Lecture 2: January 20, 1995

Lecturer: Prof. Ben Gold *Scribe: Nikki Mirghafori*

2.1 A Brief Historical Survey

Today we briefly cover the history of speaking machines (such as Voder) and music machines. The first question that comes up is, why do we discuss music and speech together? In response, let's take a look at the history of the two subjects in ancient Greece.

2.1.1 Ancient Greece

The educational system of ancient Greece was divided into two main curricula: the *Trivium* and the *Quadrivium*. The Trivium was the "verbal" discipline, comprised of grammar, rhetoric, and logic, while Quadrivium was the "numerical" discipline, comprised of astronomy, geometry, mathematics, and music. speech ($\subset$ Trivium) and music ($\subset$ Quadrivium) are similar in the following ways:

- Understanding: We may understand speech in a similar way that we distinguish between types of music.
- Analysis: We may perform a *physical* (e.g. studying the vocal tract) or *functional* (e.g. breaking speech into Fourier analysis) analysis of music and speech.
- Synthesis
- Perception: Are the pathways for music perception identical to speech perception?

2.1.2 Speaking Machines

The oldest "speaking machines" are the oracles of ancient Greece (BC). A tube carried the speech of a hidden speaker into the head of the oracle, and the oracle (supposedly) predicted the future.

In the middle ages (around 1000 AD), there were *talking heads* which were very similar to the oracles of the ancient Greece, except that the aim was to understand the speaking process rather than to influence people.

In the 1700's (Renaissance), the quest for knowledge became more important. In 1779, the Imperial Academy of St. Petersburg offered its annual prize for explaining the physiological differences between five vowels and for making an apparatus to produce them artificially. Kratzenstein, the winner of the prize, constructed acoustic resonators similar in shape to the human vocal tract and activated the resonators with vibrating reeds which, in a manner analogous to the human vocal cords, interrupted an air stream [FL72].

In his paper, Dudley [DU50] writes that the shapes of the tubes were not essential to creating the vowels, and that the vowels could be produced using a tube of varying length. It is rumored that Newton experimented with half empty beer bottles and observed the same effect.

Von Kempelen's speaking machine was the first to attempt the reproduction of all sounds. He spent 20 years working on his speaking machine (completed around 1780) and kept very detailed records of its inner-workings[1]. Sir Charles Wheatstone used Von Kempelen's notes to make his own version of the speaking machine (1879). Wheatstone, in turn, was the inspiration to the Bell Family.

Voder by Dudley (1930's) functionally (not physically) produced the speech of humans. Voder had a console with 12 keys, a wrist bar, and a pitch pedal. The wrist-bar controlled the buzz-hiss, and the ten keys controlled ten bandpass filters. (Listened to voder's original tapes: vibrato, singing, use of prosody, various pitches, etc.)

2.1.3 Music Machines

Going back to ancient Greece, from historical records it seems that they had knowledge of scales and relations between notes and the physics of strings.

The first known "music machine", was the *water-powered barrel organ* in the 1500's. The idea of the barrel organ is very similar to the contemporary music boxes, where there is a cylinder with fixed bumps which plucks cords. The main limitation is that the music box only plays one tune, and is not adjustable in real-time.

The *Player Piano* was the next music machine (1880-1900). The music was punched in *punched rolls* (similar to IBM's Babbage computer's punched cards). In the case of the street organs, punched cards were used and the power came from the organist. The machine was also capable of recording the tune the organist played by punching holes. The important point about the Player Piano is that the music came directly from data.

The *Telharmonium* by Thaddeus Cahill (1867-1934) worked based on additive synthesis. The idea behind it was that any sound can be produced if you can produce enough sinusoids and change the amplitude (envelope control). The first performance was in 1906 (before the electronic era) in Holyoke, MA. At the turn of the century, the electric dynamo was built, which then powered the Telharmonium. The main problem with this machine was that it was very large and expensive. Besides, with the invention of triode, the electronic's took over and the Telharmonium never became widespread.

Theremin (early 1920's) was a very simple machine with a transmitting antenna and a receiving antenna. The performer's hands would move between the two antennas, and the change in electromagnetic field at both the transmitter and receiver would create sounds that were similar to a female voice singing mass.

References

[F72] J. Flanagan *Speech Analysis, Synthesis and Perception*, Springer–Verlag, New York, 1972, pp. 205–215.

[DT50] H. Dudley and T.H. Tarnoczy, "The Speaking Machine of Wilhelm Von Kempelen". *Journal of Acoustical Society of America*, Vol 22.

[1] Copies of Von Kempelen's books exist today in French and German, but not in English.