SCI220 – Foundations of Musical Acoustics
Cogswell Polytechnical College
Fall 2008

Week 12 – Class Notes

The Human Voice (FMA)

Chapter 19: The Voice as a Musical Instrument

19.1 The Voice: A Source of Controllable Sound

The voice is a sound source whose pitch is controllable. In physical terms this means that the human voice can produce acoustic signals having repetition rates that can be varied over a large range. Also, because a singer can enunciate different sustained sounds while maintaining his/her pitch suggests that the amplitudes of its sinusoidal components are subject to control as well.

A *phoneme* is the distinct element for each speech sound. Phonemes are classified into five distinct groups: plosives, fricatives (sibilants), other consonants, pure vowels, and diphthongs.

Plosives consonants are produced by completely blocking the vocal tract and suddenly opening it to let a single burst of air through. For example, *p, t, k, b,d,* and *g* (the last three are known as voiced or pitched plosives).

A sibilant is a sound produced, like *s, sh, k, t, f,v, z, sh, zh,* and *th,* that involve a broadband (multicomponent) random source and can be sustained for any length of time. Voice and unvoiced pairs exist in this group.

Under other consonants several kinds of phonemes are included, such as *semivowels (w, y, l,* and *r)* and *nasals (m, n,* and *ng).*

Vowels are steady voiced sounds with definite pitch.

Diphthongs are quick transitions from one vowel and ending on another because of changes in the tongue. For example, *eh-ee* (mate)*, aw-oo* (moat)*, a-ou* (mount)*,* and *aw-ee* (oil).

When speech sounds are made, the larynx may or may not itself be vibrating to produce an oscillatory flow of air; it is this choice that makes the distinction between the voiced and unvoiced consonants.

We can think of the larynx as being a simple source which feeds into a small, very elongated room of complex shape formed by the vocal cavities. The acoustical response of these cavities will depend on the excitation of the frequency components of the source that match one or more characteristic vibrational modes of the cavity.

In the course of speaking or singing, one continually alters the shape of one's vocal cavities. The production of each particular vowel or consonant is associated with a fairly well-defined shape for the cavities, and therefore with a particular pattern of strong and weak responses to the various sinusoidal components of the airflow controlled by the vocal cords.

At the mouth opening, the oscillatory flow of air depends on the relation between the excitation frequency (from the larynx) and the various resonances of the vocal cavity. The mouth is acoustically important since it serves as the source for sounds as we hear them in the room.

19.2 The Larynx: A Self-Sustaining Oscillatory Flow Controller

The vocal cords, which do the actual vibrating in the larynx, are flap-like folds of muscle attached to the interior of the larynx in such a way as to produce a slit-like opening through which air can pass. The cords are capable of assuming a wide variety of shapes and spacing.

When one phonates (produces vocal sound) normally, the cords are given a shape and spacing that permits the aerodynamic forces which arise from air flowing between them to set them into oscillation. However, the speed of the airflow only slightly influences the frequency of this oscillation; the predominant control comes from the mass of the vocal cords and the muscle tension set up in them. The oscillation of the cords is of such a nature that they alternately approach one another and recede, bringing about a corresponding oscillatory decrease and increase in the amount of air that is permitted to flow between them. Not only can the speaker choose the frequency of oscillation of the cords (change pitch), he can also choose to have the cords swing with sufficient amplitude that they can press together during a controllable portion of each oscillatory cycle. Meaning that the tone color of the sound can also be changed.

For a sung pitch that has relatively high breath pressure and fairly close vocal cord spacing, the first 6 partials will have similar amplitudes (after this the amplitudes decrease by $1/n^2$). On the other hand, for a sung pitch with a gentle air stream and a vocal cords that never shut close, the fundamental frequency component is considerably stronger that the other harmonic components.

19.3 Sound Transmission through the Vocal Cavities and into the Room

The vocal tract, extending from the the larynx to the mouth/nose aperture, has the duty of transforming the simple air flow spectrum provided by the vocal cords into the recognizable acoustical patterns needed for speech and music. The larynx feed one point in an elongated and tubular, one dimensional “room” whose set of natural frequencies can be adjusted by the tongue and lips. The mouth aperture acts as a simple source for the excitation of the vibrational modes of a three-dimensional room in which we can imagine we are listening.

The pressure variations produced by the larynx in the vocal tract, and thence the strength of the resulting source at the mouth depend in a simple way on the adjustable resonance properties of the vocal tract. The pressure amplitudes produced for the various partials in the room surrounding the listener do not, however, have a simple proportionality to the strengths of the corresponding airflow components from the mouth. Simple sources radiating into a three-dimensional room have the fundamental property that the room-averaged sound pressure resulting from a given source strength is larger for high frequency sources that for those oscillating more slowly. This is because of the rapidly increasing number number of off-resonance room modes whose collected responses make up much of the sound in a room.

There is no corresponding increase in the number of modes at high frequencies in a one-dimensional room, which explains why we do not find a similar “treble boost” taking place at the junction of larynx and vocal tract.

For two different pitched sounds with the same pronounced vowel, we can see that both share similar frequency components. These common components in the frequency spectrum correspond to the characteristic frequencies of the particular vocal tract air column used by the subject when asked to pronounce a determinate vowel. Dips in the spectrum arise from the cancellation between the in-phase responses of a higher mode driven below resonance and of a lower mode driven above its natural frequency.

The problem of formant pattern ambiguity is managed by our hearing in such a way that it keeps track of the formant locations that might otherwise sandwich themselves between the voice harmonics. In speaking and singing, one is constantly going from one sound to another, and each formant moves smoothly from its position for one part of the utterance to that belonging to the next part. If the pitch is maintained constant throughout we have the spectrum envelope moving past the fixed voice harmonics to plot out their shapes in time, just as we earlier found that pitch fluctuations are able to explore the shape of a fixed formant pattern. In actual speech and singing, both processes are going on continually as we raise and lower the pitch of our voices and simultaneously change the formant patterns belonging to the separate parts of the word we are enunciating.

19.4 The Male Voice and the “Singer's Formant”

For human speech, it is not difficult to recognize a particular vowel pronounced by a singer at different pitch ranges. For a transposed pitch that lies an octave or two above or below a given pitch, say C4, we can recognize the vowels sounds as having a rather constant tone color. We would could detect, at a more subtle level, a trend toward what many people would call brightness or lightness in this sound as we go up the scale, and a corresponding darkening as we go down.

It is possible to identify vowels even when shifting formant frequencies as long as it lies within a 25% range. A 50% shift transposition of the formants (a musical fifth) will change the sounds enough to hinder intelligibility.

Opera singers and other who perform with large orchestra accompaniment have developed several interesting ways of coping with the problem of being heard recognizably.

Consider the LTAS (long time average spectrum) for ordinary speech and ordinary singing that has a shape similar as described in fig. 19.7. A singer (male) in this context might have problems being heard since he would sound very different from the orchestra, and is unlikely that he can overpower it through sheer vocal exertion. If the LTAS of an orchestra and an ordinary voice are quite similar, we would expect a certain amount of masking to take place.

The first acoustical alteration cultivated by the operatic singer to help him in the audibility contest is his habit of singing with a vocal cord placement and lung pressure relationship that produce short, sharp puffs of air in the output of his larynx. By this means he can strengthen the upper partials in his voice. The increased audibility of these upper partials helps us follow the rest of his voice components through the orchestra sound.

The second acoustical alteration is the so-called *singer's formant*. Male operatic singers do not sing words with quite the formant arrangement that they would use in speaking those same words. One significant alteration is a strongly marked *extra* formant lying between 2500 and 3000 Hz. This particular formant has a frequency that is independent of the placement of the other, more ordinary formants.

19.5 Formant Tuning and the Soprano Singing Voice

A soprano singer can tune her formants in such a way that while sustaining a note she can make small adjustments of her formant frequency components to make them coincide with some of her voice partial components. By approximating her tuning of the first formant, a large increase in loudness is produced for a sound at a given effort. Thus, she is able to compete with string accompaniment, but also it vastly expands her dynamic range.

19.6 Intermediate Voices and Various Musical Implications

Because the male voice has formant peaks whose widths are comparable to the distance between its closely spaced harmonics very little change in the loudness of such a voice would be expected when formant tuning takes place. The loudness contributed by a pair of partials that straddle a formant peak is not very different from that produced when one of these lies exactly on the peak while the other one is displaced some distance down along the shoulder.

The soprano makes almost no use of the singer's formant. The muscular requirements for formant tuning are sometimes incompatible with those needed for achieving the singer's formant.

Singers whose voices lie between the bass and the soprano are apt to borrow heavily from the techniques used by their higher- and lower-pitched neighbors. The alto will use frequently use the singer's formant and both tenors and altos will resort to formant tuning when using the higher parts of their registers, where the technique becomes acoustically more effective.