1 Observations

For this assignment, we consulted mainly "Elements of Acoustic Phonetics" by Peter Ladefoged, 1996.

For the purpose of this assignment, we are interested in formants up to 4000 Hz and therefore we can downsample the audio to 8000 Hz. This is enough as it will allow us to capture frequencies up to 4000Hz, as the Nyquist theorem states. The Nyquist theorem states that the value of a signal (in units of samples per second), if it is to be free of the distortion known as aliasing, needs to be equal to twice the highest frequency of a given function or signal. This also means that, for a given sample rate, the corresponding Nyquist frequency is one-half the sample rate in Hz.

When analysing an average male voice, we can observe around 1 formant per 1000Hz. In general, a formant can be represented as a damped sinusoidal wave. Knowing the two previous points is enough to predict any point in such a wave, so therefore we need at least 2 coefficients (so we can calculate those two points) per formant. This amounts to at least 8 coefficients all together. With fewer poles than that the formants might not be captured well. Our recordings are made by a female speaker.

We tested an LPC model of the orders 6, 8, 10, 12, 16, 40 (see figures 1 and 2). According to the rule of thumb[1] while choosing the number of poles (orders), 8 poles should be sufficient for our model and 10 would be best.

As expected, with model order 6, the formants in the vowel of "four" are not well captured, nor are those of the vowel of "three". There are peaks visible, and some bumps where we would expect the second formant for the vowel of "four". For the vowel in "three", the envelope is very smooth, only capturing two main peaks (and two mirrored).

With model order 8, the second formant of "four" is captured only as a kink in the peak of the first formant. The two formants are quite close in frequency (at 450Hz and 800Hz), which makes it challenging for the model to capture them.

Order 10 seems to provide the best results for the vowel of "four". It also captures all formants of the vowel of "three", even though the second formant is only represented by a slight bump.

With model order 12 we observe spurious formants captured for the vowel of "four". For

[1] "A general rule of thumb for the number of coefficients is the sample rate in kHz plus 2, e.g., 10,000 Hz = 10 kHz plus 2 equals 12" (Ladefoged, 1996)

the vowel of "three", we now see a peak at the second formant location. However, we also see a fake formant captured around 500Hz, which is a multiple of the fundamental frequency, which measures at 263Hz.

For the higher model orders we observe that a lot of the harmonics are falsely recognised as formants. Further, the LPC predicted values are all plotted quite a bit higher than the actual spectrum.

The right amount of poles would be where no harmonics are found, and only the global peaks, the formants, are captured. Too many poles result in capturing more peaks in the magnitude spectrum, which are often harmonics, or sometimes just noise. For our data, this happens once we go up to order 20 where we can observe small peaks that do not correspond to actual formants.

References

Ladefoged, P. (1996). *Elements of acoustic phonetics*. University of Chicago Press.

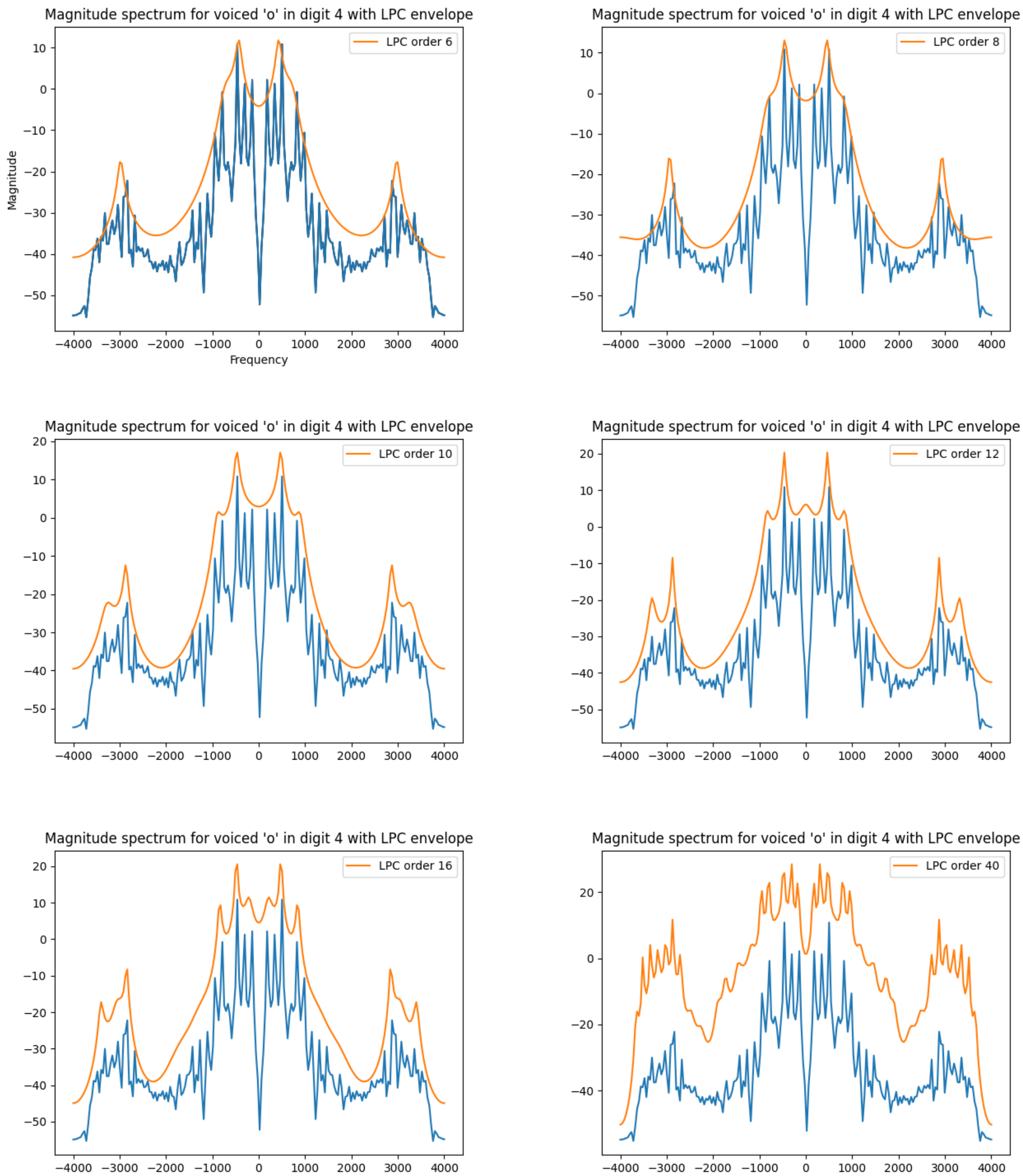

Figure 1: Effect of LPC model order on magnitude response, voiced region of digit 4

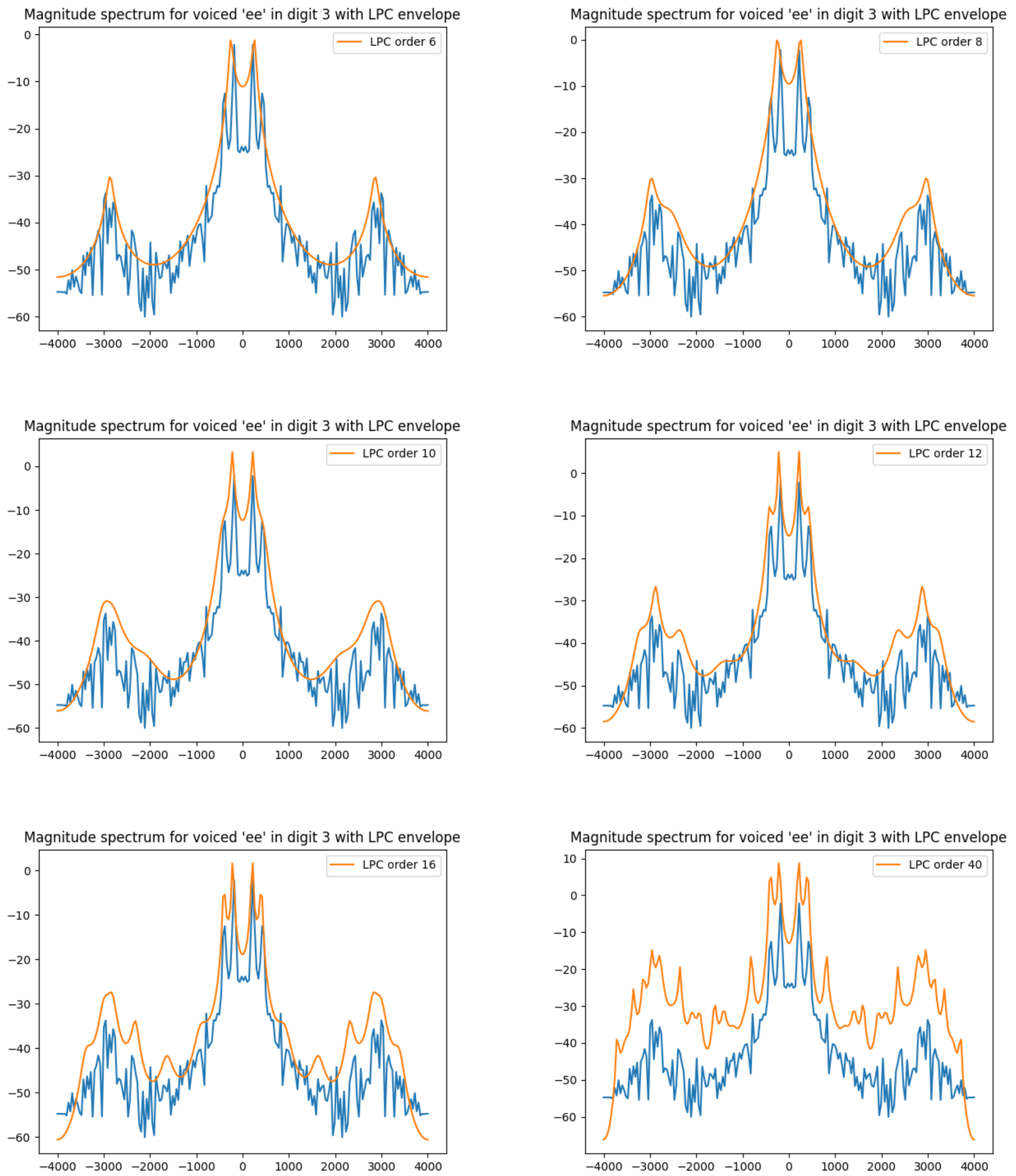

Figure 2: Effect of LPC model order on magnitude response, voiced region of digit 3