Lecture 15: Frequency-Time Analysis of Sounds using a Windowed Fourier Method

©Christopher S. Bretherton

Winter 2015

15.1 Character of music and speech

Speech and music involve producing a certain sound (set of frequencies in a given proportion) for some time interval, then producing another sound, etc. We hear the sound by air pressure oscillations vibrating our ear drum. While each sound is being produced, the pressure signal may be expected to have a well-defined power spectrum corresponding to the frequencies contained in it. That signal can be isolated by windowed Fourier analysis. The cochlea in our ear (Fig. 1) does a fascinating natural version of this. It is a coiled tapered fluid-filled tube. Pressure waves of different frequencies are damped at different rates as they move along the tube, with the lowest frequencies getting the furthest before being damped. Nerve cells along the length of the cochlea thus allows the ear to naturally separate the frequency content of the sound, essentially doing a continuous version of a windowed power spectral analysis.

15.2 A musical example

We apply windowed Fourier analysis to a short segment of Handel's Messiah, sampled for about 9 secs at 8192 Hz (samples per sec). Script **music.m** loads and plays the music, then goes through all the steps needed to use the windowed analysis to identify the different sound frequencies (musical notes) present at different times. The time-frequency plot of spectral power that is generated is called a **spectrogram**. In the first second, we can identify the notes D (in two octaves), A, and F, all tuned slightly flat (a common custom for Baroque music), which also can be seen at the start of the fourth measure of the score.

An important consideration in this type of joint time-frequency analysis is the tradeoff between window length N_w, the number of half-overlapping power spectra to average n_a, and the time resolution δt_a of the analysis. If the sampling frequency in Hz is f, the time between samples is $\Delta t = f^{-1}$. Hence, each window spans a time

$$\delta t_w = N_w \Delta t \qquad (= 256/8192 = 1/32 \text{ s} \text{ for our example})$$

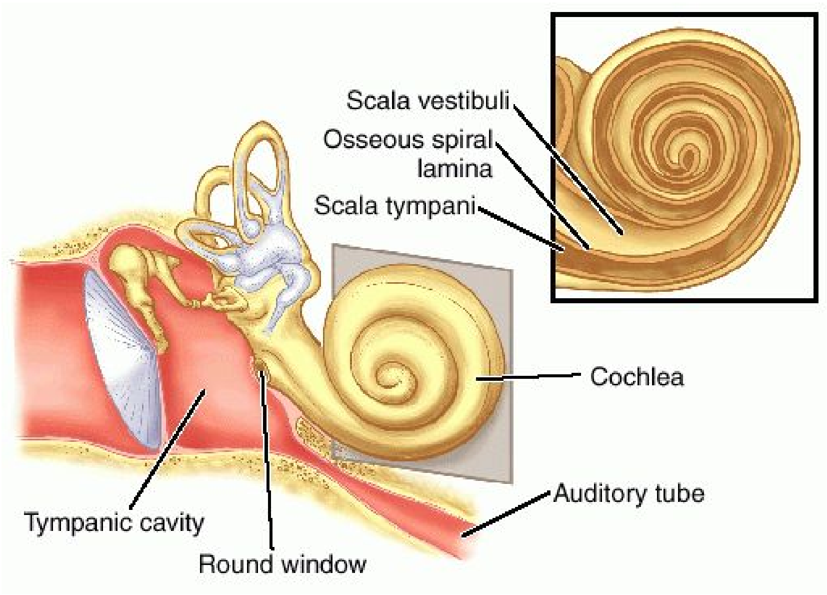

Figure 1: The cochlea (http://medical-dictionary.thefreedictionary.com/cochlea)

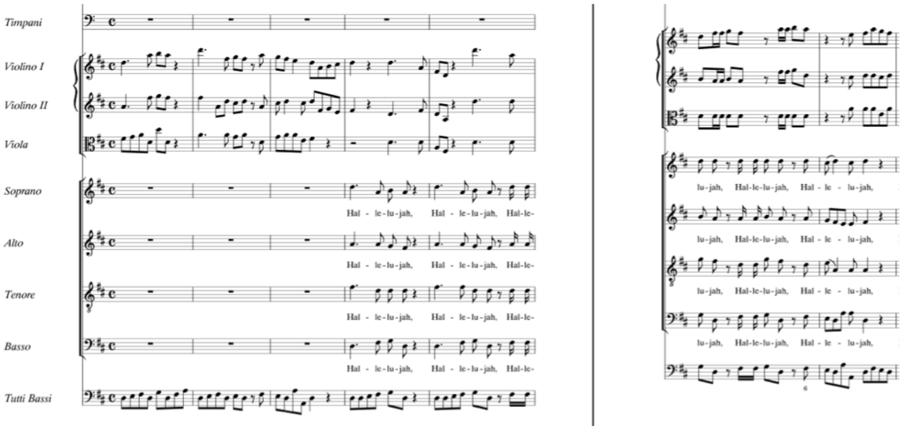
Timpani
Violino I
Violino II
Viola
Soprano
Hal - le-lu-jah, Hal - le-lu-jah, Hal-le-
Alto
Hal - le-lu-jah, Hal - le-lu-jah, Hal-le-
Tenore
Hal - le-lu-jah, Hal - le-lu-jah, Hal-le-
Basso
Hal - le-lu-jah, Hal - le-lu-jah, Hal-le-
Tutti Bassi
lu-jah, Hal-le-lu-jah, Hal - le - lu-jah,
lu-jah, Hal-le-lu-jah, Hal - le - lu-jah,
lu-jah, Hal-le-lu-jah, Hal - le - lu-jah,
lu-jah, Hal-le-lu-jah, Hal - le - lu-jah,

Figure 2: Excerpt from score of the *Messiah*. Our segment starts a little before the start of the fourth measure and after 5 seconds continues off the page.

On the other hand, the frequency resolution (the difference between adjacent DFT frequencies, given in units of Hz, so no 2π factor) is is

$$\delta f_w = 1/(N_w \Delta t) = \delta t^{-1} \qquad (= 32 \text{ Hz} \ \text{ for our example})$$

or

$$\delta t_w \delta f_w = 1$$

This is an example of the 'Heisenberg uncertainty principle' for any waves; there is a trade-off between the sampling time δt_w and the frequency resolution δf_w. To resolve notes in the 500 Hz range spaced 1/12 octave apart, we need roughly 30 Hz frequency resolution, which is what motivated the choice $N_w = 256$. On the other hand, this requires a minimum window length of 1/32 s. Furthermore, to get statistically reliable spectral estimates we need to average spectra from adjacent overlapping windows. Using $n_a = 4$ brings the noise in these estimates down enough to be useful, but this now doubles the time between spectral estimates to used for spectral averaging

$$\delta t_a = n_a \delta t_w / 2 \qquad (= \ 1/16 \text{ s} \ \text{ for our example})$$

Given that music has a typical cadence of 2 beats per second, some notes may be sustained for periods of as little as 1/8 of a second, so we need to spectrally analyze them in a time interval shorter than this. This is why we did not use a much longer window or more windows per spectral average for our analysis.