INTRODUCTION TO ACOUSTIC PHONETICS 2
Hilary Term, week 6
22 February 2006

1. Resonators and Filters

Different vibrating objects are tuned to specific frequencies; these frequencies at which a particular object prefers to vibrate are called *natural frequencies*.

If you couple a vibrating object (a driving force) with another object (a driven system), it will cause *forced vibration* in the latter. **Resonance** happens when the frequency of the driving force is close to the natural frequency of the driven system. It has two-fold effect:

1) amplitude of vibration are increased at the natural (resonant) frequencies of the driven system, i.e. the latter works as a **resonator**;
2) amplitudes at other frequencies are absorbed, decreased, i.e. the driven system works as a **filter**.

Filtering of a complex sound is the process of selective separation; some frequencies are allowed to pass through the filter while other frequencies are blocked from passing.

There are different kinds of filters. A **low pass filter** permits only frequency components *below* a specified frequency (cut-off frequency) to pass unattenuated and reduces or blocks completely frequency components above the cut-off frequency. A **high pass filter** permits only frequency components above a cut-off frequency to pass unattenuated. A **band pass filter** is a combination of LP and HP filters: it permits frequency components between two cut-off frequencies to pass unattenuated.
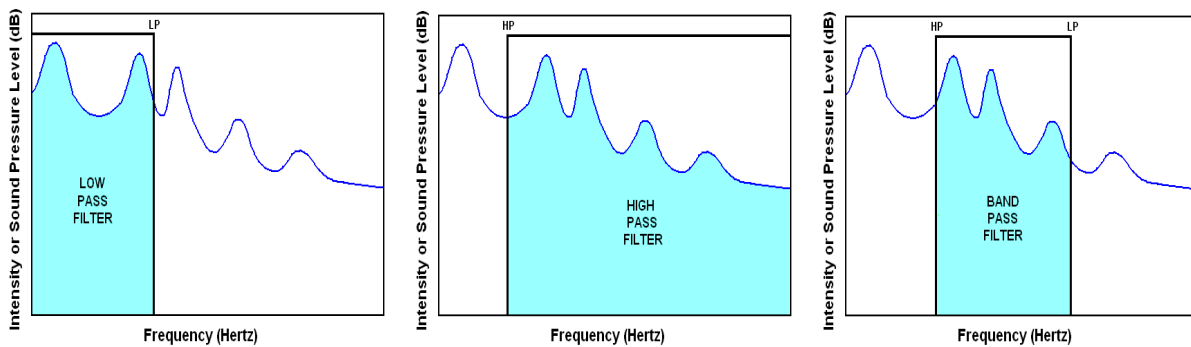

Fig. 1. Three types of filters: low pass, high pass, and band pass

Band pass filters are characterized by **centre frequency** and **bandwidth**. The latter is defined as a range of frequencies passed by the filter, which are not more than 3 dB down from the maximum amplitude (i.e. the amplitude of the centre frequency). The bandwidth of a filter may be relatively narrow or broad (see Fig. 2 and 3).
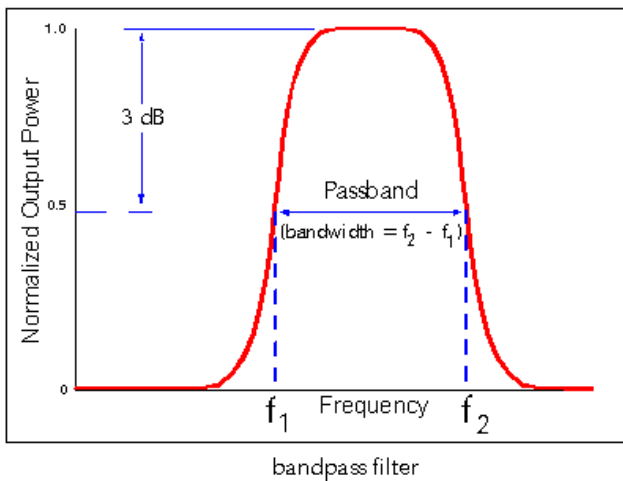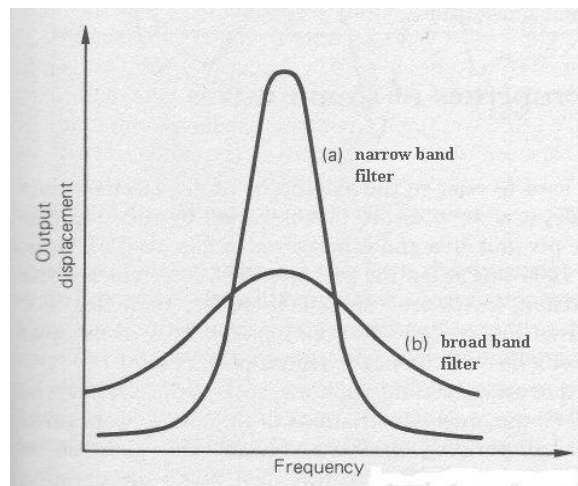
Fig. 2. Band pass filter and its properties

Fig. 3. Filters with narrow and broad bandwidths

2. Spectrograms

Series of band pass filters are used to create a visual representation of sounds called **spectrograms**. Unlike 2-D power spectra, spectrograms also represent the dimension of time, which makes it possible to capture constant variation of the acoustic signal. Spectrograms display time on the horizontal axis, frequency on the vertical axis, while amplitude is represented by means of colour density:

Depending on the use of broadband or narrowband filters, we can get either **wide-band** (typically 300 Hz bandwidth) or **narrow-band** (45 Hz bandwidth) spectrograms. There is a trade-off between temporal and frequency resolutions. In a wide-band spectrogram (Fig. 4) individual harmonics are smashed, but it has a high temporal resolution – it is possible to see individual voicing pulses as vertical striations. They are most commonly used because in majority of cases we are not interested in the changes of harmonics.
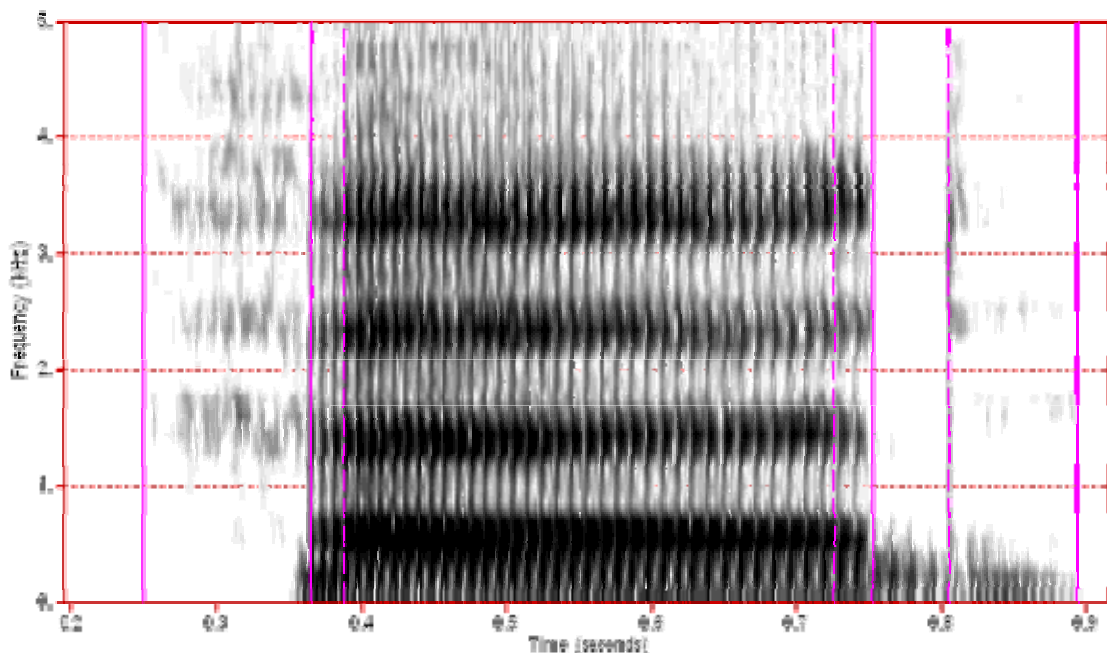

Fig. 4. Wide-band spectrogram of the word 'heard'; vertical striations show individual voicing pulses

Narrow-band spectrograms (Fig. 5) pick individual harmonic that appear as black horizontal lines, and the drops of energy between successive harmonics that appear as white horizontal lines, but temporal resolution is poor.
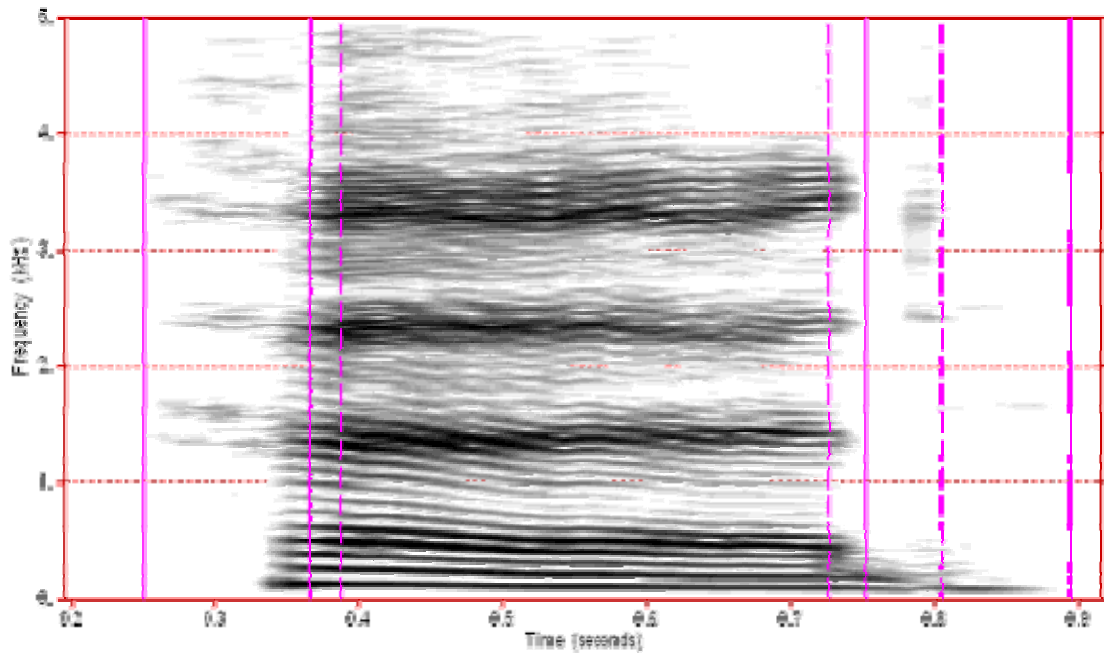
Fig. 5. Narrow-band spectrogram of 'heard', horizontal lines are individual harmonics

3. Source-Filter Theory

The acoustic theory of speech production, known as source-filter theory, postulates that sound production consists of two basic components: (1) generation of sound **source** at the glottis or at some point along the length of vocal tract, and (2) **filtering** of that source by the vocal tract.

There are three types of sound sources involved in speech production:

a)   quasi-periodic laryngeal voicing source, produced by the vocal fold vibration (present in vowels, nasals and approximants),
b)   continuous aperiodic turbulent source, produced at some point along the vocal tract (majority of voiceless fricatives),
c)   transient aperiodic noise source, produced at some point along the vocal tract (the release burst of voiceless stop consonants).

These sources may be combined, e.g. production of voiced fricatives involves both periodic laryngeal source and aperiodic noise generated in the vocal tract. Production of voiced stops combines all three sources.

Alone or in combination, these sources serve as input for the vocal tract filter which modulates this input, as different frequency components of the source are passed through the filter.

4. Source-Filter Theory for Vowels

Vowels are the product of glottal quasi-periodic source and filtering effects of the supraglottal tract. Same quality vowels have the same gross spectral shapes, irrespective of the fundamental frequency of the source (a variable that changes significantly depending on the age, sex, and emotional state of the speaker).

4.1. Voice glottal source

The air flowing out of the lungs supplies the system with the energy needed to produce sounds. When vocal folds vibrate, the rate of air flow through the glottis rises and falls and it generates a complex periodic wave.

Fourier analysis on the glottal source waveform gives us the power spectrum showing its component frequencies (see Fig. 6). Amplitude of nearly all harmonics after the second one decrease rapidly as frequency increases – it is an important characteristic of the glottal source spectrum.
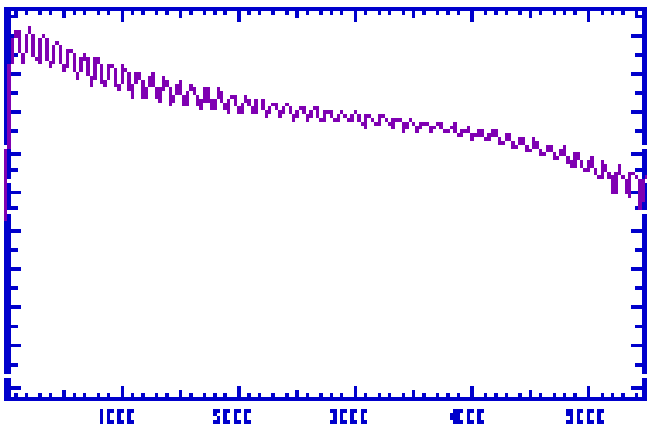

Fig. 6.Spectrum of the glottal source waveform

Glottal source waveform and spectrum vary depending on the type of phonation – modal, creaky or breathy. The differences in the waveform are due to the differences in the amount of time that the vocal folds are open during each glottal cycle. The relationship between amplitudes of the first two harmonics and general slope of the spectrum are two main spectral indicators of the type of phonation (Fig. 7-8).
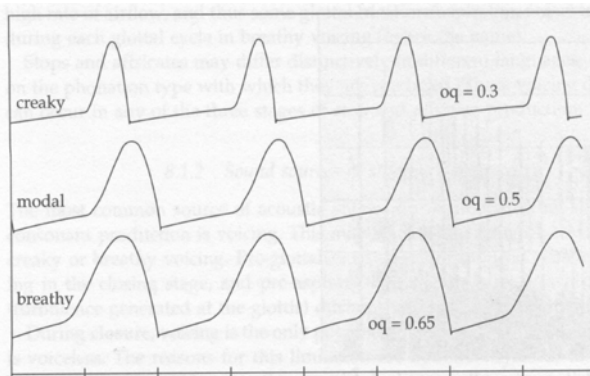

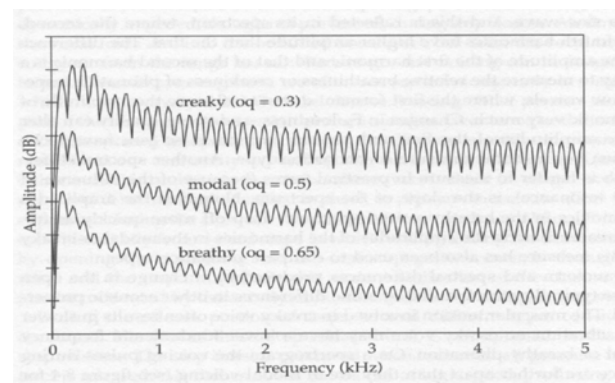Fig. 7. Glottal source waveform. From Johnson (2003)

Fig. 8. Power spectra of glottal waveforms.

f_0 of the vocal fold vibration is dependent on several factors such as mass, length and tension of the folds which are interrelated in a fairly complicated way. Typical average values for f_0 are as follows:

adult males voice: 125 Hz
adult female voice: 220 Hz
child's voice: 300 Hz

During normal speech production, the frequency of voicing varies over an octave or more (e.g., for an adult male voice the range will be from 80 Hz to 160 Hz).

4.2. Vocal tract filter

Vocal tract filter selectively passes energy in the harmonics of the source. The size and shape of the vocal tract determine for each harmonic of the source the relative amount of energy that is passed. Characteristic resonances of the vocal tract are called **formants** (F_1, F_2, F_3 etc). The vocal tract transfer function for a particular vowel is defined by the centre frequency and bandwidth of these formants. In the study of speech sounds we are mostly interested in the first three or four formants.

We can model the acoustic properties of the vocal tract as a tube open at one end (mouth) and closed at the other (glottis). Assuming that this tube is uniform in its cross-section – as in production of mid-centre schwa – we can calculate resonant frequencies of that tube (the frequencies that will produce *standing waves*) by using the following formula:

$F_n = (2n – 1)c/4L$ where *n* is the number of the formant and *L* is the length of the tube

This formula derives formants for tubes with uniform cross-sectional area only, but we need to analyse acoustic properties of vowels other than schwa and their production involves constrictions in vocal tract. One way of modelling the acoustic properties of vowels is to represent the vocal tract as a concatenation of tubes (Fant, 1960). Alternative approach is known as perturbation theory which models vowel acoustics in terms of relationship between air pressure and velocity (Chiba and Kajiyama, 1941).

4.3. Formant frequencies of the vowels

First formant frequency (F_1) is traditionally referred to as the frequency of the pharyngeal/back cavity, but in fact it is influenced by the shape of the entire vocal tract. F_1 is inversely related with tongue height: low vowels have high F_1 and high vowels have low F_1.

The size and length of the oral/front cavity are the main factors determining **second formant frequency (F_2)**. F_2 is associated with the front-back dimension: front vowels have high F_2 while back vowels have low F_2; the formant frequencies decrease through the cardinal vowels. However, the relationship is not straightforward because of the possible effects of lip rounding, which always lowers all frequencies; thus [ɑ] is "backer" than [u] but has higher F_2.

Third formant frequency (F_3) varies less then the first two formant frequencies. Lip rounding and retroflexion of the tongue have the biggest effect on this formant, both causing considerable lowering.
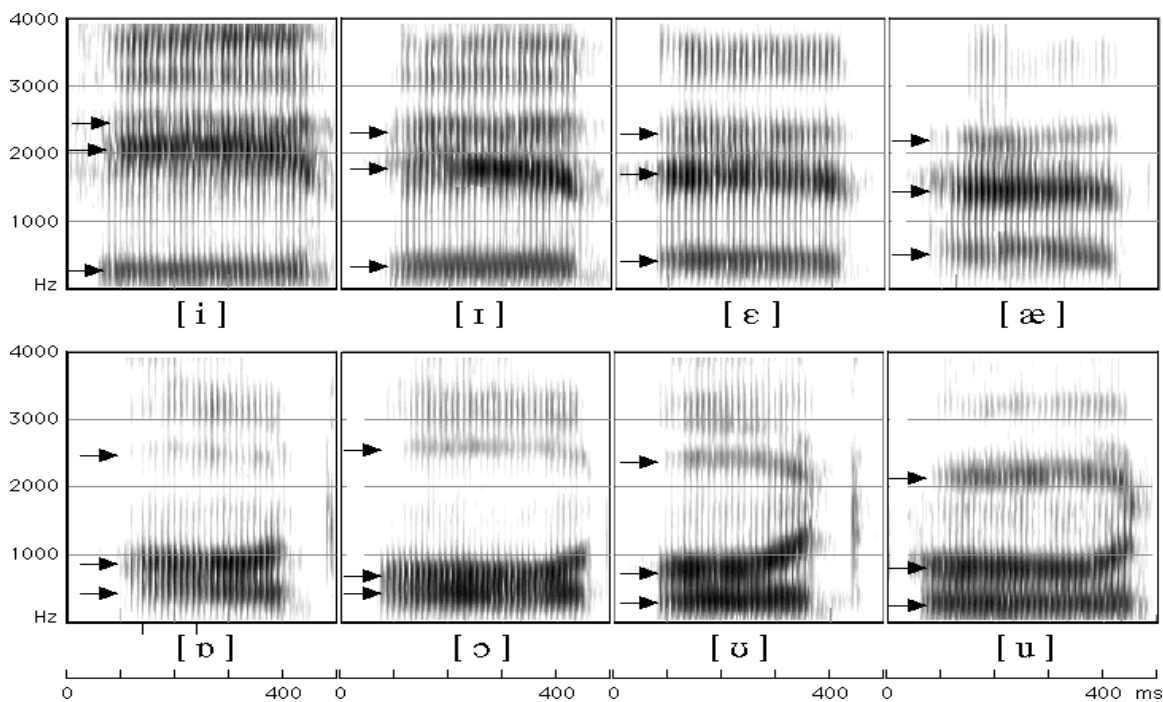
Fig. 9 Spectrograms of 8 British English vowels. From Ladefoged, Peter (2001)

Articulatory properties of vowels can be related to the first two formant frequencies by the means of formant charts that plots F_1 against F_2 (Fig. 10). Because of the inverse relation between articulatory parameters and formant frequencies the graph is represented in such a way that zero frequency would be at the top right corner.
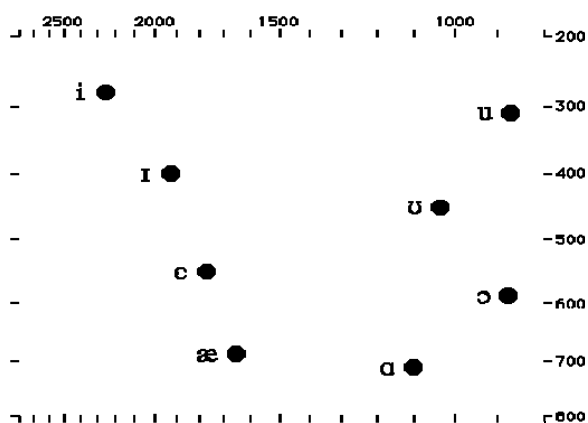

Fig. 10. A formant chart plotting first and second formants for 8 English vowels. From Ladefoged, P. (2001)

Reading:

Chiba, T. and Kajiyama, M. (1941) *The Vowel: Its Nature and Structure*. Tokyo: Kaiseikan.

Fant, G. (1960) *Acoustic Theory of Speech Production*. The Hague: Mouton.

Fry, D. B. (1979) *The Physics of Speech*. Cambridge: CUP (chapters 4-7 and 9).

Johnson, K. (2003) *Acoustic and Auditory Phonetics*. 2nd edition. Oxford: Blackwell (chapters 5-6).

Kent, R. D., Dembowski, J. and Lass, N. J. (1996) The Acoustic Characteristics of American English. In Norman J. Lass (ed), *Principles of Experimental Phonetics*, pp. 185-225.