Automatic Speech Recognition
Lecture Note 6: Sound Wave Basics

1. **Wave Equations**

   A wave equation is given by

$$c^2(\frac{\partial^2 y}{\partial x^2}) = \frac{\partial^2 y}{\partial t^2}.$$

   The solution to this equation is

$$y = f(x - ct) + g(x + ct),$$

   which apparently gives the wave velocity of c. A fast-vibrating string satisfies the wave equation, as is demonstrated in the textbook.

2. **Sound Waves**

   (cf. Feynman Lectures on Physics) The sounds that we hear correspond to pressure waves in the air. A sound source disturbs the air in its vicinity. This disturbance creates a density change, which in turn changes the pressure. The difference in pressure results in net force and set the nearby air in motion, thus relaying the disturbance. Let the displacement of the air at position x and time t be $z(x,t)$, and similarly for pressure change $p(x,t)$ and density change $\rho(x,t)$. From Newton's second law, we have

$$\rho_0 \frac{\partial^2 z}{\partial t^2} = -\frac{\partial p}{\partial x}.$$

   From the relation between pressure and density, we have

$$p = \alpha \rho,$$

   where α is a constant given the original density ρ_0 of the air. Finally, we have, from the conservation of mass, that

$$\rho = -\rho_0 \frac{\partial z}{\partial x}.$$

   Putting things all together, the displacement $z(x,t)$, and similarly for p and ρ, or partial derivatives such as $\frac{\partial z}{\partial t}$, satisfies the wave equation

$$\frac{\partial^2 z}{\partial t^2} = c^2 \frac{\partial^2 z}{\partial x^2},$$

   with $c^2 = \alpha$.

3. **Resonances of Vocal Tract and Formants**

   The boundary condition of a sound source puts constraints on the frequencies of the resultant pressure waves. In the case of vocal tract as the sound source, the configuration of the vocal tract determines which frequencies are resonating and these frequencies are called the formants (or formant frequencies). Note that this is different from the fundamental frequency, which is the vibrating frequency of the vocal cords. In an over-simplified calculation, as shown in the textbook, the formants are $500, 1500, 2500, 3500$ Hz, corresponding to roughly to the neutral vowels.

4. **Discrete-time Simulation of An Acoustic Tube**

   Refer to Figure 10.5 in the textbook, one can approximate the velocity and pressure traveling in an acoustic tube with a discrete-time linear difference equation (of order $2M$),

   $$y[n] = x[n] \pm y[n-2M],$$

   where M is the number of delays in the upper track and physically corresponding to the one-way traveling time of the wave. The transfer function is

   $$H(z) = \frac{Y(z)}{X(z)} = \frac{1}{1 \mp z^{-2M}}.$$

   The poles of these transfer functions are located on the unit circle. Let the sampling period be T, and the tube has an open end so the plus sign applies. In this case, the physical frequencies of resonance are

   $$\omega_n = \frac{(n+\frac{1}{2})\pi}{MT}.$$