# ASR: Past Paper May 2022

Completed on May 5, 2023

**Patrick Tourniaire**

# Problem 1

**Subproblem 1.** *What is the difference, if any, between pitch and fundamental frequency?*

**Answer**
The fundemental frequency is closely related to pitch, which is defined as our perception of fundemental frequency. That is, the $F0$ describes the actual physical phenomenon, whereas pitch describes how our ears and brains intrepret the signal, in terms of periodicity.

**Subproblem 2.** *When we hear the high voice of children, is it because of their shorter vocal tract, their shorter vocal fold, or both? Why?*

**Answer**
When we hear the high-pitched voices of children, it is mainly due to the shorter length of their vocal tract, rather than the lenght of their vocal folds. The pitch of a voice is determined by the frequency of the vibrations produced by the vocal folds. When the vocal folds vibrate at a higher frequency, the resulting sound has a higher pitch. However, the vocal tract, which includes the mouth, throat, and nasal cavity, also plays a significant role in shaping the sound produced by the vocal folds.
In children, the vocal tract is shorter and narrower than in adults, which results in a higher resonance frequency. This means that when the vocal folds vibrate, the resulting sound waves are amplified at a higher frequency, resulting in a higher-pitched voice. As children grow and their vocal tract elongates, their voices gradually deepen and become lower in pitch.
So, to sum up, the high-pitched voices of children are primarily due to the shorter length of their vocal tract, which amplifies the higher-frequency vibrations produced by their vocal folds.

**Subproblem 3.** *In the ideal speech production model described in class, what is the connection, if any, between formants and fundamental frequency?*

**Answer**
The fundemental frequency is the first frequency component of the glottal pulse, whereas formants are the resonance frequencies of the vocal tract. Which leads to the production and perception of certain phones, particularly vowels.

**Subproblem 4.** *Could we infer fundamental frequency from log Mel spectrograms? If so, how?*

**Answer**
Yes, it is possible to infer the fundemental frequency (also know as $F0$ or pitch) from log Mel spectrograms is to use a technique called the autocorrelation method. The basic idea behind this method is to calculate the correlation between the spectogram and a delayed version of itself, and then identify the delay that results in the highest correlation. This delay corresponds to the period of the fundemental frequency, which can be used to calculate the $F0$.
Another approach is to use a DNN based method to directly estimate $F0$ from the log Mel spectogram. This involves training a NN on a large dataset of audio recordings and their corresponding $F0$ values, so that the network learns to recognise patterns in the spectogram that are associated with different pitch values. Once the network is trained, it can be used to predict $F0$ for new spectograms.

**Subproblem 5.** *One of your friends taking the ASR course suggests we use a 3-state HMM to model the high (H) and low (L) of fundamental frequency contours. At each state, the HMM can emit a symbol H or a symbol L. You can find the transition probabilities and the emission probabilities in Tables 1 and 2, respectively. We only allow sequences that start at state 1 and end at state 3. In other words, the prior probability is 1.0 for state 1 and 0.0 for others*
*What is the joint probability of emitting HHLLL for the state sequence 12223?*

Problem 1 continued on next page. . .

**Answer**

$$Q = [1, 2, 2, 2, 3] \tag{1}$$
$$X = [H, H, L, L, L] \tag{2}$$

Using these parameters we can calculate the joint probability of the sequence in the HMM.

$$\begin{aligned} P(X, Q; \lambda) &= P(1)P(H|1)P(2|1)P(H|2)P(2|2)P(L|2)P(2|2)P(L|2)P(3|2)P(L|3) && (3) \\ &= 1.0 \times 0.75 \times 0.80 \times 0.50 \times 0.60 \times 0.50 \times 0.60 \times 0.50 \times 0.40 \times 0.75 && (4) \\ &= 0.0081 && (5) \end{aligned}$$

**Subproblem 6.** *What is the most likely state sequence given that we observe HHLLL?*

**Answer**
By maximising the joint-probability, the most likely state sequency for the above observation would be $Q = [1, 1, 2, 3, 3]$.

**Subproblem 7.** *What is the marginal probability of emitting HHLLL*

**Answer**
*TOOD:* Calculate using forward probabilities.