Article

MUSICÆ SCIENTIÆ

Musicae Scientiae
1–15

DOI: 10.1177/1029864916631393
msx.sagepub.com
SAGE

Voicelikeness of musical instruments: A literature review of acoustical, psychological and expressiveness perspectives

Emery Schubert
The University of New South Wales, Australia

Joe Wolfe
The University of New South Wales, Australia

Abstract

What makes an artificial musical instrument such as the flute, trombone or cello sound like the human voice, and which of all instruments is the most voicelike? This article reviews acoustical and psychological arguments that might explain why a musical instrument would be likened to the singing human voice. The authors could find no evidence to support the idea of any single instrument being systematically, and consistently, regarded as voicelike. The human voice is frequently referred to as an ideal to which a well-played musical instrument should aspire. However, investigations are few that go beyond speculation and introspection regarding what instrument or instruments sound voicelike and why. Although no single instrument emerged as being systematically and consistently voicelike, a program of empirical research is suggested to determine whether there currently exist (possibly culturally promulgated) beliefs about what instrument is considered most voicelike and why.

Keywords

acoustics, action-perception loop, expressiveness, mimicry, timbre, voicelike musical instruments

Music psychologists have shown increasing interest in explaining why we experience emotion in music, particularly in non-vocal music, which has no words to guide us with denotations of emotional experiences. The vocal apparatus is good at communicating emotion through prosody, and there exists strong evidence that music, whether performed by the voice or non-vocal instruments, is able to communicate these emotions (Juslin, 2000; Juslin & Laukka, 2003a; Mithen, 2005, 2009; Scherer, 1995). One disarming explanation of the ability of instrumental music to communicate emotion is that musical instruments themselves share qualities with the

Corresponding author:
Emery Schubert, School of the Arts and Media, University of New South Wales, Sydney, NSW 2052, Australia.
Email: e.schubert@unsw.edu.au

human voice. And of those, some musical instruments[1] may be more suited to resembling the sound of the singing human voice than others. For example, the cello has been claimed to be the instrument that best resembles the human voice (Juslin, Harmat, & Eerola, 2014, p. 604; Juslin & Västfjäll, 2008, p. 574). But what is the evidence for such a claim? Is it sufficient to base such an account on the assertion of a musician (as do Juslin et al., 2014, p. 619, who make reference to a statement by cellist Steven Isserlis about the cello being the instrument that sounds most like the human voice; see further, Isserlis, 2011)? In our review of the literature, we found no empirical evidence that explicitly tested what musical instrument sounds most like the human voice. However, given the apparent influence of the human voice upon instrumental music (as this review will further reveal) it was necessary to investigate the validity and roots of a conclusion such as "the cello is the most voice-like musical instrument".

Historical accounts lack consensus, even though musical instruments that sound like the human voice have been reported almost across the entirety of written history. An early reference can be found in the *Twelfth Pythian*, an ode dating from c. 490 BC by the Greek poet Pindar. Reference is made to the player of the aulos making the instrument sound like the lamentation of the human voice. The instrument is thought by some scholars to resemble a flute, others have argued that it sounds more like a reed instrument (Held, 1998; Steiner, 2013), as suggested by the reference from the poem itself: "through vocal vent its music flows, Of brass with slender reed combined" (from *Twelfth Pythian*, translation by Turner & Moore, 1852, p. 343). In the Renaissance, the viola da gamba was viewed as a special instrument because of its voicelike qualities (Danks, 1979, p. 41; Dolan, 2008).[2] In Sulzer's *General Theory of Fine Arts* (1771–1774) the entry of instrumental music places the oboe as the most voicelike instrument (Sulzer, Baker, & Christensen, 1995, p. 97). The "pureness" of the glass harmonica ("armonica", an instrument of soprano range) was said to make the instrument a rival to the human voice in the late 18th century (Hadlock, 2000, p. 513). Reuter's (2002) detailed account of historical sources describing the tone of musical instruments also pointed out that comparisons between musical instruments and the human voice have been made throughout recorded history, but especially so in the beginning of the 19th century. Of the instruments examined only the Cor Anglais, Clarinet, Saxophone, Contrabassoon and Tuba had no connections made to the human voice in the sources examined by Reuter. Those which were linked to the human voice in historical sources were the following woodwind instruments: the Flute (including Alto Flute), Oboe and Bassoon; and the following lip-reed instruments: French Horn, Trumpet, Trombone, Zink (or Cornett), Serpent, Basshorn (an instrument related to the Serpent), Ophicleide, Cornet and Bugle. The identification of a single most-voicelike musical instrument is therefore controversial. In this article we consider the evidence for these positions to help scrutinise the meaning of voicelikeness. The matter is addressed from three perspectives: (1) Acoustic evidence, in which the physical properties and mechanisms of the human voice and musical instruments are compared; (2) Psychological issues, where we examine neuroscience and perception in instrument identification; and (3) The role of musical expressiveness, specifically how concepts of "musical expressiveness" (as distinct from emotional expression) may be related to the idea of voicelikeness.

Acoustic evidence

It was not until the work of Helmholtz (1912) in the middle of the 19th century that major steps towards systematic and empirical understanding of the production of sound by the human voice were made. While interest in the human voice up until this time was notable and led to the building of mechanical devices that could generate speech sounds, such as Joseph

Faber's Euphonia (Hankins & Silverman, 1995), there was also debate and speculation on the mechanism of vocal production. Physiological and acoustical understanding of the voice has advanced significantly in the 20th century. In this section we briefly overview how the voice functions, and compare its mechanisms with those of musical instruments.

Pitch control of the voice is largely achieved by a combination of the tension and geometry of muscles in the larynx and sub-glottal pressures (for further details, see Atkinson, 1978; Garnier, Henrich, Smith, & Wolfe, 2010; Hirschberg, Pelorson, & Gilbert, 1996; Honda, Hirai, Masaki, & Shimada, 1999). The geometry of the vocal cavity (including the shape and position of tongue, lips and jaw) largely serves the function of a filter, which produces frequency bands of enhanced power in the output sound (e.g. Fant, 1960). We also note that, while the larynx is often considered as a source and the tract as a filter and that the two are largely independent, they do of course interact (e.g. Swerdlin, Smith, & Wolfe, 2010; Titze, 1994). The fundamental frequency of vibration is a complicated result of several different muscle tensions and configurations and, importantly, increases strongly with the subglottal pressure.

Helmholtz used resonators to detect or to observe the presence of component frequencies in stable, complex tones and, in particular, different vowel sounds (Helmholtz, 1912). This led to the discovery of one of the complexities and marvels of the human voice: that it could manipulate the relative amplitudes of harmonics via formants (collections of prominent harmonics, Standards Secretariat, 1994)[3] to generate different vowel sounds. These different vowel sounds could, in principle, be related to different timbres on musical instruments. For example, the horn and the bassoon are both reported to have a formant around 400–500 Hz (Smith & Mercer, 1974), and the human voice can produce a similar formant singing a vowel close to /u/ or /o/. But for a musical instrument to be able to imitate the sounds of the human voice, it would need to be able to routinely manipulate its formant structure rapidly and dramatically as functions of time, just as the vocal cavity does in the human voice. None of the traditional, western musical instruments (piano, guitar, instruments of the orchestra, etc.) can do this without some non-typical intervention, such as moving a mute smoothly but rapidly on and off an instrument, speaking/growling through a wind instrument, using an electronic effects processor and so forth. So this constancy of formants adds to the conundrum of how musical instruments might plausibly be identified as *being* voicelike from this physical, acoustic perspective.

Table 1 compares and contrasts the basic mechanics and acoustics of the voice and several classes of musical instruments (for additional, relevant acoustic information see, e.g. Fletcher & Rossing, 1998). For the brevity necessary in such a table, many simplifications and approximations are made, of which we now signal some of the most serious. In virtually all instruments, the spectral envelope varies with loudness and with pitch; here we consider only independent control giving large variation. *Portamento* is straightforward on most bowed strings and on the trombone; it is possible on discrete-pitch woodwind (e.g. Chen, Smith, & Wolfe, 2009), but usually relies on advanced techniques and is not idiomatic for these instruments. It is also possible on the (plucked) *oud*, but is not idiomatic on the (bowed) viols.

Some modification of the spectral content is possible on orchestral wind instruments using the resonances of the vocal tract. However, compared with those of speech or didjeridu playing, the effects in the radiated sound are modest (although they may seem much greater to the player) and they require advanced techniques (Li, Chen, Smith, & Wolfe, 2015).

Virtually alone among musical instruments, the didjeridu is capable of producing variable formants with magnitudes of tens of decibels. Further, they are produced and controlled using a mechanism somewhat similar to that of the voice: the player produces a formant at a particular frequency by adjusting the geometry of the vocal tract so that its acoustical impedance at

Table 1. A highly simplified comparison of the mechanical and acoustical features of voice with those of several classes of acoustical musical instruments.

	Voice	Didjeridu	Brass	Woodwind	Bowed string	Struck string	Tuned percussion
Energy input	Continuous: High pressure air from lungs	Continuous: High pressure air from lungs	Continuous: High pressure air from lungs	Continuous: High pressure air from lungs	Continuous: Steady motion of bow	Impulsive: collision with finger, etc.	Impulsive: collision with beater
Conversion of steady to oscillatory power	Self-oscillation of the vocal folds	Self-oscillation of the player's lips	Self-oscillation of the player's lips	Self-oscillation of a reed or air jet	Stick-slip friction between bow and string		
Principal pitch control	Tension and geometry of vocal folds; pressure from lungs	Resonances of the bore of the instrument	Resonances of the bore of the instrument	Resonances of the bore of the instrument	Resonances of the string	Resonances of the string	Resonances of the vibrating element (bar, bell ...)
Portamento	Yes	No	Trombone	No	Usually yes	Usually no	Usually no
Principal sound level control	Pressure from lungs; tension and geometry of vocal folds	Air pressure from lungs	Air pressure from lungs	Air pressure from lungs	Speed of bowing	Forces in the collision	Forces in the collision
Impedance matching to the radiation field	Resonances of the vocal tract	Resonances of the vocal tract and of the bore	The bore, especially the bell	The bore, especially the bell, when present	The instrument body, especially its resonances	The instrument body, especially its resonances	Usually not needed
Control of spectral tilt or brightness	Geometry and tension of vocal folds; geometry of the tract	Embouchure	Embouchure	Embouchure	Place of bowing; bow force	Site of plucking	Site of striking
Variable control of formants	Variable resonances of the vocal tract	Variable resonances of the vocal tract	Mutes (limited control via vocal tract)	Mutes (limited control via vocal tract)	Mutes		

Note. In terms of mechanisms – not output, we stress – overall similarity to the voice increases towards the left. See the text for a list of some of the most serious omissions and simplifications.

the lips is at a minimum for that frequency band and large at other frequencies (Tarnopolsky et al., 2005; Tarnopolsky et al., 2006). One reason why the didjeridu is not usually cited as a particularly voicelike instrument is, we believe, because it does not use *portamento* and indeed is almost always played at constant pitch. Additionally, this constant pitch is lower than the typical range of most human voices.[4]

Most of the discussion in this paper concerns the quiescent or sustained part of the note, because these are one of the key differences between singing and speaking. In singing, vowels are typically elongated (or more rarely shortened) to achieve the rhythm of the song. Further, the fundamental frequency of the vowel is usually held approximately constant in singing, to convey the pitch assigned to that syllable. However, the transients of a musical note, especially the initial or attack transient, are salient (Berger, 1964; Thayer, 1974) and have some parallels with consonants, especially occlusives and fricatives. In both cases, the initial amplitude of the fundamental is sometimes briefly exceeded by that of inharmonic components or of higher harmonics or subharmonics. Also, the initial transient in both cases often includes a relatively loud burst of broad-band noise (e.g. Bello et al., 2005). Once again, in the case of the voice but not of instruments, the broad-band noise exhibits strong formants that can be widely varied. Further, the first and second formants of the subsequent vowel vary with time as the lip aperture or other constriction closes or opens. These variations are characteristic of different consonants (Clark, Yallop, & Fletcher, 2007), another feature that is not possible on musical instruments.

While the order of the columns in Table 1 could be said to rank the mechanical similarities to the voice, at least approximately, it does not resemble the order of voicelikeness found in the literature (see introduction). A limited functional analogy can be made between the voice and a bowed string instrument (e.g. Askenfelt, 1991) by comparing elements and their roles. In the violin, the body serves as the acoustic impedance matcher that transmits power from the vibrating string (high impedance) to the radiation field in the air (low). It does this most effectively at certain resonances of the bridge, body and the air within it: these produce formants (again, meaning enhanced frequency bands of sound power, Standards Secretariat, 1994) in the output sound (see also Traube & Depalle, 2004a, 2004b for some parallels with the guitar). For the human voice, the vocal tract, as an acoustic duct, acts as an impedance matcher that transmits power from the larynx to the radiation field. It does this most efficiently at its resonances, and thus produces (highly variable) formants in speech and singing. For the present argument, the important difference is the severe limitation to the possibility of rapidly varying the formants of a musical instrument.

Consider the comparison of the singing voice with both the flute and violin (e.g. Askenfelt, 1991; Hirt, 2010, pp. 19–20; Reilly, 1997, p. 434). The flute and violin are instruments whose excitation mechanisms, resonators and impedance matchers are completely different at the mechanical level. What acoustical features might explain this perceived similarity? First, all three are harmonic and produced sustained tones, features that they share with all wind instruments. We further note that all three, at sufficiently high pitches, have a dominant fundamental with relatively weak overtones.

An important limitation in the instrument–voice analogy concerns the control of pitch. String, woodwind and brass instruments all use an acoustic resonator to control the pitch. For string instruments, this pitch control resonator is the string itself, whose steady vibration is established at one or more of the resonances due to standing waves of the string. For brass and woodwinds, acoustic resonances due to standing waves in the instrument bore largely control the steady vibration of lips, reed or air jet and thus control the pitch.

At the functional level, the vibrating elements of brass instruments have strong similarities to the voice. In both, the source of sound is the modulation of the breath by two vibrating tissues: the player's lips for brass and the vocal folds for the voice (for reviews, see, e.g. Fletcher & Rossing, 1998; Titze, 1994). Furthermore, in both cases these tissues are acoustically coupled to resonant acoustic ducts on both the upstream and downstream side. The difference is that one or more of the resonances in the bore of a trombone largely control the frequency of vibration of the lips, whereas the resonances of the (much shorter) vocal tract modify the amplitudes of higher harmonics, contributing to timbre rather than pitch.

So the voice differs from wind and string instruments both in lacking a pitch-control resonator, and in being capable of rapid changes in the resonator that produces formants. As we have previously argued, non-vocal musical instruments are good at producing stable pitch but are not so good at varying timbre; the human voice (without considerable practice) is not so good at keeping pitch steady and independent of loudness but good at manipulating timbre (Wolfe, 2002, 2007; Wolfe & Schubert, 2012). On the grounds of current understanding of acoustics and instrument anatomy and mechanism alone, selecting an artificial musical instrument as being voicelike cannot be soundly based on operating principles.

Psychological issues

Neuroscientific evidence suggests that, under certain conditions, perception of the human voice is privileged, activating brain regions that are not activated when auditioning a wide range of other stimuli, including musical instruments (Levy, Granot, & Bentin, 2001, 2003). However, if a non-vocal musical instrument were able somehow to activate some or all of those regions, the listener might interpret the instrument as sounding voicelike. It has been suggested that this mental privilege arises simply because people are typically exposed to human voices more frequently than to non-vocal musical instruments. Consequently, we become experts at processing vocal sounds through experience (Chartrand & Belin, 2006). According to this viewpoint, there would be no inherent, specialist voice-processing pathway: rather, some well-practised pathways. No neuroscientific studies have been cited that specifically map the regions uniquely activated by specific, human voice sounds, and compare them with the brain activation caused by matched, controlled musical instrument sounds. That is, the stimuli used in the abovementioned neuroscientific and behavioural studies to date might not be suitable for addressing the question of interest here because they are of limited musical relevance, in that they employ single tones (Levy et al., 2001, 2003), or sequences of three tones (Chartrand & Belin, 2006), without a conventional, longer musical context.

Our question concerns the relationship between musical instruments and the human voice, so the more ecological comparison would be between musical instruments and the human voice as generators of music. However, many psychological and neuroscientific studies investigate implicit *processing* and physiological changes caused by vocal information, ignoring the *phenomenal experience* of similarity between voice and instrument sound, a limitation shared with the acoustics explanation. The mental cognition and the physical/physiological operation of the brain are the central issues, and the individual concerned may not have conscious access to either of those. Also relevant is the cognitive-behavioural aspect of such processing, such as the influence upon memory. These issues were dealt with in a study by Weiss, Trehub, and Schellenberg (2012).

Weiss et al. (2012) demonstrated that melodies are remembered better when presented by a singing voice rather than played on an artificial musical instrument. In their study, folk melodies from the United Kingdom were recorded using a female alto singing the melodies with a

"la" syllable, with additional recordings made of the same melodies using three acoustic artificial musical instruments (piano, marimba and banjo) and MIDI generated instruments. For the MIDI generated versions, the authors generated the stimuli using the pitch, duration and amplitude patterns of the vocal version, however, the expressive performance parameters such as vibrato were not retained (M. Weiss, personal communication, July 2, 2014). In an exposure phase, participants listened to 16 melodies in four different timbres and were encouraged to focus on the music by being asked to answer a question about the emotion expressed by each stimulus. In the second phase they heard the 16 melodies again, but this time intermingled with 16 new melodies. Their task was to rate their confidence that they had heard the melody before. The superior recall of melodies that were sung in the exposure phase was explained in terms of the additional arousal and vigilance evoked by the human voice over other musical instruments.

These arousal and vigilance inducing characteristics were thought by Weiss et al. to contribute to deeper cognitive processing of vocal stimuli than other kinds of stimuli. Of particular relevance is the conclusion that "subvocal activity or related motor imagery could have enriched participants' representations of the vocal melodies" (2012, p. 4). This conclusion argues that vocal sounds are privileged because we are able to mimic them better than sounds from musical instruments. That is, humans possess the same apparatus as the vocal source they are perceiving. And so, consistent with action-perception psychological processing models, mental representations are shared between the perception and action of singing.

This idea is consistent with Prinz's (1997) perception-action model which proposes that the mental processing for a perceived action (in this case, utterances of the human voice) is shared by the processing required for the action (here, vocal production. See also Galantucci, Fowler, & Turvey, 2006). Principles of mimicry, contagion and empathy are all invoked by such an explanation (Preston & De Waal, 2002). Shared action-perception circuits can be employed to explain the results and to present a model explaining why voice and musical instruments appear to be treated as categorically different. Juslin (e.g. Juslin, 2000; Juslin & Västfjäll, 2008) argued that the human voice plays an important role in communicating emotion via contagion (e.g. a sad vocal utterance can make the listener become sad). Therefore, he argues, an artificial musical instrument's sound is able to communicate emotion contagiously because it bears some critical resemblances to the human voice, particularly in terms of expressive capabilities (a point to which we shall return). The Weiss et al. (2012) study provides some evidence for this argument.

Of course, if a musician is proficient at playing a non-vocal instrument, the shared action/perception mental circuitry could be activated as a result of perceiving that non-vocal musical instrument (Bishop & Goebl, 2014; see also Keller, 2012; Novembre, Ticini, Schütz-Bosbach, & Keller, 2012). And so the argument that the human voice is processed in a privileged way can still be explained by experience and expertise rather than innate (brain hardwiring) advantage. This does not alter the significance of the processing of the human voice, but it does draw attention to the possibility than the neural substrates that are shared by perception and production of a particular musical instrument are alone unlikely to provide an explanation for perception of voicelikeness. They provide information on shared expertise in action and in perception, regardless of the instrument involved.

No studies were cited that accounted for the voice being able to dynamically manipulate formants to produce different vowel sounds, and understandably so. The differences (behavioural at least) would be trivial – comparing a human voice with changing formant structure, such as a diphthong or a consonant-vowel pair, versus a traditional musical instrument playing a tone with a fixed and stable pitch. Instead, researchers have taken a more conservative

approach to see whether "stable" tones and vowels are processed according to the timbre source (voice vs artificial instrument) (Chartrand & Belin, 2006; Levy et al., 2001, 2003). In that respect, the identification of any differences in timbre source processing is remarkable. However, none of the studies referred to explicitly control for microtonal nuances, which may present additional biasing cues, as they may be particularly prevalent in the voice, but also to some extent in string instruments and the trombone as these instruments can perform pitch slides relatively easily over a wider range of pitches.

Another way of explaining the *perceived* similarity between the singing voice and a musical instrument is that top-down processing overrides the bottom-up signal similarities. Consider the study by Sarris and Tzevelekos (2008). The authors were interested in conducting an acoustic examination of the thesis that the "open throated" singing style of some of the vocal music in the Balkans resembles musical instruments used in that region. Although they found several signal similarities between vocal performance and *gaida* (bagpipe) playing techniques, they also observed that it was part of the culture to link the *ganga* polyphonic song of Hercegovina to the qualities of the *gaida*. This top-down, cultural imposition of instrument sounds needs to be taken hand in hand with evidence of physical mechanics and other bottom-up characteristics discussed above. And furthermore, the expressiveness of the musician also plays a role in determining whether a musical instrument sounds like a singing human voice, arguably an additional form of top-down processing because expressiveness in music is driven by culture rather than physics (Fabian, Timmers, & Schubert, 2014). We therefore turn our attention to the role of expressiveness in determining the voicelikeness of a musical instrument.

The role of musical expressiveness

Our review to this point has examined literature from acoustics and psychology to see what the state of the art is in our knowledge about the voicelikeness of musical instruments. Acoustic analysis is firmly rooted in a bottom-up perspective on voicelikeness, because looking at the physics involved in producing the sounds may give clues about what makes a musical instrument sound like the human voice. The psychological approach is a mixture of bottom-up and top-down explanations because it tries to account for perception of sound that is partly a result of its physics, and partly a result of the way we are influenced by factors that are not directly related to the signal, but to matters such as culture and memory. We call these top-down, and in this section we take the top-down aspect further by turning our attention to the influence of expressiveness on the idea of a musical instrument being voicelike. Two things need to be noted in this part of the review. One is that the investigation involves analysis of acoustic, psychological and musicological arguments about *expressiveness* and the other is that we restrict our use of the term expressiveness to "musical expressiveness" rather than "the expression of emotion" (as proposed by Schubert & Fabian, 2014; for excellent reviews of the latter, see Juslin & Laukka, 2003b; Scherer, 1995).

Consider the editorial comment written in response to Uvedale Price's *The Picturesque*, first published in 1794, with a short discussion about music at the end of the fifth chapter. In the 1842 edition, the editor (who was presumably the Scottish author Sir Thomas Dick Lauder, but is listed as the second author of the 1842 edition) made an extended remark in response to Price's assertion that "the human voice is the most beautiful and melodious of all sounds":

> But why does the human voice affect us more powerfully than the sound of a musical instrument? Is it because its tones are finer, more delicate, or more powerful? I suspect not. The most magnificent human voices can be excelled in all these particulars by certain instruments, when played on by the

> best performers. The greater influence which the human voice possesses over us, arises from the circumstance of its being the human voice. For, as the influence which instrumental music has over us, arises from the association which its tones awaken with the feelings and passions of human nature, so it follows, that the human voice, as being more immediately connected with these, must be in itself a superior vehicle for their expression. It has also the immense advantage of being able to give utterance to those sentiments of poetry, with which the notes have been harmoniously associated. In support of this view, the experience of every one must bear witness to the fact, that it is by no means always the finest voice, considering it as an instrument, that most deeply touches the human heart, and that feeling and powerful expression, will always awaken more chords of sympathy, and more general emotions in the minds of the auditors, than the finest toned voices can possibly do without it. Nay, the very power which instrumental music possesses over us, depends entirely on the extent to which this mental feeling and expression can be imitated. (Price & Lauder, 1842, p. 109)

The quote by the editor has two ideas that are particularly relevant, and that have been hinted at in the above overview. First is the possibility that an artificial musical instrument may actually sound more beautiful than the human voice – a reflection of the pro-instrumental, absolutist aesthetic that peaked in 19th century European high-art music (Dahlhaus, 1978/1989). The second concerns the expressive power of the performer. From an aesthetic perspective, it is the expressive capacity of the performer that is more important than whether the instrument itself sounds like the human voice, with the ultimate goal of touching the human heart – the emotions. Expressive capabilities of the performer might be a necessary requisite to allow exploitation of the full expressive, and therefore potentially voicelike, capacity (if any) of a musical instrument. Therefore, expressiveness might make an independent but critical contribution to the voicelikeness of a musical instrument. Goehr put it like this:

> The seemingly simple prescription that instrumental playing should approximate to the condition of singing is not [...] simply a demand that the violin *sound like* a human voice. It is a demand that a violinist should sing as a singer sings, where the analogy between the violinist and singer depends upon an elusive metaphor of musicality usually expressed with all its Romantic and metaphysical grandeur. (Goehr, 1998, p. 123, emphasis in original)

The famed clarinettist of Mozart's time, Anton Stadler, played with an ensemble Mozart's Serenade for 13 wind instruments, K361 (the *Gran Partita*). In writing of the event in 1784, Johann Schink accoladed Stadler: "Never would I have thought that a clarinet could be capable of imitating the human voice as deceptively as it is imitated by you. Truly your instrument has so soft and lovely a tone that nobody who has a heart can resist it'" (cited by Lawson & Stowell, 1999, p. 110).[5] This quote clearly highlights the importance of the player's expressive capacity to be able to make the instrument sound voicelike. However, it is worth staying cognisant of possible psychological artefacts; the concert was also a tribute to Stadler, who already had a reputation as a very fine player. And so, it is possible that we are also seeing a halo effect (a top-down effect, where the individual attributes individual characteristics a value commensurate with their global judgement: see Nisbett & Wilson, 1977). That is, resemblance to the human voice is symptomatic of the pleasure that the player brings to his audience. It makes the job of separating player capacity and instrument voicelikeness more complex.

Some musical instruments have characteristics apart from timbre, air source and vibration mechanism that make them resemble the singing voice more than others. Any instrument that is able to glide in pitch is an example, and includes most string instruments, and the trombone, because it allows them to mimic easily microtonal inflections such as *vibrato* and *portamento*,

which the human voice can do relatively easily (Schubert & Wolfe, 2013). But when matched up with a capable performer, some instruments thought to have digital pitch control can be manipulated to better mimic the human voice by application of pitch sliding, such as the *portamento* required for part of the opening of Gershwin's *Rhapsody in Blue* as performed by the solo clarinet. Interestingly in this example, the vocal apparatus plays a crucial link in how the clarinet can achieve such a wailing, possibly voicelike quality (Chen et al., 2009), and is required in some advanced techniques used in 20th and 21st century compositions (e.g., see Read, 1993). Even restricting the definition of voicelikenesss to what the artificial musical instrument is capable of doing is limited until we discover the full range of capabilities of the instrument in the hands of the expert player.

The piano and many keyboard instruments with acoustic sound excitation mechanisms are not capable of producing sustained tone, *portamento*, *vibrato* and pitch/loudness independent control of formant structure (with the clavichord being an exception) regardless of the expertise of the player. The piano does not have the same range of note attack options as the human voice, nor is it able to sustain or louden a sound once the hammer strikes the string. One might expect that the more of these voicelike things that an artificial musical instrument is able to do, the greater its *potential* to sound voicelike. Still, if a musician judges any instrument as sounding voicelike, even a piano, it may be that this is what they hear – without further justification required.

Conclusion

Imposing qualities of the human voice on the performance of non-vocal musical instruments has interested musicians and music lovers for millennia, and in that respect it is interesting that it has received fairly limited attention in scientific fields of enquiry, and in particular music psychology. From an acoustic perspective, the human voice has some mechanical similarities to brass (lip reed) instruments, and to a lesser extent to reed instruments. But one significant difference is the absence in the voice of a resonator to control the pitch (Table 1). Because of the voice's readiness for *portamento*, instruments with greater pitch flexibility may be considered better candidates, such as many (especially unfretted) string instruments and the trombone. However, these observations all omit the rapid time-variation in spectral envelope that is characteristic of the voice, but relatively rare in standard acoustical musical instruments. Further research could more explicitly examine the bottom-up characteristics of a non-vocal auditory signal that will sound most voicelike. Comparisons of radiated spectra at different frequencies at equal amplitudes could be made, encompassing low, mid and high pitch and soft, moderate and loud dynamics. Or specific comparisons between the formant structure of the vowels in voice could be made with the resonances of winds, brass or strings. This would enable researchers to see whether the properties of the steady tone in the voice are more similar to certain instrument families and instruments than others. But the scope of such research will be limited because of the difficulty in generalising the findings to more complex musical contexts, and because of the absence of the consideration of top-down influences on the perception of voicelikeness.

The article also acknowledged that judgements of voicelikeness rely on perceptions that are to some extent controlled by top-down psychological phenomena too, including the role of culture. If an individual hears an instrument as being voicelike, then all arguments about acoustics and other justifications vanish. The question then is a more cultural/psychological one – is there agreement as to which instruments sound more voicelike than others? Future research could make inroads into this matter by gathering survey data on the musical instrument that

sounds most like the human voice. The literature review above suggests that consensus will be unlikely. But identifying the factors that may lead to a bias toward one musical instrument rather than another may have interesting implications.

The current literature review was not able to identify a single musical instrument or even a class/family of instruments that was consistently, and throughout history linked to being voicelike. Instead, voicelikeness may be another way of saying something positive about an instrument – what can be viewed as a top-down assessment rather than a bottom-up acoustical argument. If someone likes a musical instrument, they may say, in addition to liking the sound of the instrument, other things that embellish the generalised liking. One embellishment is to refer to the resemblance of the instrument to the human voice, and this comes about because historically, and possibly across numerous cultures and styles of music, the singing voice is seen as potentially the most perfect, superb musical instrument (e.g. Hirt, 2010, pp. 19–20).

The perfection assumption of the human voice has an obvious flaw, because not all singing voices are equal. Some people can sing better than others, some sing in different ranges, and some sing with different vocal characteristics that are more amenable to a given style of performance than others (such as a country and western singer attempting a Wagner Opera role, and vice versa). This article has argued that the critical issue missing here is that of expressiveness. Expression is an important component of what makes an instrument sound voicelike, but it requires the addition of a variable – the player. In addition, the ability of the performer to exploit the (sometimes unknown) expressive potential of an instrument is needed so that the instrument can better resemble the equally expressive singer (as distinct from any singer).

In sum, when we refer to voicelikeness in an expressive sense, we are probably referring to resemblance with an idealised vocal expression, rather than a typical or inferior one. Resemblance from a purely perceptual standpoint may be a matter of finding the just noticeable difference in tone between the voicelike instrument and an otherwise matching voice. The review of the psychological literature suggests that such a study has not been conducted, despite much interest in similarity rating of timbres in general through the work of Grey, Wedin and others (e.g., Grey, 1977; Kendall, Carterette, & Hajda, 1999; Lakatos, 2000; Wedin & Goude, 1972). The presence of a specialist, localised voice-processing region of the brain has so far produced equivocal evidence, and it appears the prevalence and importance of the voice means that we are better at processing it for reasons that can be explained by experience, rather than something intrinsically special about the voice. The relationship needs to be understood in terms of aesthetics, culture, expressiveness (including the role of the performer) and through an informative theoretical framework, such as the action-perception model that explains why the voice is a good but not the sole candidate for privileged, expertise-based, mental processing.

Acknowledgements

Comments and suggestions made on an earlier draft by Patrik Juslin are greatly appreciated. Thanks to Bettina Rosche and Georg Ramm for assistance with German text translation.

Funding

This research was supported by the Australian Research Council (FT120100053) held by author ES.

Notes

1. The term “musical instrument” will be used here as a synonym for “artificial acoustical musical instrument”. That is, it can refer to all musical instruments that exist outside the human body, but

require human intervention to activate directly the sound production, usually by inputting energy from the breath, and/or doing work with the hands, which then leads to energy radiating through the atmosphere. This definition thus excludes electronic instruments, such as the Theremin, and in general, synthesisers (especially singing synthesisers), samplers, effects (such as effects pedals used on electric guitars) because the human production (when it occurs) and the corresponding sound output is mediated by an electronic/electromechanical interface. These electronic instruments present a less intriguing case because they can be or have been deliberately designed to mimic the human voice (Cook, 1996, 1998; Feugère, d'Alessandro, & Doval, 2013; Sundberg, 1989; Traube & D'Alessandro, 2005; Traube & Depalle, 2004a, 2004b). Also, in this article, when we refer to the human voice, we refer to the prosodic or melodic aspects of the singing voice, rather than the speaking voice (see, e.g. Patel, 2008; Patel & Iversen, 2003; Sammler et al., 2009; Sammler et al., 2013).

2. The viola da gamba is a fretted string instrument played in the same position as the cello.
3. The historical and formal definitions of formant are broad maxima in the spectral envelope of a sound. In speech science, however, formant is sometimes also used to mean a resonance that gives rise to the spectral maximum. In this paper, formant is used with its formal (Standards Secretariat, 1994) meaning. Smith and Mercer (1974) give examples of the formant frequencies of musical instruments.
4. The pitch range of the voice and musical instruments is not well defined. Music for choirs is rarely written below about E2 (~80 Hz) or above about G5 (~800 Hz), because relatively few men or women respectively sing below or above these limits. Solo parts and solo singers regularly go well beyond these. For instruments, the range of the 88 key piano (A0 at 27.5 Hz to C8 at 4290 Hz) covers approximately the range from the lowest note of the contrabassoon to the highest note of the piccolo.
5. It is worth noting that the *Gran Partita* requires very large interval leaps that would in fact be difficult for the human voice to imitate, specifically those traversing a descending 17th (two octaves and a third) and an ascending 15th (two octaves) in the first basset horn part in the eighth bar of the *Adagio* movement: F5-D3-E3-E5.

References

Askenfelt, A. (1991). Voices and strings: Close cousins or not? In J. Sundberg, L. Nord, & R. Carlson (Eds.), *Music, Language, Speech and Brain: Proceedings of an International Symposium at the Wenner-Gren Center, Stockholm, 5–8 September 1990* (pp. 243–256). New York, NY: Macmillan.

Atkinson, J. E. (1978). Correlation analysis of the physiological factors controlling fundamental voice frequency. *The Journal of the Acoustical Society of America, 63*, 211–222.

Bello, J. P., Daudet, L., Abdallah, S., Duxbury, C., Davies, M., & Sandler, M. B. (2005). A tutorial on onset detection in music signals. *Speech and Audio Processing, IEEE Transactions on, 13*, 1035–1047.

Berger, K. W. (1964). Some factors in the recognition of timbre. *The Journal of the Acoustical Society of America, 36*, 1888–1891.

Bishop, L., & Goebl, W. (2014). Context-specific effects of musical expertise on audiovisual integration. *Frontiers in Psychology*, 5. doi: 10.3389/fpsyg.2014.01123

Chartrand, J.-P., & Belin, P. (2006). Superior voice timbre processing in musicians. *Neuroscience letters, 405*, 164–167.

Chen, J.-M., Smith, J., & Wolfe, J. (2009). Pitch bending and glissandi on the clarinet: Roles of the vocal tract and partial tone hole closure. *The Journal of the Acoustical Society of America, 126*, 1511–1520.

Clark, J., Yallop, C., & Fletcher, J. (2007). *An introduction to phonetics and phonology*. Malden, MA: Blackwell.

Cook, P. R. (1996). Singing voice synthesis: History, current work, and future directions. *Computer Music Journal, 20*, 38–46.

Cook, P. R. (1998). Toward the perfect audio morph? Singing voice synthesis and processing. In *Proceedings of the 1st International Conference on Digital Audio Effects (DAFX), Barcelona*. Retrieved 29 January 2016 from https://ccrma.stanford.edu/~serafin/UVA/MUSI445/Cook98.pdf

Dahlhaus, C. (1989). *The idea of absolute music* (R. Lustig, Trans.). Chicago, IL: University of Chicago Press. (Original work published 1978)

Danks, H. (1979). *The viola d'amore*. Halesowen, UK: Stephen Bonner.
Dolan, E. I. (2008). ETA Hoffmann and the ethereal technologies of "nature music". *Eighteenth Century Music*, *5*, 7–26.
Fabian, D., Timmers, R., & Schubert, E. (Eds.). (2014). *Expressiveness in music performance: A cross cultural and interdisciplinary approach*. Oxford, UK: Oxford University Press.
Fant, G. (1960). *Acoustic theory of speech production* (Vol. 2). The Hague, the Netherlands: Mouton & Co. N. V.
Feugère, L., d'Alessandro, C., & Doval, B. (2013). Performative voice synthesis for edutainment in acoustic phonetics and singing: A case study using the "Cantor Digitalis". In *Intelligent technologies for interactive entertainment* (Vol. 124, pp. 169–178). Cham, Switzerland: Springer International.
Fletcher, N. H., & Rossing, T. D. (1998). *The physics of musical instruments*. New York, NY: Springer-Verlag.
Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*, 361–377.
Garnier, M., Henrich, N., Smith, J., & Wolfe, J. (2010). Vocal tract adjustments in the high soprano range. *Journal of the Acoustical Society of America*, *127*, 3771–3780.
Goehr, L. (1998). *The quest for voice: On music, politics, and the limits of philosophy*. Berkeley: University of California Press.
Grey, J. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, *61*, 1270–1277.
Hadlock, H. (2000). Sonorous bodies: Women and the glass harmonica. *Journal of the American Musicological Society*, *53*, 507–542.
Hankins, T. L., & Silverman, R. J. (1995). *Instruments and the imagination*. Princeton, NJ: Princeton University Press.
Held, G. F. (1998). Weaving and triumphal shouting in Pindar, Pythian 12.6–12. *The Classical Quarterly (New Series)*, *48*, 380–388.
Helmholtz, H. L. F. (1912). *On the sensations of tone as a physiological basis for the theory of music* (A. J. Ellis, Trans., 4th ed.). London, UK: Longmans, Green, and Co.
Hirschberg, A., Pelorson, X., & Gilbert, J. (1996). Aeroacoustics of musical instruments. *Meccanica*, *31*, 131–141.
Hirt, K. M. (2010). *When machines play Chopin: Musical spirit and automation in nineteenth-century German literature* (Vol. 8). Berlin, Germany: Walter de Gruyter.
Honda, K., Hirai, H., Masaki, S., & Shimada, Y. (1999). Role of vertical larynx movement and cervical lordosis in F0 control. *Language and Speech*, *42*, 401–411.
Isserlis, S. (2011, October 27). The cello's perfect partner: The human voice. *The Guardian*. Retrieved from http://www.theguardian.com/music/2011/oct/27/steven-isserlis-voice-and-cello-series
Juslin, P. N. (2000). Vocal expression and musical expression: Parallels and contrasts. In A. Kappas (Ed.), *Proceedings of the Sixteenth Conference of the International Society for Research on Emotions* (pp. 281–284): Quebec City, Canada: ISRE Publications.
Juslin, P. N., Harmat, L., & Eerola, T. (2014). What makes music emotionally significant? Exploring the underlying mechanisms. *Psychology of Music*, *42*, 599–623.
Juslin, P. N., & Laukka, P. (2003a). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*, 770–814.
Juslin, P. N., & Laukka, P. (2003b). Emotional expression in speech and music. *Annals of the New York Academy of Sciences*, *1000*, 279–282.
Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, *31*, 559–575.
Keller, P. E. (2012). Mental imagery in music performance: Underlying mechanisms and potential benefits. *Annals of the New York Academy of Sciences*, *1252*, 206–213.
Kendall, R. A., Carterette, E. C., & Hajda, J. M. (1999). Perceptual and acoustical features of natural and synthetic orchestral instrument tones. *Music Perception*, *16*, 327–363.
Lakatos, S. (2000). A common perceptual space for harmonic and percussive timbres. *Perception & Psychophysics*, *62*, 1426–1439.

Lawson, C., & Stowell, R. (1999). *The historical performance of music: An introduction*. Cambridge, UK: Cambridge University Press.
Levy, D. A., Granot, R., & Bentin, S. (2001). Processing specificity for human voice stimuli: Electrophysiological evidence. *Neuroreport*, *12*, 2653–2657.
Levy, D. A., Granot, R., & Bentin, S. (2003). Neural sensitivity to human voices: ERP evidence of task and attentional influences. *Psychophysiology*, *40*, 291–305.
Li, W., Chen, J.-M., Smith, J., & Wolfe, J. (2015). Vocal tract resonances and the timbre of the saxophone. *Acta Acustica united with Acustica*, *101*, 270–278.
Mithen, S. (2005). *The singing neanderthals: The origins of music, language, mind and body*. London, UK: Weidenfeld & Nicolson.
Mithen, S. (2009). Holistic communication and the coevolution of language and music: Resurrecting an old idea. In R. Botha & C. Knight (Eds.), *The prehistory of language* (pp. 58–76). Oxford, UK: Oxford University Press.
Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35*, 250.
Novembre, G., Ticini, L. F., Schütz-Bosbach, S., & Keller, P. E. (2012). Distinguishing self and other in joint action: Evidence from a musical paradigm. *Cerebral Cortex*, *22*, 2894–2903.
Patel, A. D. (2008). *Music, lanugage, and the brain*. Oxford, UK: Oxford University Press.
Patel, A. D., & Iversen, J. R. (2003). Acoustic and perceptual comparison of speech and drum sounds in the North Indian tabla tradition: An empirical study of sound symbolism. In J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)* (pp. 925–928). Barcelona, Spain: Universitat Autònoma de Barcelona.
Preston, S. D., & De Waal, F. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, *25*, 1–20.
Price, U., & Lauder, T. D. (1842). *The picturesque: With an essay on the origin of taste and much original matter*. Edinburgh, UK: Caldwell, Lloyd & Co.
Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, *9*, 129–154.
Read, G. (1993). *Compendium of modern instrumental techniques*. Westport, CT: Greenwood Press.
Reilly, E. R. (1997). Quantz and the transverse flute: Some aspects of his practice and thought regarding the instrument. *Early Music*, *25*, 429–438.
Reuter, C. (2002). *Klangfarbe und Instrumentation*. Frankfurt, Germany: Peter Lang.
Sammler, D., Koelsch, S., Ball, T., Brandt, A., Elger, C. E., Friederici, A. D., … Wellmer, J. (2009). Overlap of musical and linguistic syntax processing: Intracranial ERP evidence. *Annals of the New York Academy of Sciences*, *1169*, 494–498.
Sammler, D., Koelsch, S., Ball, T., Brandt, A., Grigutsch, M., Huppertz, H.-J., … Elger, C. E. (2013). Co-localizing linguistic and musical syntax with intracranial EEG. *Neuroimage*, *64*, 134–146.
Sarris, H., & Tzevelekos, P. (2008). "Singing like the gaida (bagpipe)": Investigating relations between singing and instrumental playing techniques in Greek Thrace. *Journal of Interdisciplinary Music Studies*, *2*, 33–57.
Scherer, K. R. (1995). Expression of emotion in voice and music. *Journal of Voice*, *9*, 235–248.
Schubert, E., & Fabian, D. (2014). A taxonomy of listeners' judgments of expressiveness in music performance. In D. Fabian, R. Timmers, & E. Schubert (Eds.), *Expressiveness in music performance: A cross cultural and interdisciplinary approach* (pp. 283–303). Oxford, UK: Oxford University Press.
Schubert, E., & Wolfe, J. (2013). The rise of fixed pitch systems and the slide of continuous pitch: A note for emotion in music research about portamento. *Journal of Interdisciplinary Music Studies*, *7*, (1–2), 1–28.
Smith, R., & Mercer, D. (1974). Possible causes of woodwind tone colour. *Journal of Sound and Vibration*, *32*, 347–358.
Standards Secretariat (1994). American National Standard Acoustical Terminology (12.41), ANSI S1.1–1994 (R2004) C.F.R. Melville, NY: Acoustical Society of America.
Steiner, D. (2013). The Gorgons' lament: Auletics, poetics, and chorality in Pindar's Pythian 12. *American Journal of Philology*, *134*, 173–208.

Sulzer, J. G., Baker, N. K., & Christensen, T. S. (1995). *Aesthetics and the art of musical composition in the German enlightenment: Selected writings of Johann Georg Sulzer and Heinrich Christoph Koch*. Cambridge, UK: Cambridge University Press.

Sundberg, J. (1989). Synthesis of singing by rule. In M. Mathews & J. Pierce (Eds.), *Current directions in computer music research* (pp. 45–55). Cambridge, MA: MIT Press.

Swerdlin, Y., Smith, J., & Wolfe, J. (2010). The effect of whisper and creak vocal mechanisms on vocal tract resonances. *The Journal of the Acoustical Society of America*, *127*, 2590–2598.

Tarnopolsky, A., Fletcher, N., Hollenberg, L., Lange, B., Smith, J., & Wolfe, J. (2005). The vocal tract and the sound of a didgeridoo. *Nature*, *39*, 436.

Tarnopolsky, A. Z., Fletcher, N. H., Hollenberg, L. C., Lange, B. D., Smith, J., & Wolfe, J. (2006). Vocal tract resonances and the sound of the Australian didjeridu (yidaki) I. Experiment. *The Journal of the Acoustical Society of America*, *119*, 1194–1204.

Thayer, R. C. (1974). The effect of the attack transient on aural recognition of instrumental timbres. *Psychology of Music*, *2*, 39–52.

Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, NJ: Prentice Hall.

Traube, C., & D'Alessandro, N. (2005). Vocal synthesis and graphical representation of the phonetic gestures underlying guitar timbre description. In *8th International Conference on Digital Audio Effects (DAFx'05)* (pp. 104–109), Madrid, Spain.

Traube, C., & Depalle, P. (2004). Phonetic gestures underlying guitar timbre description. In S. D. Lipscomb, R. Ashley, R. O. Gjerdingen & P. Webster (Eds.), *Proceedings of the 8th International Conference on Music Perception and Cognition, Northwestern University, Evanston, Illinois, August 3–7* (pp. 658–661). Adelaide, Australia: Causal Productions.

Traube, C., & Depalle, P. (2004b). Timbral analogies between vowels and plucked string tones. In IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'04) (pp. 293–296 vol. 294). New York, NY: IEEE.

Turner, D. W., & Moore, A. (1852). *The odes of Pindar*. London, UK: Henry G. Bohn.

Wedin, L., & Goude, G. (1972). Dimension analysis of the perception of instrumental timbre. *Scandinavian Journal of Psychology*, *13*(3), 228–240.

Weiss, M. W., Trehub, S. E., & Schellenberg, E. G. (2012). Something in the way she sings enhanced memory for vocal melodies. *Psychological Science*, *23*, 1074–1078.

Wolfe, J. (2002). Speech and music, acoustics and coding, and what music might be "for". In D. B. K. Stevens, G. McPherson, E. Schubert, & J. Renwick (Eds.), *Proceedings of the 7th International Conference on Music Perception and Cognition* (pp. 10–13). Adelaide, Australia: Causal.

Wolfe, J. (2007). Speech and music: Acoustics, signals and the relation between them. *Proceedings of ICoMCS December, Sydney, Australia* (pp. 176–179).

Wolfe, J., & Schubert, E. (2012). Stable, quantised pitch in singing and instrumental music: Signals, acoustics and possible origins. *Acoustics 2012 Fremantle: Acoustics, Development and the Environment*. Perth, Australia: The Australian Acoustical Society, Western Australian Division.