Acoustic Waves

Notes on Speech and Audio Processing

Chia-Ping Chen

Department of Computer Science and Engineering
National Sun Yat-Sen University
Kaohsiung, Taiwan ROC

Introduction

- The sound we hear, especially speech and music, is a perception of air waves produced by the source.
- The pattern of air vibration is characteristic of the instrument. Instruments are excited in several ways
  - plucked instruments: guitar, harp
  - bowed instruments: violin, cello
  - struck instruments: piano, xylophone
  - tube resonance: brass, wind instrument
  - human vocal tract
- The vibrations can be mathematically described by the wave equations, with solutions determined by the boundary conditions.

Wave Equation for String

- We begin our discussion with the simplest case – the vibration of a string.
  - In Figure 10.1, consider what happens in the infinitesimal dx. From Newton's second law

$$F_y = S(\tan\phi_2 - \tan\phi_1) = S(\frac{\partial^2 y}{\partial x^2})dx = ma = (\epsilon dx)\frac{\partial^2 y}{\partial t^2}$$

$$\Rightarrow c^2(\frac{\partial^2 y}{\partial x^2}) = \frac{\partial^2 y}{\partial t^2}, \text{ where } c = (S/\epsilon)^{1/2}.$$

  - The above is the wave equation. You can verify that the solution is $y = f(x - ct) + g(x + ct)$, and c is velocity of the wave.

Discrete-Time Traveling Waves

- For computer simulation, we need to discretize both time and space. This is shown in Fig 10.2.
- If an external stimulus is applied, a wave traveling to the right is created in the upper track and another traveling to the left is created in the lower track. The output of the system at any time is the pattern of all the sums shown in the figure. This is the solution.
- Since both x and t are now discrete, we can replace x by mX and t by nT, leading to the solution

$$y(mX, nT) = Af(mX - cnT) + Bg(mX + cnT).$$

Boundary Conditions

- The behavior of a traveling wave at the interface is specified by the boundary condition.
  - If the end is fixed, then a wave is inverted there, and continues traveling in the opposite direction.
  - If the end is loose, then a wave is not inverted there, and continues traveling in the opposite direction.
- For string waves, both ends are fixed. The situation is illustrated in Figure 10.3.

The Physics of Sound Wave

(cf. Feynman's Lectures on Physics) We hear sounds due to air pressure change at the ear. The generation of sound wave consists of the following processes.

- 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. The disturbance pattern of the air is thus relayed.

Equation of Motion

- More precisely, let the displacement be $z(x,t)$, the pressure change be $p(x,t)$, and the density change be $\rho(x,t)$.
- Considering what happens in the infinitesimal volume of air Adx. From Newton's second law

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

- Further, p and ρ are approximately linearly related

$$p = \alpha\rho,$$

Sound Wave Equation

- From the conservation of mass

$$(\rho + \rho_0)dx(1 + \frac{\partial z}{\partial x}) = \rho_0 dx \;\Rightarrow\; \rho = -\rho_0 \frac{\partial z}{\partial x}.$$

- Putting 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 $c^2 = \alpha$.

Discrete Simulation

- All physical variables satisfy the same wave equation, and their solutions have the same form.
- In particular, we can write

$$u(x,t) = u^+(x-ct) - u^-(x+ct)$$

$$p(x,t) = Z_0(u^+(x-ct) + u^-(x+ct))$$

- This is illustrated in Fig 10.5. The upper track contains $u^+(mX - cnT)$, and the lower track contains $u^-(mX + cnT)$. -1 at the right end signals the right-end is closed. $+/-$ at the left means the tube may be open/close at the left end.

Acoustic Tube Resonances

- Consider the discrete acoustic tube model with linear difference equation $y(n) = x(n) \pm y(n-2M)$. The frequency response is

$$|H(e^{j\omega T})| = |\frac{Y(e^{j\omega T})}{X(e^{j\omega T})}| = \frac{1}{|2\cos(M\omega T)|}.$$

- The closer ω is such that $\cos(M\omega T) = 0$, the larger the magnitude response.
- The frequencies at which $|H(e^{j\omega T})|$ peaks are called resonance frequencies.

Formants

- The human vocal tract can be modelled by an acoustic tube: one end is the glottis and the other end is the lip(s).
- A formant is a frequency with high energy in the spectrogram.
- For the neutral vowel, the formants are found to be around $500, 1500, 2500, 3500$ Hz.
- If the length of an acoustic tube is 17cm, then the resonance frequencies are also $500, 1500, 2500, 3500$ Hz!

Acoustic Tube Models for Phonemes

- The area function from X-ray tracing for phoneme /i/ is given in Fig 11.1. This area function can be approximated by a concatenation of cylinders.
- Our goal is to establish the relationships between the acoustic tube structures and the resonance modes (frequencies).
- (Referring to Fig 11.2) Let u_k, p_k be the velocity and the pressure of section k. Then from earlier analysis,

$$u_k = u_k^+ - u_k^-,$$
$$p_k = Z_k(u_k^+ + u_k^-).$$

Two-Tube Model I

- Let the delay from section k to section $k+1$ be M samples, then

$$u_{k+1}^{+} = z^{-M} u_k^{+}, \quad u_{k+1}^{-} = z^{M} u_k^{-}.$$

$$\Rightarrow \begin{cases} u_{k+1} = u_{k+1}^{+} - u_{k+1}^{-} = z^{-M} u_k^{+} - z^{M} u_k^{-} \\ p_{k+1} = Z_{k+1}(z^{-M} u_k^{+} + z^{M} u_k^{-}). \end{cases}$$

- Therefore, (p_k, u_k) and (p_{k+1}, u_{k+1}) are related by

$$\begin{bmatrix} p_k \\ u_k \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} p_{k+1} \\ u_{k+1} \end{bmatrix}.$$

Two-Tube Model II

- Going one step further, one can relate (p_k, u_k) and (p_{k+2}, u_{k+2}) by

$$\begin{bmatrix} p_k \\ u_k \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} p_{k+2} \\ u_{k+2} \end{bmatrix}$$

- Setting $p_{k+2} = 0$ (when the mouth is open),

$$u_{k+2} = \frac{u_k}{B_{22}},$$

  so the formants occurs when $B_{22} = 0$.

- Results of two-tube model are shown in Fig 11.4.

Excitation Mechanism in Speech

- The vibration of vocal folds provides periodic excitations for vowel and vowel-like sounds.
- Glottis can remain open (with relaxed folds) when air passes through it. Constrictions in the vocal tract produces turbulence, resulting in noise-like voiceless sounds.
- Constriction and vibration can occur simultaneously, for voiced consonants.
- Transient is another kind of excitation for stops and plosives.