Music, Language, Speech and Brain
Proceedings of an International Symposium at the Wenner-Gren Center, Stockholm, 5-8 September 1990. Edited by Johan Sundberg, Lennart Nord, and Rolf Carlsson. Dept of Speech Communication and Music Acoustics, Royal Institute of Technology, Stockholm, Sweden. Macmillan Academic and Professional Ltd., 1991. ISBN 0-333-56429-4, ISSN 0-0083-7989.

Excerpt pp 220-231:

Analogies in the production of speech and music

J. Liljencrants

Abstract

The production mechanisms of musical sounds and of speech have many common features. Obvious parallels to speech is found among the wind instruments, but the same basic laws apply to most other musical instruments, perhaps most prominently so when the filter is described as a wave equation in space and time. Then the difference between instruments becomes largely one of boundary conditions. Some design features of wind instruments are illustrated in the light of perturbation theory and the relations between pitch and spectrum are discussed.

Source and filter

To describe and model a sound production mechanism it is convenient to view it as a system divided into source and filter. The source is some device that provides energy to the system. The physical filter is a bounded medium like the interior of a tube, a string, or a plate in where this energy propagates and reverberates. Part of this filter energy radiates away and this is modelled by some radiation impedance that represents the conversion into resulting output outside sound.

Here we use two convenient general concepts, impedance and transfer function. Impedance means the ratio between force and velocity in a mechanical case, pressure and flow in an acoustical, or voltage and current in an electrical. The acoustical radiation impedance for instance gives a measure on how much sound power you get from a specified amount of flow. The source and radiation impedances that terminate the filter ports determine how the filter energy is transmitted or reflected, they form boundary conditions and are important components to determine the transfer function. This function is the comprehensive result that shows how the ratio of output to input varies with frequency, the spectral shaping.

In voiced speech the source is at the vocal folds where the periodic sequence of air puffs constitute the acoustic source input signal. This goes into the oral cavity which fulfills the filter function. The cavity is an acoustic line with the distinctive property that its cross sectional area is variable in space and time by use of the articulators, mainly the tongue, jaw and lips. The filter output is the mouth opening where the signal meets the radiation impedance. Part of the signal is there reflected back into the filter line while the other part is radiated. In the brass and reed woodwinds the layout is very much alike, except the articulatory temporal adjustment of the filtering line is in terms of length rather than area.

In all these cases the acoustic tube is essentially closed at the source end (the transmission line is open circuit in impedance terms) since the source impedance is high, and open at the other end, the radiation impedance is low. This differs from the flutes and labial organ pipes where also the source impedance is low, so that the filter tube is effectively open at both ends.

The source-filter concept is equally well suited for other musical instruments. In the percussion the source signal is the impulse from a hammer blow. The filter can have a simple one-dimensional line structure like a string or a bar, or a more complex multi-dimensional one like a membrane or plate. In the string instruments the source signal may come from a plectrum, hammer or a bow. Here also the filter can be seen as two cascaded parts, the string and the soundboard. Often however the string characteristics are well known to the extent they are included as part of the source.

The wave equation

Propagation of sound in all media is gouverned by the wave equation pioneered by Jean Lerond d'Alembert in the mid 18th century. For one dimension x, like in a line, and a wave quantity Φ it is

$$\delta^2\Phi/\delta t^2 = c^2 * \delta^2\Phi/\delta x^2$$

Its prominently useful feature is two basic facts: that its general solution is two waves that travel in opposite directions, and that it gives a measure on the propagation speed c of those waves. This is mostly constant and is found from the ratio of a stiffness measure to mass or density. For a gas of pressure p, density p and adiabatic constant γ we have $c^2 = \gamma p/\rho$. For liquid and solid media γp is replaced by the bulk modulus B, or, in case the wave goes along a thin bar you have to use Young's modulus E. The same comes again for the wave speed in a string tensioned with a force F and having the mass M per unit length $c^2 = F/M$, and even again for a taut membrane, now with tension F per unit length and mass M per unit area.

Except this it is very frustrating to try to use the wave equation because it is the same for any field quantity you care to insert: pressure, velocity, velocity potential, displacement, density, temperature, or whatever. Also it may appear hard to grasp since the function Φ is arbitrary. But this is only the mathematical way to say that a wave may convey any kind of signal.

In passing we can note that if we elect to observe a single frequency, so that time differentiation can be represented by multiplication by $j\omega$, then the wave equation can be written in the form

$$-k^2\Phi = c^2 * \delta^2\Phi/\delta x^2$$

called the Helmholtz equation, here also using the wave number $k = \omega/c$. But still you get nothing to solve from it until you specify something more, namely some boundary conditions. The conventional simple thing is to prescribe that the two waves for instance cancel (or alternatively be equal) simultaneously at two different places. Not until then you get anything more tangible, now you can at least find a number of values of Φ that satisfy. These are called the resonance frequencies or the eigenfrequencies. Their manifestation as peaks in the signal spectrum is often called formants.

The filter resonances

Applying the Helmholtz equation to a uniform line, terminated at its ends you get particularly simple results since the solution is just two travelling sinusoidal waves. The sum of these you have all seen many times as illustrations to standing waves. Depending on if the two boundary terminations are the same or if they are dual you get two different sets of resonances. If the terminations are the same, for instance both short circuit as in the open organ pipe and the traverse flute, or both open circuit as with a string, then all resonances are integer multiples of the fundamental having one half wavelength within the line. The other archetype comes with one end open circuit and the other short circuit. This case is associated with the speech vocal tract, the brass instruments and the reed woodwinds. With a uniform tube the fundamental resonance then has a quarter wavelength in the line and the higher resonance frequencies are odd multiples of the fundamental.

A general observation is that the wave speed and the length of the line determines an average frequency spacing between the resonances, $\Delta f_{av} = c/2L$. For speech this average is about 1 kHz contrasting to the long brass instruments where they are close enough even to subdivide a low register octave.

In open acoustical lines another characteristic is that with high frequencies the resonances tend to disappear. This is because the radiation impedance rises with frequency up to the point where the wavelength approximates the tube circumference. From then on the radiation load essentially matches the line impedance so there is little reflection back into the line.

The cylindrical clarinet uses the fundamental quarter wave resonance for its main register, and the high overblown register is located at the next resonance, one octave and a fifth, that is three times higher frequency. To fill the chromatic scale between those the tube is perforated at many places so that it can be effectively shortened. Such an opening short circuits the propagating wave and will effectively disconnect the downstream part of the tube. You need some 18 steps on this journey with grip holes basically covering 2/3 of the length of the instrument. To make it possible for a performer with only two hands to manage this Boehm invented his famous mechanism with numerous link operated cover pads.

When we come to the more complicated situation of a tube with area varying along its length the classical approach is to use an additional term in the wave equation which then is called the horn equation. This was developed by several workers in the 19th century but became widely known only in 1919 when treated by Webster. A primary interest at that time was to find out how to make loudspeaker horns. The horn equation can be used analytically on some simple shapes like the conical and exponential, but dealing with more general shapes this becomes too difficult and you have to tackle it numerically. Another approach is the one Fant (1960) took in the 1950s to handle the vocal tract transmission. He set up models with a few cascaded line segments. Each segment was a straight tube that could be analytically treated as well as the few impedance mismatches at the joints.

A continuation in spirit of the computer age is to subdivide the line into a large number of small uniform segments of equal length but differing areas. This offers a possibility to simulate the line by two sequences of delay units, one for forward waves and one for backward, and where the delay time step corresponds to the wave travel time through one segment. Between the steps you insert scatterers, elements to take care of the transmission and reflection of the partial waves in accordance to the impedance mismatch at each joint.

I will not go inside the related very extensive modem theories and methods for numerical signal processing. I will just mention that there exist simple conversions from the exemplified bidirectional line into a canonical digtal filter and the other way round. The filter model can also be inverted. Then it can be used as a research tool to inverse filter a natural signal, that is, take away the signal features imposed by the production filter, and retrieve the source signal. In the seventies Markel, Gray (1976) and others developed the theory of linear prediction along these principles such that you can compute filter coefficients from samples of the sound wave. This field is still maturing and expanding, nowadays under the name of system estimation.

The bending waves in plates and shells are difficult to treat because the stiffness measure to determine wave speed not only depends on the material, but also on its thickness and, even worse, on the wavelength. Thus the waves have dispersion, speed depends of frequency. The associated wave equation has a higher order spatially, and the solutions for cases simple enough to be analytically treated hold Bessel functions rather than the simple circular sines and cosines. An immediate consequence is that the resonance frequencies are never harmonic unless you make them such by control of thickness and shape.

Transient and continuous

It is important to recognize that the resonance frequencies are are the ones at which the system oscillates after it is excited, but then left to itself, that is when there is no further source signal. From the viewpoint of source time duration it is now pertinent to divide the musical instruments into transient and continuous. Typical transient sources are with plucked and beaten strings like the guitar, the piano and the percussion instruments, while the bowed strings and the wind instruments, including the voice, have a continuous source.

In the transient instruments the prime source property is its magnitude and the width of its spectrum which is determined by the hardness of the hammer. This tells which resonances can be excited at all. In some instruments you can however select the place of excitation to be at the nodes of one or more resonance modes, so that these modes are essentially not excited.

Now once the filter has been excited it goes on oscillating at its resonance frequencies. The point is that these are defined by the filter, not the source. In the case of strings they indeed are harmonic except for the small deviations caused by bending stresses. It appears to be very general that to be musically attractive and to give clear sensations of pitch also the transient instruments should have a close to harmonic distribution of their resonances. The means to get there is to control the wave speed and impedance throughout the resonating body. The details on where to do what are complicated and delicate and generally constitute the deeper trade secrets in fabricating the instruments. Well known examples are the shaping of a horn bell, the variation in membrane thickness in certain drums, the tuning of metallophone bars or church bells by selectively cutting material away, not to speak of the magic blacksmith work in making a gong or a steelpan.

In the contrasting class of instruments with a continous periodic source the harmonic structure is forced on the system by the source. For a stationary situation we can view each of the source harmonics modified by the corresponding value of the filter transfer function, just as in speech. Still one often wants the filter to have its resonance peaks close to the harmonics in order to enhance them as much as possible to produce a rich spectrum. As an example fit a stiff steel string to a violin. When you bow it all harmonics are strict multiples of the fundamental. But because of the string inharmonicity the high harmonics are out of tune with the string filter resonances. So they will not build up to their expected level, the resulting sound is effectively lowpass filtered and becomes dull.

But for music to be musical it cannot be very stationary. So the transients that always accompany a change in the source will excite the filter resonances, harmonic or not. These transients at the onset of a note are highly characteristic of an instrument and are often difficult to implement satisfactorily in synthesizing its sound. Many sources also contain a noise component, quite conspicuous for instance in the flutes. The noise has no harmonic components but instead gets a direct coloring from the filter transfer function. These spectral peaks need not coincide with the harmonics of the periodic signal.

In string instruments the filter between the string and the ambient air is one or several sound boards combined with resonator cavities. This gets complex enough to make it impossible to do an elementary analysis. The speed of the bending waves changes over the area of the board due to its varying thickness or to the presence of stiffening bars and even if the board is isotropic the bending wave speed also depends of frequency. The boundaries of the board do not have a simple shape and the boundary conditions may vary unsystematically, clamped at some places, free to bend at others, etc. Finally the radiation to the ambient is distributed over the entire board, so also the air load and the directional characteristics become very complicated. Characteristics like input impedance and transfer of this kind of filter is conveniently studied with modern tools like a dual channel Fourier analyzer. The two channels may then represent the source and the filter output derived from an accelerometer on an exciting hammer and a microphone. You can get results of this kind, here the transfer from a driving point of the string, that is the bridge, to ambient sound pressure. These two are quite like each other although the instruments differ somewhat in design and size, it is a violin and a piano.

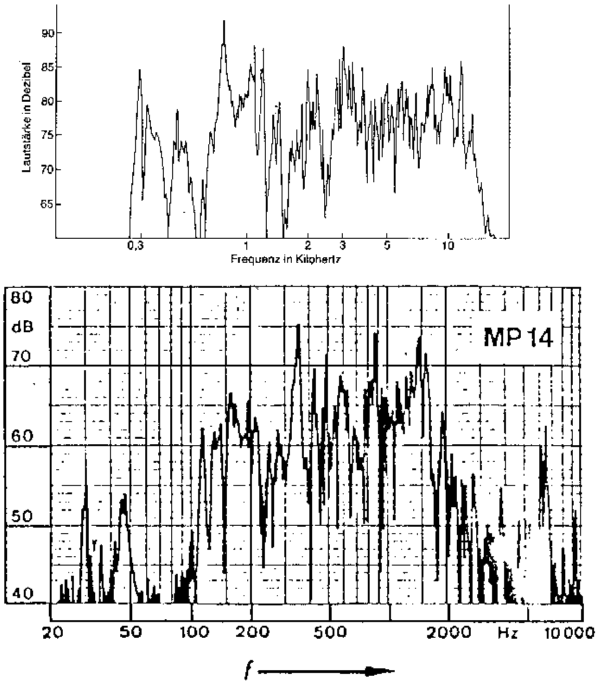

Fig 1. Transfer functions from bridge to air for top a violin (Hutchins (1981)), and bottom a piano (Wogram (1981)). The graphs are displaced in frequency by a factor of 5 to approximately normalize for the difference in soundboard size.

Area perturbation

The standing wave pressure and flow (or force and velocity) directly link the potential and kinetic parts of the wave energy. In the sixties Manfred Schroeder (1967) used an arcane theorem by Ehrenfest to create what we know as perturbation theory. This theorem tells the relative change in a resonance frequency coming from a small local change in impedance. The sign of the frequency change is related to whether the static or kinetic energy dominates at the point of change. The meaning of this has the same character as these curves, usable as a guide when the resonances of a church bell are to be tuned in a lathe. But the instrument maker would rather be interested in a resolved answer, namely the perturbation theory which tells what kind of change in shape that is required to move one, and only one of the resonance frequencies.

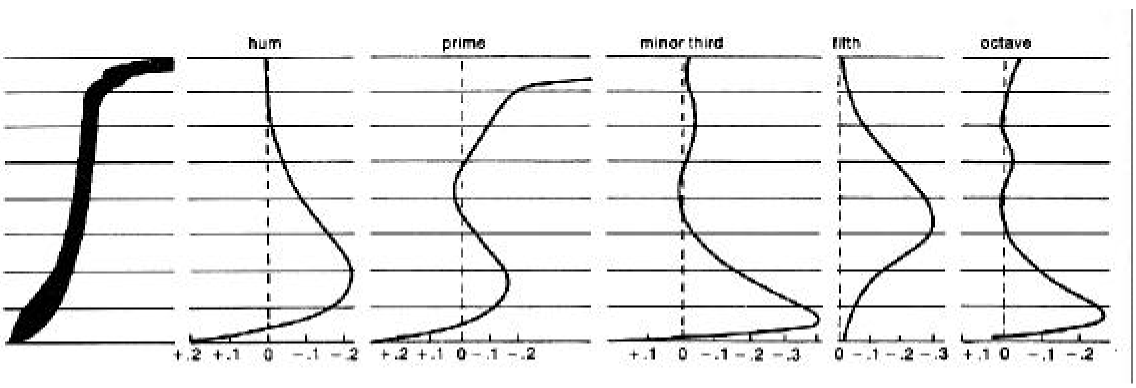

Fig. 2. Bell tuning curves show how the first five partials change in frequency when metal is removed from various inner locations on the inside surface, which appears in cross section at the left. (van Heuven, from Rossing (1988)).

For the simple uniform tube case it comes out that we have to modify the area in cosinusoidal shape having half the wavelength of the standing acoustic wave. Let for instance the diameter get smaller at all pressure maxima where the potential energy dominates and larger at all pressure nodes where the kinetic energy dominates. Then both measures tend to raise the resonance frequency. This perturbation is orthogonal to all the other resonances, so these remain unchanged.

Let us try to perturb a straight tube as to shift all but the fundamental resonances upwards to make them come at the places for a closed tube. The second resonance should be moved up 1/3, the third 1/5, and so on. We might then guess our total area function could be a sum of spatial cosines, amplitudes falling correspondingly with frequency. Such a sum, using the wanted integer spaced frequencies, indeed does look like a trumpet when you take the square root to see it as diameter rather than area! By a decrease in the lowest nonzero frequency cosine amplitude the narrow part can come out as almost cylindrical rather than flared. This also matches the fact that the fundamental resonance is out of tune and left unused in the brasses.

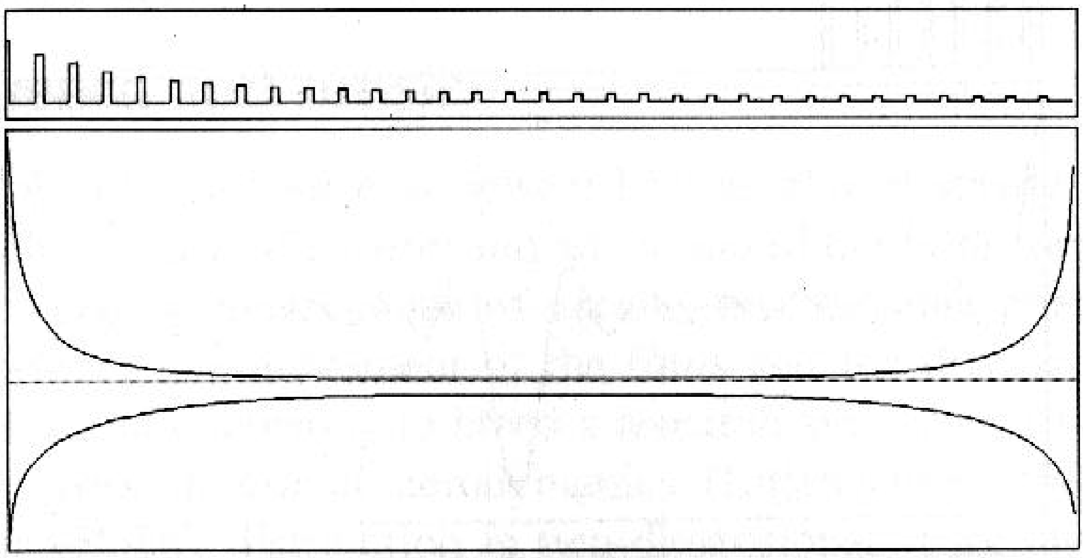

Fig. 3. A trumpet-like shape as generated by a cosine series, motivated by perturbation theory reasoning. Top the spatial frequency spectrum, bottom the shape viewed as area and diameter (square root of area).

If we instead perturb substantially to raise also the fundamental resonance it is easy to similarly approach the conical shape typical to the saxophone. In this and other conical instruments the fundamental is used and the overblown register at next resonance comes only one octave higher.

The input resistance to the filter has superficially the same shape as the transfer function, only scaled with the radiation resistance. This is not self-evident or universal, but provided the internal filter losses are small it can be inferred as the driving power we expend at the filter input has no other way to escape than into sound radiation. For an archetype we can represent the impedance with pulses, located at the usable resonances along the frequency axis. They are some ten in number with the brasses, only a few with the woodwinds.

Let us now Fourier transform this schematic impedance into an impulse response at the mouthpiece. This then represents the pressure at the mouthpiece as a function of time, following the input of a unit area flow impulse. The remarkable consequence of our moving the resonances into contiguous integer multiples is that the reflected pulse now arrives in phase after travelling once fore and back in the tube. The reshaping of the uniform tube into a trumpet has made the input look as if also the remote end is closed what regards nonzero frequencies. This also explains why the instrument is not appreciably detuned when you insert a mute into the bell: the end wall of the mute is located approximately at this virtual closure. Hence the resonance frequency of the mute itself as a Helmholtz resonator is not important to the tuning, indeed this resonance normally falls right in the middle of the operating frequency range. A main function of the mute is to reduce the mouth radiating area which decreases the low frequency radiation appreciably.

So far we have looked only at one-dimensional waves. This is not very accurate when a horn flares significantly and the wavefront becomes curved. To cope with this using analytical mathematics tends to get rather complicated, one might remember heroic work by Benade and Jansson (1974) in the seventies to bring the theory forward for brass instruments.

For a general case in more than one dimension we will have to resort to numerical methods. The space can for instance be subdivided into a number of finite

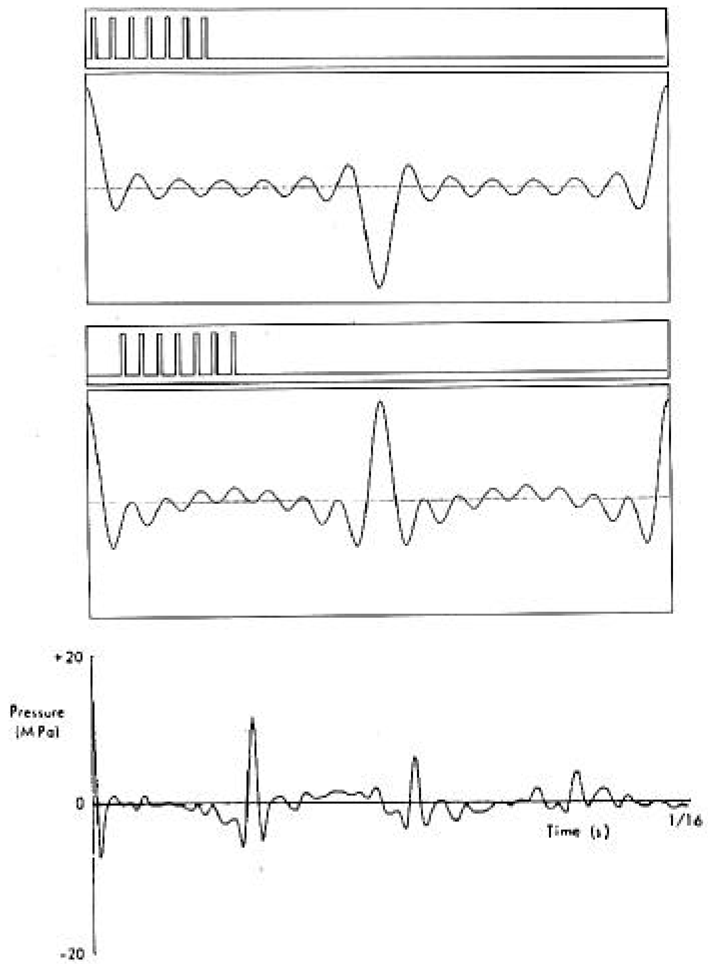

Fig 4. Idealized input impedances and their corresponding impulse responses. Top for a cylindrical tube, closed at one end. Middle for a trumpet-like tube where the second and higher resonances have been shifted to be at integer multiple frequencies. Bottom impulse response for a trombone as measured by Elliot and Bowsher (1982).

elements, each with simple differential expressions defining the relations between time, place, pressure and flow. The whole space is subsequently brought together and mapped by the simultaneous solution of a massive equation system involving all the elements. The finite element method emerged in structural mechanics and is directly applicable to the complex movements and tensions in soundboards. For the somewhat easier problem of acoustical radiation in a homogeneous space the boundary element method is now emerging an attractive alternative. Then we need not subdivide the space since nothing very special happens there except the mere radiation transport. We can limit ourselves to the bounding surfaces, which is very beneficial to the volume of computation, and the interior field can be fairly straight forward inferred from what goes on at the boundaries.

Sources in speech and music

The vocal fold speech source is powered by an almost steady lung pressure. By delicate muscular control of tension and adduction of the folds they can be made to vibrate. This comes as consequence of various aerodynamic phenomena at the narrow glottal passage. The movement of the folds and the driving mechanism is sufficiently complex and interesting to breed a research area of its own. This picture from my current work in glottal aerodynamics (Liljencrants (1989)) shows an example of the air velocity distribution in two-dimensional space around the vocal folds in a simple model. Both in speech and wind instruments the driving pressures inserted are so high that the corresponding kinetic air velocities are in the range well over 10 m/s. One consequence of this is that when the flow passes a significant constriction you almost always get flow separation from the wall, the flow forms a jet. Mostly also the Reynolds numbers are high enough that there comes instability and turbulence downstream the jet, you see it here as it is developing into a vortex street. Such instability makes some noise and is important in the labial pipe as an initial seed to start oscillation.

The brass wind instruments are the ones most closely like the speech apparatus. The vocal folds have their counterpart in the player's vibrating lips. These have a less sophisticated anatomical structure which calls for exercise to make them vibrate properly. Common to the vocal folds and the oscillating lips is that they are comparatively weakly influenced by the sound pressure in the following filter since the downstream sides of their vibrating members work against it only at small areas. A major part of the forces rather come from the aerodynamics inside the passage, but these in turn do depend on the acoustic pressure which is a component in the total pressure drop across the passage.

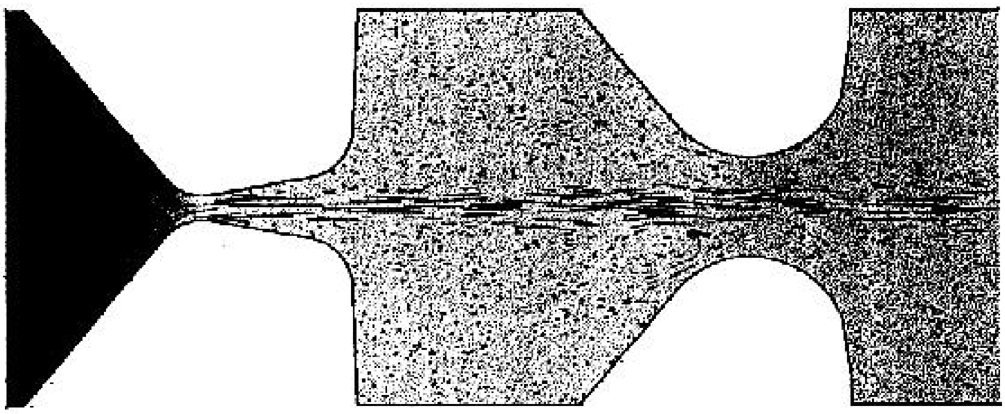

Fig. 5. Simulation of two-dimensional flow between the vocal folds and the following false folds, a layout similar to the one in a brass instrument mouthpiece. Pressure is shown by the gray level and flow as particles with velocity vectors. It is seen how the air jet develops instability.

The reed source is to some extent similar, but an important difference is that the entire inner surface of the reed is exposed to the acoustic pressure in the tube. This makes the source flow very strongly influenced by this pressure and it becomes easy to view the reed mechanism as a negative resistance. An increase in acoustic pressure tends to open the reed slit and allow an increased flow in direction against that pressure.

In the labial instruments like the flutes and organ pipes the source flow is also very much controlled by the filter. Here a thin band of air is blown across the aperture by the power supply and is split by the labium edge. This band has essentially no compliance and mass to oppose the resonant pressure or flow in the filter. But when there is a flow into the filter across the jet this will deflect the jet toward the inside, and the flow is thus incremented with the jet flow. This is in turn converted into pressure as the jet slows down inside the tube and part of its kinetic energy is converted into potential. It can then be regarded as a negative admittance; a flow into the filter is accompanied by an amplifying pressure change. As compared to our earlier source examples this one has a very low impedance level, when connected to a tube the tube practically remains open ended when we regard it as a resonator.

Pitch to formants relation

A prominent difference between spoken voice and the musical instruments is the pitch in relation to the filter resonance frequencies. The low pitch in speech makes the densely spaced harmonics more or less completely fill the envelope of the transfer function and there is no special frequency relation between the pitch and the resonances. Crudely neglecting prosody, the information is conveyed by the movements of the formant resonances. The pitch and harmonics just serve as a carrier to fill the spectrum to make all formants audible. A time domain correspondence is that the pitch period is long compared to the sound travel time through the resonator. Indeed the vocal tract may be viewed as a transient instrument. It is primarily excited at the closure of the vocal folds after which the sound reverberates several times in the mouth cavity. When the vocal folds open for the next cycle the reverberation is heavily damped until the next excitation. This is similar to the alternate action of hammer and damper when you play a rapid series of notes on the piano. It also has a similar function, namely to suppress excessive reverberation that would have smeared out the sound timewise to conceal the 'message'.

Now interaction between source and filter comes to issue, in a first place since the source-filter concept assumes time-invariant linear conditions. Still the concept is useful, but we will have to introduce unorthodox connections to represent that some signals control element values or other signals in the system. A simple way to cope is to let certain elements vary slowly according to some smoothed function of the signal state. An example may be to assume incremented formant bandwidths in proportion to some average loss in the glottal passage. A more complicated way is to let the control go directly without smoothing. For instance to model that both the frequency and bandwidths of the formant oscillations change dynamically with the momentary value, of the glottal impedance. Then the model becomes nonlinear which among other things will open the possibility for chaotic behaviour, perhaps a candidate to explain jitter and shimmer in the voice.

In the instruments an important aspect of the interaction is that the filter signal to an essential degree controls the source pitch. In the labial pipe this is obvious since the source mechanism directly depends on the filter signal. In the reeds there is the strong coupling to the inner surface of the reed from the acoustic pressure in the tube. The lesser coupling to the heavy lip valves in the brasses is made up for by a correspondingly high filter impedance, that is narrow bore in the tube, which also corresponds to a very high internal sound level. It is illustrative to interchange the mouthpieces of a clarinet and a trumpet. With the trumpet mouthpiece on the low impedance clarinet there is little interaction, so the player can easily vary the pitch with his lips independent of the fingering and tuning of the filter, just like in speech. Oppositely, with the clarinet mouthpiece on a trumpet the pitch is securely locked to the fundamental 'pedal' note and it is hard to get any other unless you open a water valve to damp out this resonance.

Thus in the instruments the pitch coincides with a resonance, usually one of the lower, and the higher resonances are left to color the harmonics of the pitch. One could twist this into that the roles of pitch and spectrum are interchanged, the pitch conveys the information and the spectrum becomes a prosodic element. The singing voice is a bridging phenomenon. To make the first formant frequency coincide with pitch is much used to render a high output sound level, but because of the weak coupling this is not automatic, instead it is one of the singer's acquired skills.

References

Benade, A. H., Jansson, E. (1974): On plane and spherical waves in horns with non-uniform flare. Acustica 31, pp. 79-98.

Elliot, S. J., Bowsher, J. (1982): Input and transfer response of brass wind instruments. JASA 72-6, pp. 1447-1460.

Fant, G. (I960): Acoustic theory of speech production. Mouton, S'Gravenage. 2nd ed. 1970.

Hutchins, C. M. (1981): The acoustics of violin plates. Sci. Am. 245-4 (Oct.), pp. 126-135.

Liljencrants, J., Fant, G. (1975): Computer program for VT-resonance calculations. STL QPSR 4/1975, pp. 15-21.

Liljencrants, J. (1989): Numeric simulation of glottal flow. Proc. Congr. of Vocal Fold Physiology. Stockholm.

Markel, J. D., Gray, A. H. (1976): Linear prediction of speech. Springer.

Rossing, T. D. (1984): Acoustics of bells. American Scientist, 11, pp 440-47, reprinted in Rossing (1988).

Rossing, T. D., editor (1988): Musical acoustics. Am. Assoc. of Physics Teachers.

Schroeder, M. R. (1967): Determination of the geometry of the human vocal tract by acoustic measurements. JASA 41, pp. 1002-1010.

Wogram, K. (1981): Acoustical research on pianos. Das Musikinstrument 24, reprinted in Rossing (1988).