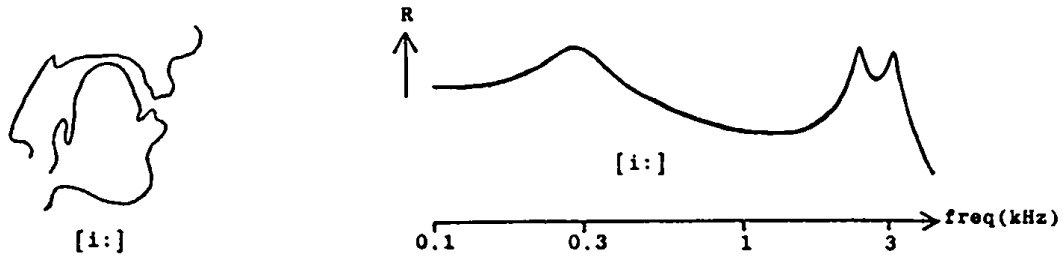


Same data plotted on F1-F2 plane looks like vowel quadrilateral

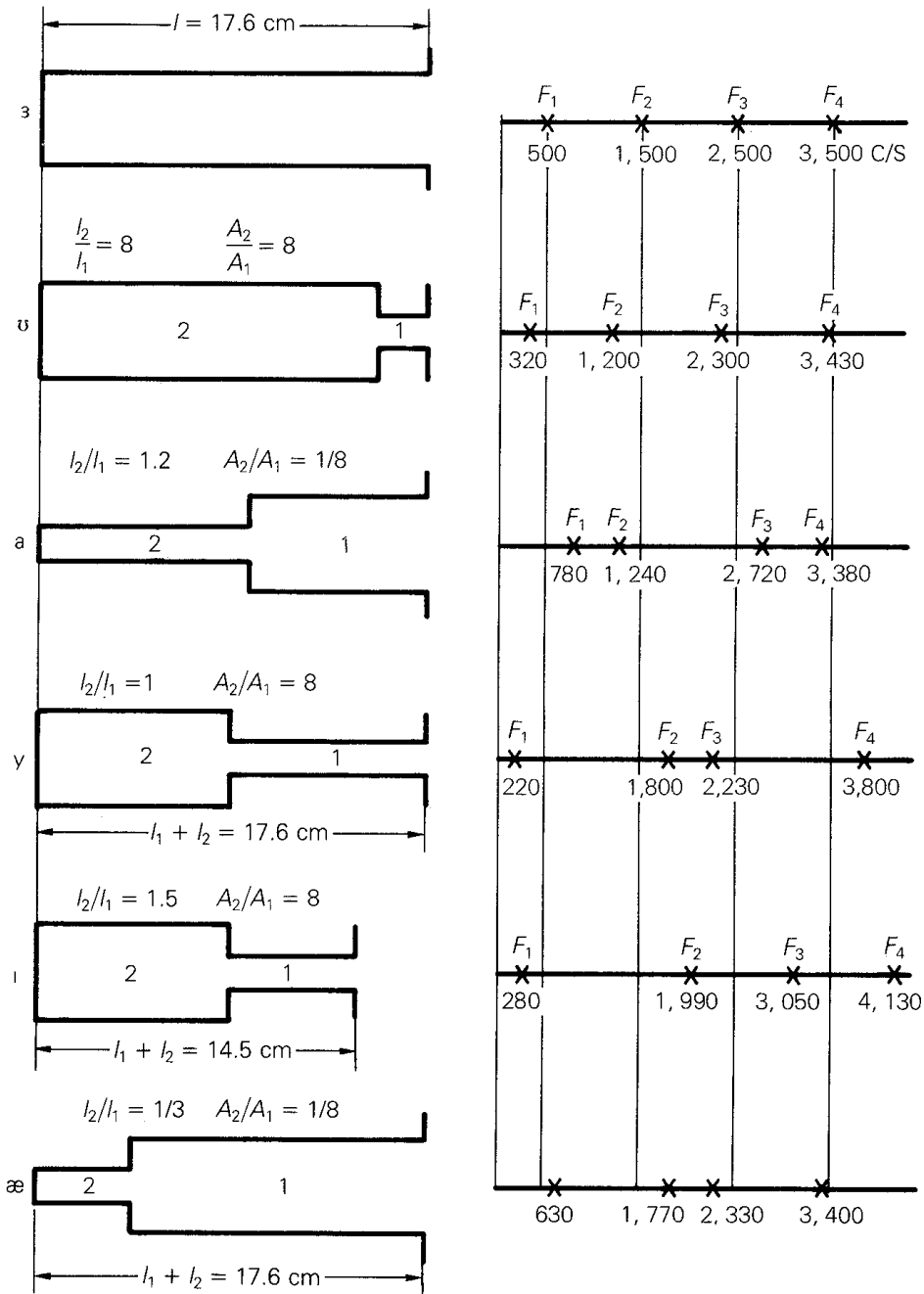


Notice that patterning is similar for adult (closed circles) and child (open circles), but absolute frequencies are not.

Figure 2-3.3 Two Tube Resonators Approximating Vowel Articulations



What is the connection between the X-ray cross-section and the resulting frequency response?



Two-tube resonators have resonant frequencies that are very similar to vowel formant frequencies (Clark & Yallop, after Fant 1960).

Lab 2-3: Analysis of Natural Vowels

Introduction

Vowel formant frequencies are known to vary for a number of causes operating at different levels in the speech chain, we can summarise these as: (i) because of *phonological* differences, e.g. different lexical pronunciations across accents, (ii) because of *phonetic* differences, e.g. choice of articulation, or speed of articulation, and (iii) because of *acoustic* differences, e.g. different speakers have different-sized vocal tracts.

We can measure formant frequencies fairly accurately by investigating the *spectral envelope* of the averaged short-time spectrum of the signal in the centre of a vocalic region. In this experiment you will measure the formant frequencies of a number of vowels embedded in a sentence of your own speech, and compare the frequencies to your classmates and to some measurements made from children's speech.

Scientific Objectives

- to investigate how formant frequencies vary for different vowels.
- to investigate how much formant frequencies vary for the same vowel across speakers.

Learning Objectives

- to gain more experience of how speech signals look on a spectrogram.
- to practise making measurements from live speech, and to appreciate the difficulties that can arise.

Method

Your recording of the utterance 'Gene had thirty-four smart boots' has been acquired on to the computer system and a wide-band spectrogram has been printed for you.

You can use the SFSWin program to find and display your recording. You can use the facility to zoom and to replay selected portions of the signal to relate parts of the signal to parts of the utterance. You can also measure formant frequencies from a smoothed spectral cross-section display. (<http://www.phon.ucl.ac.uk/resource/sfs/>)

Observations

1. Annotate your spectrogram with a *broad phonetic* transcription.
2. Using the smoothed cross-section display, measure the first three formant frequencies for the vowels, writing your results in a table like the one overleaf.
3. Plot your vowel formant frequencies and those of your group on an F1-F2 graph. Plot F1 going top to bottom on the *right* side of the paper, and F2 going right to left across the *top* of the paper. Allow a range of 200-1500Hz for F1, and 500-4000Hz for F2.
4. Plot your vowels on the class F1-F2 diagram.
5. See if you can see any consistency within your group for F3 values. What can you infer from this about the importance of F3 for vowel perception?
6. You are given a child's production of a similar sentence. Make measurements of F1 and F2 for the child's vowels and compare these with the adult values. Plot these in a different colour on your F1-F2 graph. Interpret what you see.
7. If you have time, record and measure spectrograms of some simple diphthongs such as
 - [aɪ] in 'buy'
 - [oɪ] in 'boy'
 - [əʊ] in '(rain)bow'

See if the formant frequencies match up to those of the monophthongal vowels you measured in 1. You can draw these as lines on your F1-F2 graph.

Vowel	F1 (Hz)	F2 (Hz)	F3 (Hz)
[i:] in 'Gene'			
[æ] in 'had'			
[ɜ:] in 'thir'			
[ɔ:] in 'four'			
[ɑ:] in 'smart'			
[u:] in 'boots'			

Concluding Remarks

The class F1-F2 diagram will probably show a large amount of variation in the measured formant frequencies for each vowel. Suggest a number of reasons why the values vary and assign each cause to a *phonological*, *phonetic* or *acoustic* level of description.

How are listeners able to identify a vowel given such speaker-to-speaker variation?

Examination Questions

These are questions from past exam papers. You may like to write outline answers to these, or to discuss them in tutorial.

1. Describe **three** causes of variability in the acoustic realisation of phonological vowels. Show with examples how each affects the measured formant frequencies. [2004/5]
2. Explain with examples how regional accent, prosodic context, consonantal context and vocal tract size can affect the measured formant frequencies within a vowel category. [2005/6]
3. Why are formant frequencies useful for measuring the acoustic quality of vowels? Explain, using examples how separate changes in (i) accent, (ii) prosody and (iii) speaker can lead to changes in the measured formant frequencies of a (phonological) vowel. [2008/9]