Alex Kelley, Kevin Dade, Andrew Kang
Speaking Synthesizer Write up
Math 5: Prof. Barnett
Final Project

Wolfgang von Kempelen was an 18th century Hungarian inventor who served under Empress Maria Theresa of Vienna. Among other things, he is famous for his work with phonetics, and for creating a machine capable of simulating human speech. Von Kempelen's Talkbox, as the device came to be known, was intricate by necessity. A bellows supplied constant airflow, reeds filled the role of the vocal cords, and an elaborate system of whistles, percussive devices, and openings allowed for the articulation of consonants. Most importantly, when demonstrating his device, von Kempelen would manually control the shape of a leather tube in order to imitate the vowel sounds.

While all of the elements were necessary to truly "speak", the leather tube most obviously resembles a component of human anatomy. When a human talks, the vocal cords generate a pitched sound that is then filtered through the vocal tract, which is essentially a tube. Constricting the vocal tract at different locations changes the resonant properties of the pipe, and thus filters the sound in different ways. A simple example of this would be humming while opening the mouth as wide as possible, and relaxing the throat. Popular among dentists, the "aah" phoneme is one of the most relaxed, basic sounds a human can make.

The "aah" phoneme is one of three vowels that were studied; we also studied the 'ooh' and 'ihh' vowels. In linguistics the study of vowels has led to various descriptions of the syllabic sounds we make. The "aah" vowel is described as a low back unrounded tense vowel. The 'low back' portion of this description, explains the area in which the sound seems to resonate within our mouths. Whether the vowel is rounded or unrounded is a description of whether or not the vocal tract is compressed. Vowels are also described as tense or lax, which describes whether vowels require a particularly significant amount of effort or 'friction' in our vocal tract. The "ooh" vowel is described as a high back rounded tense vowel and the "ihh" vowel is described as a high front unrounded lax vowel. These vowel sounds are created from the human vocal tract in the following ways. The "aah" sounds is created when the vocal tract is clear and relatively uniform in width, the "ooh" sound is created with the round constriction of the sound at the mouth, and the "ih" sound is created with the flattening constriction of sound near the soft palate. Using this knowledge we attempted to recreate the constrictions and shapes of our vocal tract using Von Kempelen's leather tube.

Our results were marginally successful. For the “aah” sound, which is the most unrestricted and open sound, we did not put pressure on the tube. The spectrogram for this sound made vocally is while the spectrogram from the tube is as followsWe tried to make the tube roughly the same length as the vocal tract, 17 cm. As one can see, the formant pattern was the same in both cases. An unrestricted tube of length 17cm will have formants at 340/(.17\*4), or with the end correction, a little bit over 500 Hz. The second formant will be around 1500 Hz. For the vocal spectrogram, the first two formants are at 700 and 1550 Hz. The resulting spectrogram for the device showed the first two formants at 800 and 1400. The top two formants were also comparable. Theoretically they should be at 2500 and 3500 for a completely round tube. Since our vocal tract, even when saying “aah” is not round, predicting formants accurately is difficult. Also, the pipe we used is imperfect in this way. The glue on the inside creating the rounded leather is raised in some places, and the leather’s natural position is more of an oval.

 The next sound we replicated was the “ooh” sound. This is made mainly by closing one’s lips. Since the mouth is a node, restriction at a node would cause all of the resonant frequencies of the pipe, the formant frequencies, to lower in pitch. The vocal spectrogram is as followswhile the device’s replication is this:

The general pattern of formants is what our brain recognizes as a vowel. For “oo” the pattern is formant at a low frequency, with a group of higher frequency formants separated by a large gap, as seen in the spectrograms above. The lowest formant for the vocal “oo” is at around 400. It makes sense that this one is lower than the fundamental frequency of the vocal tract because the lips at the end close at a node. By similarly closing the end of the device, we achieved the same result.

 For the “ee” sound, one primarily has to flatten the back of their throat by the tongue. The spectrogram shows the first formant at around 375 Hz, therefore the constriction is closer to the node of the first formant. The subsequent formants are higher, showing that the constriction is closer to an anti-node for the other formants. The difference between the “ee” sound, which we originally tried to emulate, and the “ihh” sound is that there is slightly less constriction to the front of one’s mouth. Since the tube did not have optimal malleability, we could not accurately model the “ee” sound. Instead, we came closer to the “ihh” sound. Because the “ihh” sound has less constriction above the tongue, approximately where many higher formant frequencies have nodes, the upper formants are not as high. This can be seen in the spectograms. This is the vocal spectrogram for the “ee” sound, followed by the “ihh” spectrogram. This is result from the device:

The spot that models the sound optimally is the beginning, where one can make out a base formant with several more at higher frequencies, like the pattern for the “ihh” sound.

 While the spectrograms show that this experiment was reasonably successful, the actual difference in vowel sounds was barely audible. This can largely be attributed to the crude replication tube used. The tube was neither wide enough nor flexible enough, and this directly affected its effectiveness. Since the tube was so difficult to manipulate, we had difficulty making precise changes to the tube’s shape. Another factor that could have hindered our experiment is the reed used to generate sound. Reeds are consistent with von Kempelen’s model, but our device produced a high, squeaky timbre; hardly resemblant of the deeper, rich sound that human’s use to filter into speech. All in all, we met with limited success. The graphical analysis does favor our experiment, but in terms of our original goal – to audibly simulate vowels – we fell a little short. The device was well constructed though, and embodies an 18th century achievement that was completely groundbreaking.