Stable, quantised pitch in singing and instrumental music: signals, acoustics and possible origins

Joe Wolfe (1) and Emery Schubert (2)

(1) School of Physics, The University of New South Wales, Sydney, 2052
(2) School of the Arts and Media, The University of New South Wales, Sydney, 2052

ABSTRACT

Unlike most artificial instruments, the voice usually has no resonator to stabilise the pitch. Singing pitch depends on geometry and muscle tension in the larynx but also, strongly, on the sub-glottal pressure. Consequently, pitch and loudness of the voice are strongly correlated, which is why a *messa di voce*—a gradual increase and decrease in loudness at constant pitch—remains a difficult exercise. Why do we sing in a style that is suited to musical instruments, but arguably much less suited to the voice itself? This paper discusses the advantages of digitised pitch in musicial signals for storage, processing and harmony. It then uses acoustical and musicological arguments to support the hypothesis that styles of singing with stable, categorical pitch, which is controlled independently of loudness, may have evolved since and because of the development of artificial musical instruments. Because stable pitch instruments are at least tens of thousands of years old, and probably much older, it is possible that this influence on song is similarly ancient. We argue that, through the generational transmission of memes, the mimicking of artificial instruments may have given rise to the 'unnatural' fixed pitch singing, which consequently became the one of the dominant styles in Western and other musics.

THE VOICE *vs*. OTHER MUSICAL INSTRUMENTS

Detailed written analyses of musical tunings suggest that, since classical Greece, instrumental and often vocal music has used discrete pitch intervals: what the psychologist would call categorically perceived intervals, and the physicist might call quantised intervals. For at least several centuries, Western singing has required stable, categorical pitch, independent of loudness.

The human voice is qualitatively different from most other musical instruments: artificial musical instruments almost always have a resonator that stabilises and largely determines the pitch (e.g. Fletcher and Rossing, 1997). This paper will explain how music that requires independent control of pich and loudness, while suited to artificial instruments, is less suited to the voice. We then make suggestions about how this musical style became ubiquitous.

Pitch and loudness of musical instruments

In plucked string instruments, a stretched, uniform string is excited by striking—an impulsive and therefore broad-band mechanism for energy input. In bowed strings, the strongly non-linear, stick-slip interaction of bow and string is capable of regeneration over a large frequency range. In both cases, the fundamental frequency and therefore pitch is usually controlled by the stable, high-Q resonances of a string. Once tuned, the player determines the pitch by choosing a string, and controlling its length. These are usually held constant during a note, and it is easy to play a pitch that is, to a good approximation, independent of loudness. The loudness is determined by the excitation mechanism: it is controlled independently of pitch by another set of parameters (and, for instruments like the violin or guitar, by the other hand: left for pitch, right for loudness). A *decrescendo* on a note—a gradual decrease in loudness at constant pitch—is the natural idiom for plucked strings. For bowed strings, both *crescendo* and *decrescendo* at stable pitch are easy: one holds the string length constant and increases or decreases respectively the speed of the bow[1].

In wind instruments, a lip, a reed or an air-jet is part of a non-linear mechanism that converts the steady input of energy from the lungs into acoustic energy, while a column of air provides the resonances that largely determine the pitch. In these instruments, the pressure and flow of air largely determine the loudness[2]. Again, two very different gestures control the pitch and loudness, which can be varied almost independently with ease.

Pitch and loudness in the voice

In contrast, pitch and loudness are strongly correlated in the voice (Sundberg, 1987). The voice usually operates at frequencies below those of the resonances of the tract. It follows that, for most singing, the pitch is not stabilised by interaction with a resonance at the desired frequency. Further, higher harmonics of the voice are not usually precisely tuned to or by a resonance (Henrich *et al*, 2011). (High singing voices are possible exceptions (Joliveau *et al*, 2004, Garnier *et al*, 2011)).

[1] The stability of pitch in musical instruments is not absolute. On string instruments, very high bowing speeds require substantial increases in bowing force (called 'bow pressure' by players). Large increases in force can change the pitch.

[2] In wind instruments, the pitch can also increase or decrease as a function of blowing pressure (especially in the recorder). However, the changes in pitch are usually fractions of a semitone, because the dependence of pitch on blowing pressure in wind instruments is much smaller than that of the voice.

The pitch depends on the tension and geometry of the vocal folds (varied by the vocalis and cricothyroid muscles). However, because of the strongly nonlinear nature of the aeroacoustic and mechanical interactions that produce phonation, the pitch also depends strongly on the subglottal pressure in the air from the lungs. The latter dependence is easily demonstrated by slapping the chest of a singer while he or she is sustaining a note: the sudden, brief increase in pressure causes a brief but large increase in pitch.

Loudness also increases with increasing pressure. Consequently, the pitch and loudness of the voice are strongly correlated. This is illustrated in a phonetogram or voice range profile (Gramming and Sundberg, 1988): singers are asked to produce a *crescendo* and *decrescendo* on each note of their range and the range of sound levels is plotted as a function of pitch (Figure 1). Phonetograms usually show two distinct, overlapping areas, one for vocal mechanism M1, the normal or modal voice for men and the chest voice for women, and one for M2, the falsetto voice for men or the head voice for women. In each area, strong correlation of sound level with pitch is observed.

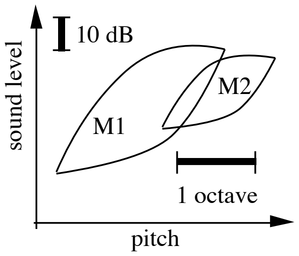

Figure 1. An idealised phonetogram for a man. Scales at constant loudness are horizontal lines on such a plot, *crescendi* and *decrescendi* on a note are vertical lines. For a woman, the M1 region would typically be smaller and the M2 region larger (from Wolfe and Schubert, 2010).

In musical phrasing, there is often a global correlation between pitch and loudness: high notes in a phrase are often played more loudly (*e.g.* Friberg *et al*, 2006). However, independence is also required in music with digitised pitch (as found in much Western music). First, passages covering a range of pitches often need to be sung with little variation in loudness. Further, and more importantly, expressive phrasing often requires a *crescendo* on one note at constant pitch or a *decrescendo* at constant pitch, especially in the last note of a phrase when that note is sustained.

To sing *crescendo* or *decrescendo* at constant pitch, singers must learn to make exactly compensating adjustments to vocal fold tension and subglottal pressure: as the pressure is lowered during a *decrescendo*, the muscular tension must be adjusted continuously to maintain constant pitch. This is difficult, although that difficulty is less obvious to those who have practised it regularly from a very young age. Nevertheless, a *messa di voce*, comprising a slow *crescendo* followed by a *decrescendo* on the same note, is an exercise practised by classical Western singers throughout their lives. They also practise scales at constant loudness, which similarly require compensating adjustments.

What sorts of sounds suit the voice?

The strong correlation between loudness and pitch is well suited to the normal prosody of speech: an emphasised syllable often has a local maximum both in pitch and in loudness, especially in non-tonal languages, and declarative sentences often end with falling pitch and decreasing loudness. Viewed from the other direction, this correlation in prosody is likely to have evolved to suit the acoustics of our voices.

However, the discussions above show that, unlike artificial musical instruments, the voice is not suited to performing music with quantised or categorical pitch and with expressive use of independently varying loudness.

Even the digitisation of pitch itself adds a difficulty for the voice. Musicians often opine that fretted string instruments (e.g. guitar and lute) and wind instruments with tone holes and valves are relatively easy to play in tune compared with continous pitch instruments, such as the (unfretted) violin family and the theramin.

Music and musical conventions and taste in Western music demand not only digitised pitch, but also this independent control, but in a strong sense neither is natural for the voice.

Not all vocal musics have digitised pitch. Examples include the 'tumbling strains' found in Australian aboriginal music (Sachs & Kunst, 1965) and songs with clear speech-like elements, such as the Maori *haka* (List, 1963) and certain Thai Phake songs (Morey, 2010). Furthermore, in popular music, melodic patterns resembling the leap followed by a descent of the tumbling strain can be found in the Blues and several recordings by Bob Dylan (Mellers, 1981).

So, given the abundance and venerability of musics that suit the acoustical constraints of the voice more than does fixed pitch music, one may well ask: What are the advantages of the latter style, and why does it dominate Western and world culture? We have previously proposed that categorisation or quantisation of pitch and time allow efficient compression of the information content of music (Wolfe, 2003). This allows it to be remembered easily, notated efficiently and transmitted accurately. It also facilitates the construction of more elaborate compositions. Finally, digitised music with stable pitch also facilitates harmony. Here we review that argument briefly.

DIGITAL MUSIC AND ITS NOTATION: QUANTISATION OF PITCH AND TIME

The familiar Western music notation is fundamentally digital in frequency and time. (Much Indian music, Shakuhachi music, Blues and Sprechstimme are also usually transcribed using a digital pitch convention.) Usually, a set of rather less than 100 discrete pitches is used. In the commonly used equal temperament, which has 12 equal semitones to the octave, the log frequency scale used for pitch means fundamental frequencies are quantised in factors of $2^{1/12}$. Time is quantised by the tempo marking and time signature, although integral fractions, especially $(1/2)^{\text{integer}}$ and 1/3, are common.

This digitisation allows music to be represented, remembered, written, stored and transmitted in very compact or compressed forms: compression that greatly exceeds that of sound file compressions such as .mp3. Three examples are given in Figure 2. The fact that musicians can often remember many long works and analyse musical structures rapidly suggests that they also use a digitised representation, from whose elements the various digital notations may have arisen. The information content of music and its origin is discussed in detail by Wolfe (2003).

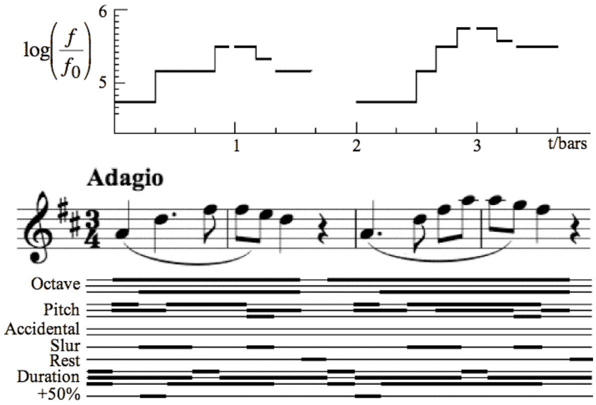

Figure 2. Three different representations or codings of quantised music. The first is a semi-log plot of fundamental frequency *vs* time. The frequency reference is the note C0, with standard frequency $f_0 = 16.3$ Hz. The large tics are octaves and the small tics equal-tempered semitones, with frequency ratio $2^{1/12} \cong 1.059$. On the time axis, the large and small tics are bars (measures) and beats respectively. The second plot is Western music notation: the vertical axis is approximately logarithmic. While the horizontal note spacing is approximately linear, the note shapes are the digital code for internote time. The third panel is a parallel binary coding of the standard notation. The horizontal axis is signal clock time, which is monotonically but not proportionally related to elapsed time: typically it is much faster, but is expanded here for clarity. Three bits give the octave (most significant bit at the top), the next five the pitch. One bit denotes connection to the previous note (a slur) and another silence (a rest). Three bits code for the negative $\log_2$ of the note duration (𝅝, 𝅗𝅥, ♩, ♪ *etc*.) and one bit (a dotted note) can extend it 50%. (One duration code, say 111, could be reserved to toggle from music to text and back, so as to allow for textual expression and tempo markings.) The example shown is the main theme from the slow movement of Mozart's clarinet concerto (after Wolfe, 2003.)

ADVANTAGES OF QUANTISED MUSIC

Signal compression and fidelity

Digital signals have two well-known advantages over analogue. First, they are, up to a point, virtually immune to noise. For musical themes, this would facilitate transmission: provided that the performer's intonation is good enough for the listener to recognise that a given interval is a third, not a fourth, and that this note is half a beat rather than a full beat, the theme is transmitted without error. Further, especially if the theme satisfies some familiar 'rules' of melody (Temperley, 2008), a listener sharing that musical culture is reasonably likely to remember it and to be able to transmit it, undegraded. (Once digital notation for quantised music was developed, it too benefitted from the insensitivity of digital signals to noise.)

Somewhat analogously, the text of speech may be compactly stored once the speech is digitised or categorised into phonemes (elemental speech sounds, which roughly equate to consonants and vowels). However, while there are similarities between the two coding systems, there are also constrasts and complementarities. Simplifying considerably, written text in alphabetic languages codes the phonemes. Acoustically, phonemes correspond to features of the spectral envelope and transients, features that are classified as timbre in music, and not usually notated. In contrast, written music notates the pitch and rhythm, features which are considered as prosody in speech, and not notated in text (Wolfe, 2002; 2007).

It is often argued that the digitisation of time in music is related to rhythms of walking, dance, repeated actions in harvesting and other activities, and even that of the heart: a large proportion of the tempo markings used in music fall in the range of heartbeats and in the range between slow walking and running (see also McAuley, 2010). It is easy to imagine how a song sung while walking, dancing, threshing grain or performing some other repeated chore might acquire digitised time elements.

Second, the digital signal is readily compressed, which allows rapid and accurate transmission. Or, put another way, it permits longer, more elaborate signals to be produced, remembered, stored, analysed and notated with any given information capacity. (See Wolfe, 2003 for more details.) Perhaps it is this that has allowed both the development and the enjoyment of complicated, multi-voice music with intricate structural details that appears to be restricted to digitised music.

Harmony

Quantised pitch has another important consequence. When two notes are simultaneously held steady over time, consonance and dissonance, and the phenomena of interference beats and roughness, are more obvious. The categorisation pitch intervals in Western music are usually related to harmony, and usually explained as being the result of it. For example, the notes of the C major scale (the 'white notes' on the piano keyboard) comprise the component notes of three major triads: that on the root note and those on the notes four and five scale steps above it. The 'black notes' (sharps and flats) arise simply from using the same major triad sequence on a starting note other than C. (The process of reducing the number of steps per octave to 12, as Western keyboards usually do, is complicated in detail because of the problem that 12 perfect fifths with frequency ratio 3:2 is not exactly equal to seven octaves: $(3/2)^{12}/2^7 = 1.014$, a difference of a quarter of a semitone. The various compromise solutions to this problem are called temperaments and the considerable literature expounding and analysing them is reviewed by e.g. Burns (1999) and Barbour (1972).) We give a discussion of this, with sound files (Wolfe and Hatsidimitris, 2012).

We are not the only animals whose songs include discontinuities in pitch: songs of birds and whales also exhibit jumps in pitch. Birdsong has intervals that, in some notations, are often approximated to the nearest note on a scale. However, a recent study of the nightingale finch (Araya-Salas, 2012) shows that the discontinuities in pitch do not correspond significantly to harmonic intervals. It appears that the categorisation in harmonic or scale intervals occurs in the perception by human ears, rather than in the production by birds. Further, the individual notes show glides, and birdsong rarely exhibits *crescendo* or *diminuendo* on more than one steady pitch.

Learning and teaching

Fixed pitch singing may be difficult in performance, but it is perceptually simple. Further, some of the simplest examples of it in Western cultures are the wordless songs that mothers sing to children (see also Trehub *et al*., 1993; Unyk *et al*., 1992). In such songs, pitch, timbre and phoneme are typically held constant during a note, then the pitch is varied but the timbre and phoneme repeated. We have argued elsewhere (Wolfe, 2002; 2007) that while this is not its conscious intention, the effect of such singing could be an example of an effective reductionist strategy for teaching aspects of auditory perception and attention that will later by useful to the child in understanding speech.

Speech is a complicated acoustical signal, with rapid, complex, transient variations in amplitude and spectral envelope, usually delivered with rapidly and continuously varying fundamental frequency. Removing much of the variability and varying the fundamental frequency in a simple, controlled way, may have had important consequences in the development of some of the auditory perception skills that have allowed language to become a channel that uses rapid and complex variations to convey information at a rapid rate. A Western style lullaby with its repeated and elongated syllables is enjoyable and soothing to listen to, but at the same time may fast track the perception of those phonemes.

How and why did we learn quantised music?

How did we learn to use our voices in a music style that is, in a sense, unnatural for the voice, but natural for (artificial) musical instruments? One possibility is that we could have learned to sing the tunes played on instruments.

Bone flutes, with tone holes that would have allowed musical scales, are at least 35,000 years old (Conard *et al*, 2009). However, fixed pitch instruments may be much older. Many modern and most ancient musical instruments would not survive to leave an archaeological record.

Consider the aulos, a roughly cylindrical pipe with either a single or double reed. A simple version of such an instrument can be made from a (hollow) reed or rush—or today from a drinking straw. Cuts at one end, with several possible geometries including that shown in Figure 3, produce a 'blown closed' valve like the oboe reed. This ‘instrument’ can be made to produce just one or a few pitches near the resonances of the column of air. Tone holes covered by the fingers allow more.

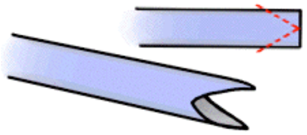

Figure 3. A simple aulos may be made from a hollow reed or drinking straw by cutting along the dotted lines as shown.

Artificial musical instruments may therefore be very old indeed. Singing with quantised pitch could be very old too, though, in the period before writing, it also fails to leave an archaeological record. This makes it impossible to answer definitively the intriguing question: Which came first?

It is often thought that our artificial musical instruments have been invented to make music to imitate that of the voice (e.g. Hubbard, 1808; Seashore, 1942). We have argued elsewhere (Wolfe & Schubert, 2010) that, instead, our ancestors may have developed fixed pitch music, and perhaps harmony too, on artificial instruments. The origin of fixed pitch singing may be more recent.

Musical memes

A meme is the name given to a cultural element transmitted from person to person. (The name draws on an analogy with genes.) Unlike genes, successful memes may be readily and extensively spread within a generation. We have suggested that a phrase or tune using quantised pitch and rhythm is likely to be a more successful meme, because of the parsimonious use of information and the ease of accurate transmission allow for its rapid transmission. Further, the appeal of artificial instruments and their music, and in particular the mimicking of them by the voice, has led to the vocal music in which pitch is quantised and in which pitch and loudness are independently controlled. The development of modern music notation, and the digital representation of music may have emerged from psychological principles founded in meme transmission (Wolfe & Schubert, 2010). Notation, in turn, may have influenced music and advantaged digitisable styles. 'Ear worms' (those highly and even annoyingly memborable snatches of music that lodge in one's mind for minutes or days) may exist in non-quanitised music, but those in digitised music would have an innate advantage, because of their smaller information storage requirements.

A plausible story?

We cannot know the history of the development of digital pitch in music but, from a synthesis of the above discussion, we propose a plausible story.

The first vocal music probably had flexible pitch that varied smoothly. It may not have been much distinguished from language. This music or 'musilanguage' (Mithen, 2005; Brown *et al.*, 2004) might have had discontinuities in pitch, a little like bird song, but these were not tied to harmonic ratios and did not usually have *crescendi* and *decrescendi* at constant pitch.

Early fixed pitch instruments appear. These may have included end blown flutes, transverse flutes, aulos and others having stable pitches. These may be much older than 35,000 years, but have not been preserved in the archeological record.

Very soon after, the first tunes with several categorised pitches appear. Perhaps the earliest of these imitate birds, but with the difference that they have stable pitches. Some of these tunes 'catch on' (like 'ear worms'): they are successful memes. Singers start imitating the stable pitch and singing phrases that use it.

Musicians (players and singers) start noticing beats, dissonance and consonance. Early harmonic ideas appear and influence the notes used in scales. Singing starts to include *crescendi* and *decrescendi* at constant pitch, in spite of its difficulty.

Or even, perhaps, because of its difficulty: difficulty in games and some other activities can be an attraction, rather than an obstacle. In Western music, the seventeenth century is considered by some (see, for example, Shera, 1939) as the era of the 'rise of the virtuoso'—a period in which complex musical techniques were accumulated and codified.

A more recent example of the natural proclivities of the voice being subordinated to styles that could have arisen on artificial instruments is the trend to eliminate portamento from singing, particularly in the early 20th century (Potter, 2006).

CONCLUSIONS

Whether or not the history of singing has followed the possible evolution sketched in the previous section, the acoustic arguments presented above should at least leave us open to the idea that music on fixed pitch instruments has influenced singers for a very long time, perhaps even longer than the influence in the more commonly acknowledged reverse direction.

REFERENCES

Araya-Salas, M 2012, 'Is birdsong music? Evaluating harmonic intervals in songs of a Neotropical songbird' *Animal behavior*, vol. 84, pp. 309-313.

Barbour, JM 1972, *Tuning and temperament: A historical survey*, Da Capo Press, New York.

Brown, S, Martinez, MJ, Hodges, DA, Fox, PT & Parsons, LM 2004, 'The song system of the human brain.' *Cognitive Brain Research*, vol. 20, pp. 363-375.

Burns, EM 1999, 'Intervals, scales, and tuning' in *The psychology of music*, edited by D Deutsch, Academic Press, San Diego, pp. 215-264.

Conard, NJ Malina, M & Münzel, SC 2009, 'New flutes document the earliest musical tradition in southwestern Germany' *Nature* vol. 460, pp. 737-740

Fletcher, NH 1992, *Acoustic Systems in Biology*, Oxford University Press, New York.

Fletcher, NH & Rossing, TD 1991, *The Physics of Musical Instruments*, Springer, New York.

Friberg, A, Bresin, R & Sundberg, J, 2006 'Overview of the KTH rule system for musical performance.' *Advances in Cognitive Psychology,* vol. 2, pp. 145–161.

Gramming, P & Sundberg, J 1988, 'Spectrum factors relevant to phonetogram measurement', *J. Acoust. Soc. Am.* vol. 83, pp. 2352-2360.

Garnier, M, Henrich, N, Smith, J & Wolfe, J 2010, 'Vocal tract adjustments in the high soprano range' *J. Acoust. Soc. Am*. vol. 127, pp. 3771-3780.

Henrich, N, Smith, J & Wolfe, J 2011, 'Vocal tract resonances in singing: Strategies used by sopranos, altos, tenors, and baritones' *J. Acoust. Soc. Am*. vol. 129, pp. 1024-1035.

Hubbard, J 1808. *An essay on music: pronounced before the Middlesex musical society, Sept. 9, AD 1807, at Dunstable, (Mass.)*. Manning & Loring, Boston.

Joliveau, E, Smith, J & Wolfe, J 2004, 'Tuning of vocal tract resonances by sopranos', *Nature*, vol. 427, p. 116.

List, G 1963, 'The boundaries of speech and song.' *Ethnomusicology* vol. 7, pp. 1-16.

McAuley, JD 2010, 'Tempo and rhythm' in *Music Perception*, edited by MR Jones, AN Popper & RR Fay, Springer, New York.

Mellers, W 1981, 'God, modality and meaning in some recent songs of Bob Dylan.' *Popular Music* vol. 1 pp. 143-157.

Mithen, S 2005, *The Singing Neanderthals: The Origins of Music, Language, Mind and Body*. Weidenfeld & Nicolson, London.

Moelants, D 2008, 'Hype vs. natural tempo: a long-term study of dance music tempi.' in *Proceedings of the 10th International Conference on Music Perception and Cognition*. Sapporo, Japan.

Morey, S 2010. 'Syntactic variation in different styles of Tai Phake songs.' *Aust. J. Linguistics* vol. 30 pp. 53-65.

Potter, J 2006, 'Beggar at the door: the rise and fall of portamento in singing', *Music and Letters* vol. 87, pp. 523-550.

Sachs, C & Kunst J 1965, *The wellsprings of music*. McGraw-Hill, New York.

Seashore, CE 1942, 'In Search of Beauty in Music', *The Musical Quarterly* vol. 28, pp. 302-308.

Shera, FH 1939, *The Amateur in Music*. Books for Libraries Press, Freeport, NY.

Sundberg, J 1987 *The Science of the Singing Voice,* Northern Illinois Univ. Press, De Kalb, IL.

Temperley, D 2008, 'A probabilistic model of melody perception.' *Cognitive Science* vol. 32 pp. 418-444.

Trehub, SE, Unyk, AM & Trainor LJ 1993. 'Maternal singing in cross-cultural perspective.' *Infant behavior and development*, vol. 16, pp. 285-295.

Unyk, AM, Trehub, SE, Trainor, LJ & Schellenberg, EG 1992. 'Lullabies and simplicity: A cross-cultural perspective.' *Psychology of Music*, vol. 20, pp. 15-28.

Wolfe, J 2002, 'Speech and music, acoustics and coding, and what music might be 'for''. *Proc. 7th International Conference on Music Perception and Cognition*, Sydney. K Stevens, D. Burnham, G. McPherson, E. Schubert, J. Renwick, eds., pp 10-13.

Wolfe, J 2003, 'From ideas to acoustics and back again: the creation and analysis of information in music'. *Proc. Eighth Western Pacific Acoustics Conference*, Mebourne. (C. Don, ed.) Aust. Acoust. Soc., Melbourne, Aust.

Wolfe, J 2007, 'Speech and Music: acoustics, signals and the relation between them' International Conference on Music Communication Science, Sydney, (Schubert E, Buckley K, Eliott R, Koboroff B, Chen J & Stevens C., eds.) pp. 176-179.

Wolfe, & Hatsidimitris, G 2012 'Interference and consonance' http://www.animations.physics.unsw.edu.au/waves-sound/interference

Wolfe, J & Schubert, E 2010, 'Did non-vocal instrument characteristics influence modern singing'. *Musica Humana*, vol. 2, pp. 121–138.