SPEECH AND MUSIC: ACOUSTICS, SIGNALS AND THE RELATION BETWEEN THEM

Joe Wolfe

School of Physics, The University of New South Wales, Sydney

ABSTRACT

The codings of speech and music are different and in some ways complementary. The voice operates on acoustical principles distinctly different from those of (other) musical instruments. This paper explains how and poses the question: are the different codings the result of the different acoustics, or *vice versa*? What if the instruments came first? This paper develops the conjecture that the pitch stability of notes so important in Western (and other) music may have come first, not from older, unaccompanied song, but from imitation of or performing with musical instruments.

1. INTRODUCTION

Musical instruments are poor at producing speech, and not all voices are good for music. Why is it so, and might it tell us something important about music?

Music and speech have much in common and many researchers have looked for a common origin and other relations between them. Others, including Peretz [1], have stressed the biological differences. Elsewhere, I have described the complementarities and symmetries in the coding of music and speech and used these to frame conjectures about the possible origin of music [2]. Here I discuss the very different acoustical principles that give rise to some of the differences.

The music of many cultures depends on notes whose pitch is stable and independent of loudness. The stability facilitates perception that may be both precise and categorical. Variation of loudness at constant pitch is important in expression and phrasing.

Further, these notes have almost periodic waveforms and therefore almost harmonic spectra. This is important both to harmony and to the precision of pitch perception. Non-vocal musical instruments produce stable, periodic waves using physical phenomena that are not used by the voice and that are rare in the natural world, i.e. rare except in objects made by people.

Phonetic information in speech is largely carried by the envelope of the spectrum: how much energy is carried in a particular frequency band. Speech uses both broadband sources (roughly corresponding to components of consonants and to whispers) and a swept frequency narrow band source (vowels and consonants) to convey this information. Stable, harmonic notes carry much less spectral information. In much music, the broadband component is of relatively minor importance, and extensive portamento (swept frequency) is relatively rare.

To discuss these ideas briefly, broad generalisations are necessary. First, however, let me mention a number of provisos and exceptions. I am not discussing the rhythmic aspect of music, nor most percussion instruments. Non-periodic sounds can be used in (different) harmony systems, such as gamelans and synthetic systems. Although categorical tone perception is used in speech, especially in tonal languages, the precision and number of categories are small. Consequently, relatively little information is carried thus, as is demonstrated by the observation that whispered Mandarin can usually be understood. Extended portamento is idiomatic to the theramin, the saw etc and easy (but used sparingly) on trombone, violin etc. Broadband components are important in identifying many instruments. Let us now return to the broad argument.

Instruments have evolved (in the engineering sense) to suit music, and some music, too, has evolved to suit instruments. Speech has evolved to suit the voice and possibly the voice has evolved to suit speech. Of course, the trained voice can produce both. This raises the interesting questions: *Might the stable pitch aspect of song have developed to imitate or to perform with instruments*? One might alternatively ask whether some instruments have been chosen or developed, in part, to imitate the melodic aspects of song or whether they might have arisen independently and converged. Here, however, let us discuss the more interesting conjecture that the instruments came first.

To discuss this, let's begin with some acoustics.

2. ACOUSTICAL BACKGROUND

Linear Oscillations

In studying vibration, the relation between force F or pressure p and displacement x is critical. In a linear system, the change in F or p is proportional to the change in x. Linear systems have the feature of superposition, with the consequence that the total output is simply related to the sum of the inputs: the inputs do not 'interact' with each other to produce new outputs. An elastic object such as a spring (or a string, the body of a violin, or the air that carries sound) is linear for small vibrations: double the force and you double the deformation.

High Q and low Q. In linear oscillators, the quality factor, more commonly just called Q, is defined as 2π times the ratio of the energy stored in an oscillating system to the energy lost in one vibration. A low friction pendulum swinging in air has a high Q: it swings many times before most of its mechanical energy is lost. The same pendulum swinging in a liquid has a lower Q. A

high Q resonator (pendulum, string, drum membrane, body of air, section of wood) has a resonant frequency f_0 and a narrow bandwidth Δf: it responds only to a very narrow band of frequencies near f_0, whereas a low Q resonator responds more weakly but over a broader range of frequencies. (In tuned musical instruments and the voice, the energy radiated as sound is small compared with thermal, viscous and other losses, so a low Q does not usually result from producing a lot of sound.)

Multiple resonances: harmonic and non-harmonic. Many extended elastic objects (strings, membranes, violin bodies, the air in a flute or the vocal tract) resonate at several different frequencies. These examples are all (very nearly) linear, because of the elastic behaviour of their components and the fact that the pressures and movements associated with sound are small.

In a very small subset of cases, rarely found in nature, the resonances fall in a harmonic series, *i.e.* at frequencies f_0, $2f_0$, $3f_0$ … nf_0 , n an integer. In music, an important case is the uniform, flexible string between two fixed supports. Because its resonances are harmonic, it readily supports vibrations with periods $T_1 = 1/f_0$, $T_2 = 1/2f_0$… $T_n = 1/nf_0$. Consequently, a plucked string on an instrument produces a sound that is almost exactly periodic (with period T_1) and has an almost harmonic spectrum. Tap the elastic body of the violin, and the resulting sound is inharmonic and has no clear pitch. Pluck its elastic string and the sound is almost exactly periodic and harmonic. The harmonicity, however, is sensitive to the uniformity, flexibility and support conditions: even some piano strings may be noticeably inharmonic. Consequently, systems that produce periodic or harmonic sounds in this manner are rarely found in nature.

Harmonic resonances are also produced by cylindrical or conical air columns, with further important consequences in musical instruments. Other shapes (e.g. the bores of brass instruments, bells and some other tuned percussion) can have harmonic resonances if suitably designed or evolved [3].

Nonlinear Oscillations

Friction gives a familiar example of strong nonlinearity: push the refrigerator and at first there is no displacement. Beyond a threshold, it ‘goes with a rush’. So, although the violin strings and body are inherently linear, the friction in the bow-string interaction makes that interaction strongly nonlinear. Other nonlinear oscillators are the reeds of woodwind and the air jets of the flute family.

The vocal folds are key elements in a highly nonlinear oscillator, whose operation shares some features with the lips of a player of brass instruments. The vocal folds have some linear properties – in physical models, each of the vocal folds is often treated as one or more mass-and-spring oscillators. Nonlinearity in their oscillation, however, comes from (at least) two sources. One is the loss, in turbulence, of kinetic energy of air passing between them, which produces a pressure difference proportional to the square of the flow velocity. A stronger nonlinearity arises when the folds collide with each other.

If parameters such as the average sub-glottal pressure and muscular tension are maintained constant, and if they fall in a certain range, then the vibration of the vocal folds is periodic (mechanisms 1 to 3; chest, head or flageolet voice). For other parameter ranges, the vibration is irregular (mechanism 0 or croak voice, chaotic screams and children's cries etc). Nonperiodic vibration is much rarer in singing than in speech and other utterances.

When periodic motion is produced, the nonlinearity gives rise to nonsinusoidal vibrations: in other words they have harmonics, and the strength of the higher harmonics depends on the strength of the nonlinearity and on vibration amplitude. In contrast, the superposition possible in linear systems allows sums of vibrations at any frequencies.

Thus the periodic vibration of a bowed string and a plucked (harmonic) string come from quite different phenomena: the nonlinear bow-string interaction (with constant control parameters) produces the periodic motion. In the plucked string, the periodicity only occurs if and because the resonances are harmonic. A string whose resonances are inharmonic (for instance due to nonuniform wear or accumulation of mass, or to finite bending stiffness) is aperiodic when plucked, but may be bowed to produce a periodic sound (and hence harmonic spectrum).

3. THE VOICE *vs.* INSTRUMENTS

Almost all acoustic musical instruments have high Q, highly linear resonators that determine the playing frequency[1]. The voice does not.

Instruments. In plucked strings (and in many percussion), the playing frequency is determined by the linear resonator alone. In contrast, the instruments that can produce sustained notes have a nonlinear mechanism. However, in instruments but not the voice, the pitch is determined by a resonator.

For example, the bow-string contact produces nonlinear oscillation, but (over a limited range of parameters [4]) its pitch is governed by the resonances of the string. The nonlinear vibrations of flute air jets, reeds in woodwinds and lips of brass players are controlled by the resonances of the air column (squeaks and altissimo ranges sometimes excepted).

Further, in most of these instruments, the parameters that determine the frequency are easily held constant. These features allow the production of a sustained note with a frequency largely independent of loudness, without compensating adjustment of those parameters.

The voice. The vocal tract is a highly linear, waveguide resonator with a moderate value of Q, but it does not

[1] Free reed instruments such as the accordion, sheng, bawu etc do not neatly fit this scheme. However, the pitch is chiefly determined by the mechanical resonance of the reed, which again holds the pitch constant during a *crescendo*. The steady pitch cooing of doves and the songs of some frogs do have a linear, elastic system determining pitch.

control the pitch of the voice[2]. To hold a constant pitch in a strong crescendo and decrescendo requires considerable adjustment of the parameters of the vocal folds, which is why a *messa di voce* is difficult and is often an important part of a singer's training.

Harmonic resonances. Most tuned instruments have a series of resonances that fall in harmonic or nearly harmonic ratios. This means that even linear instruments, such as plucked strings, bells and some drums, can produce complex sounds with nearly harmonic frequency components. For nonlinear instruments, automatic coincidence of higher harmonics and higher resonances means more stability of the pitch and higher power in the high harmonics. There is no similar phenomenon in the voice[2].

Broad band components. A further and very important difference is this: In music, broadband sources, which do not have a pitch, have a secondary role (examples are components of the starting transients of many instruments, the breath sound in wind instruments, and part of the sound of untuned percussion) [2]. Where the envelope of the broadband spectrum is variable, it is usually not independent of the harmonic components. These features and the difficulty in controlling them (and sometimes the limited capacity for *portamento*) make instruments poor at speech.

Acoustic instrumental music sounds unnatural without broad band components, but they make little difference to recognition of melody or harmony. In speech, in contrast, broadband sources are important in most phonemes and vital to comprehensibility. Further, whispering shows that speech (even in tonal languages) can be understood with only broadband signals.

Pitch stability. The most important difference, however, is pitch control by the resonator. Although a strict *messa di voce* is difficult on wind instruments, the pitch change produced in the absence of regulation of the parameters of the nonlinear components is rather smaller than it would be for the voice. Consequently, playing a sequence of notes with pitches almost independent of loudness requires relatively simple and almost independent adjustment of parameters.

Digital pitch control. Further, many instruments have keys, tone holes, valves or frets that give nearly digital control of pitch. Examples of these include Paleolithic flutes, fragments of which are currently among our earliest indications of artificial instruments [7]. Instruments with continuous pitch, such as violin and theramin, are often judged difficult to learn.

Learning singing. To be able to sing in tune and to control pitch and loudness independently, one has to learn to control parameters of the vocal folds and the subglottal average pressure in subtle combination: changing pitch at the same loudness (or *vice versa*) requires modification of several parameters. Further, one needs a rather precise 'muscle memory' of the parameter values required for entries and for changes. Fortunately, we have plenty of practice.

Do other species learn thus? Most bird and whale song has extensive portamento and little stable pitch. This does not prevent our categorical perception (or composers' stylisations) from mapping the songs into discrete pitches, of course! Further, strong variations of loudness at constant pitch are very rare. Some species with fixed pitch, such as cicadas, produce it by an entirely separate mechanism, i.e. stridulation, which is not relevant to the voice or to instruments.

Learning instruments. Compared with singing, playing in tune and controlling pitch and loudness independently would seem to require less complicated control on nonlinear instruments with resonator control (e.g. violin, trumpet), where relatively fine adjustments are required to counter the dependence of pitch on loudness. It is much easier on the linear, digital instruments (e.g. guitar), though of course these instruments have other difficulties.

Does this mean that singing is difficult, that the voice is hard to 'play'? That question is obscured by our long familiarity and very early exposure to its use. Nevertheless, I think an argument may be made that artificial instruments are, in a sense, better suited for performing music in tune, all else equal.

4. LISTENING

In the human head, music and speech are detected and processed by the same hardware and (at least) some similar low-level processing. Nevertheless, there are considerable differences in processing. For instance, I find it a difficult mental exercise to judge the pitch range of a sentence.

A strong example comes from cochlear implants (CIs). Users who achieve excellent scores in speech or even in isolated word recognition fare poorly in melody recognition and report no sense of harmony. Most CIs are configured to provide stimulation at about 20 different places in the cochlea, but they discard most of the time and frequency information that is associated with pitch and harmony. Although the reported sensation of pitch by CI users depends on both rate and place of stimulation in the cochlea, harmony and the precise perception of pitch seem impossible without detailed temporal information [8,9].

Detection of vibration rate (rather than place), which we need to comprehend music, could be considered as an extra complication in the auditory system: it seems to require a qualitatively different mechanism (something like the neural equivalent of autocorrelation [*e.g.* 10]).

So, if one can understand speech and recognise most ambient sounds without rate perception, why do we have an elaborate system of pitch detection whose performance

[2] In some ranges and styles, tract resonances fall close to one of the harmonics of the vocal folds [5,6], and singing is easier in this condition, but the resonances are tuned to the harmonic, not *vice versa*. I contend that the definition of singing does not include whistling or the high component of overtone singing.

approaches the theoretical (Fourier) limit[3]? The simple answer may be that this system is useful in distinguishing low level, periodic sounds in a noisy environment. This suggested answer is consistent with the observation that this task is difficult for users of CIs, who usually have only very limited frequency information.

5. DISCUSSION

The preceding section has given a brief answer to my first question: Why is it so? The other questions take us beyond physics, and beyond the professional competence of this author. So I shall extend the list of questions and leave the explicit answering to others.

Are there no (acoustic) instruments that are good at speech? A wa-wa muted trumpet can say 'wa wa' and some other vowel combinations. Purely acoustic speech synthesisers have a 200 year history [11,12]. These are acoustic models of the voice and so, although designed for speech, some are capable of song. Their purpose is research on the voice: they have not become popular as musical instruments.

Robot speakers apart, do we have instruments without resonances to control the pitch? The resonances of a megaphone are not strong enough to control the lips of a 'player's' lips, the way a trumpet does. The acoustic resonances of duck call do not control its pitch. So we don't consider either a musical instrument or even capable of music. An instrument whose pitch were completely variable and dependent on amplitude would probably take longer to master than the theramin and so may never have become or has not remained popular.

Have there been such instruments in the past? Would we recognise them as instruments? Would they last? A simple double reed, with no resonator, might not last for, nor attract the attention of, the archæologist. It is the tone holes in the Paleolithic flutes that have identified them.

Might the stable pitch aspect of song have developed to imitate instruments or to perform with them? If that were the case, then one would expect very ancient cultures without tuned instruments to have song with relatively little constant pitch. Some examples exist, but I leave this expectation be confirmed or refuted by a quantitative ethnomusicological study.

Until it is, let us imagine a people with a portamento singing style that discovers instruments with stable pitch under digital control – flutes or reed instruments. These instruments evolve and perhaps their pitch becomes more stable. The makers learn to imitate or to accompany these sounds with their voices.

But what is the advantage in producing and perceiving stable pitch? I have suggested that it might be a game to learn aspects of hearing perception [2]. Humming or wordlessly singing to an infant presents an auditory task in which several parameters are held constant – a common pedagogical approach. This might give the children of singing parents a small competitive advantage. A substantial advantage is not necessary for preservation of a trait: music might be, to some extent (and as many musicians hope), an object for sexual selection – an auditory and mental peacock's tail.

I have argued that part of music's attraction and power could lie in the complementarity codings in speech and music [2]. This paper has set out the acoustical differences that lead to much of that complementarity.

[3] In a note lasting T seconds, the minimum uncertainty in frequency cannot be much less than (1/T) Hz.

6. REFERENCES

1. Peretz, I. "The Nature of Music from a Biological Perspective", Cognition, Vol. 100, 2006, 1-32.
2. Wolfe, J. "Speech and music, acoustics and coding, and what music might be 'for'" *Proc. 7th Intl. Conf. Music Perception and Cognition, Sydney,* K. Stevens, D. Burnham, G. McPherson, E. Schubert, J. Renwick, eds. 2002, 10-13.
3. Fletcher, N. H., and Rossing, T. D. The Physics of Musical Instruments, Springer-Verlag New York, 1998.
4. Schelleng, J.C. "The bowed string and the player" J. Acoust. Soc. Amer., Vol 53, 1973, 26-41.
5. Joliveau, E., Smith, J., and Wolfe, J. "Tuning of vocal tract resonances by sopranos", Nature, Vol 427, 2004, 116.
6. Henrich, N., Kiek, M., Smith,. J., and Wolfe, J. "Resonance strategies in Bulgarian women's singing", Logopedics Phoniatrics Vocology, in press.
7. Dauvois, M., Boutillon, X., Fabre, B. and Verge, M.P. "Son et musique au paelólithique", Pour la Science, Vol 253, 1998, 52-58.
8. Fearn, R., and Wolfe, J. "The relative importance of rate and place: experiments using pitch scaling techniques with cochlear implantees" Annals of Otology, Rhinology and Laryngology, Vol 109, 2000, 51-53.
9. Fearn, R. "Music and pitch perception of cochlear implant recipients" PhD thesis, Univ. New South Wales, Sydney, 2001.
10. Meddis, R., and O'Mard, L. P. "Virtual pitch in a computational physiological model", J Acoust Soc Am. Vol. 22, 2006, 3861-69.
11. Dudley, H., and Tarnoczy, T.H. "The speaking machine of Wolfgang von Kempelen", J. Acoust. Soc. Amer., Vol 22, 1950, 151-166.
12. www.takanishi.mech.waseda.ac.jp/research/voice/