What Makes an Instrument Sound Sad?
Commentary on Huron, Anderson, and Shanahan

JONNA K. VUOSKOSKI [1]
University of Oxford

ABSTRACT: Huron, Anderson, and Shanahan investigated the hypothesis that instruments that are deemed most capable of expressing sadness would also be judged better able to generate acoustic features similar to those used to convey sadness in speech. The judgments of these acoustic features accounted for approximately half (51.3%) of the variance in the judgments of *sadness capacity*. I argue that the explanation rate may have been curtailed by choices made in the operationalization of the acoustic features, the overlap and relatedness of three of the acoustic features used (*mumbling, dark timbre,* and *lowest pitch*), as well as the practical omission of features such as legato articulation. Furthermore, the method used by Huron and colleagues may have inflated the effect of cultural conceptions on the judgments of *sadness capacity*. I also argue that *low energy* – albeit a fundamental feature of sadness as an emotion – is not the meaningful factor underlying the set of acoustic features most correlated with *sadness capacity*. Instead, I suggest that the only acoustic variable significantly predicting evaluations of *sadness capacity – pitch-bending* – best reflected an instrument's capability to manipulate timbre, pitch, loudness and articulation in ways that match and exaggerate the features of sad vocal expression.

IN a correlational questionnaire study published in this issue, Huron, Anderson, and Shanahan investigated which acoustic features are associated with an instrument's judged capability to convey sadness. They hypothesized that those instruments that are judged most capable of expressing sadness would also be judged better able to generate acoustic features similar to those used to convey sadness in speech. This is a well-founded hypothesis, as previous studies (for a review, see Juslin & Laukka, 2003) have shown that similar acoustic features – slow tempo, low sound level, dark timbre, low pitch, small pitch and sound level variability, and slow tone attacks – are indeed used in both music and speech to convey sadness.

Huron and colleagues carried out two separate questionnaire studies to address the issue. In Study 1, graduate music students and faculty members were asked to evaluate – in light of their past music listening experience – how frequently each of 44 different instruments is used to convey sadness, and how well these instruments are able to produce a sad sound. In Study 2, undergraduate music students were asked to evaluate – again, in light of their knowledge and past experiences – the capacity of the same 44 instruments to produce certain acoustical effects; namely how easy it is to play very quietly or slowly on the different instruments, how easy it is to "bend the pitch" (i.e., play small intervals), and how easy it is to make a dark timbre or make the instrument sound like it's "mumbling". In addition, Huron and colleagues annotated the lowest pitch each instrument could produce. These six acoustic features were selected on the basis of previous studies on sad speech prosody.

The features that most strongly correlated with an instrument's capacity to produce a sad sound were *pitch-bending, mumbling, dark timbre*, and *lowest pitch*, respectively. However, a regression analysis revealed that *pitch-bending* was the only significant predictor of *sadness capacity*. The six acoustic features accounted for approximately half (51.3%) of the variance in *sadness capacity*, suggesting that additional factors might be at play. However, there may also have been factors in the study design and methods that restricted the potential explanation rate. In the following section I am going to address some of these potential factors.

RELEVANCE OF THE ACOUSTIC FEATURES CHOSEN

The appropriate operationalization of the acoustic features under investigation is of paramount importance in a study utilising questionnaires as its main method of data collection. Furthermore, this type of a method relies heavily on the participants' knowledge and past experiences with different instruments, as well as on their understanding of the questions asked. As the hypothesis of Huron and colleagues specifically states that an instrument's capability to convey sadness should be associated with its capability to produce acoustic features similar to those present in sad speech, it is reasonable that they approach the operationalization of these features from a "speech" point-of-view. However, Juslin and Laukka (2003) have already carried out an extensive review of the acoustic features utilized in the expression of emotion in both speech and music, and determined which of the features are shared between the two modalities.

In light of the results of Juslin and Laukka's review, especially the feature *mumbling* (i.e., "How easy is it on this instrument to make it sound like it's mumbling?", p. 35) appears somewhat problematic. In terms of sadness-related cues, Juslin and Laukka consider *low precision of articulation* (which could be seen as an equivalent to *mumbling*) as a speech-specific cue, while *legato articulation* is a similar music-specific cue. Thus, one might ask whether *mumbling* is an appropriate descriptor in a musical context, and whether it is related to legato articulation. A notable portion (66.7%) of the participants of Study 2 expressed some degree of confusion regarding the definition of *mumbling*. Based on the profile of correlations between the different features, as well as on the ratings given for the different instruments, it appears that the ratings regarding the ability to "mumble" reflect a combination of low pitch and low spectral centroid (i.e., dark timbre) rather than legato articulation. Thus, it may be that *mumbling* ended up being a redundant variable (i.e., not explaining any additional variance beyond dark timbre and low pitch), while legato articulation/slow tone attacks was almost entirely omitted from the set of predictors.

If the acoustic feature *mumbling* was not quite music-specific enough, then the feature *lowest pitch* was perhaps taken too far from its original context. Although sad expression has been associated with lower pitch (on average) in both speech and music (see e.g., Juslin & Laukka, 2003), this does not imply that there is a linear (negative) relationship between pitch and sadness. Indeed, previous studies have shown that sad melodies are only slightly lower in average pitch: Huron (2008) found that minor-key themes were only 1.1 semitones lower on average than major-key themes, and Bresin and Friberg (2011) demonstrated that, when asked to adjust the acoustical parameters (tempo, sound level, articulation, phrasing, pitch, instrument, and attack speed) of melodies to best communicate sadness, participants chose an average pitch of A#4. As very few of the instruments included in Huron and colleagues' study (e.g., the piccolo and the glockenspiel) are not able to play in this range, it seems that *lowest pitch* is not a very accurate measure of an instrument's ability to play in the "optimal range" for sad expression. This speculation is supported by the relatively low correlation ($r = .31$) between *lowest pitch* and *sadness capacity*. It is also worth noting that timbre and pitch perception are not independent, and that the timbral feature *brightness* is best explained by the spectral centroid of a tone (see e.g., Schubert & Wolfe, 2006). This means that low pitches will also have darker timbres, as is also indicated by the high correlation between *lowest pitch* and *dark timbre.*

Finally, the feature *pitch-bending* (i.e., "How easy is it on this instrument to bend the pitch? [to play small intervals]", p.34) seems to refer to the capability to play micro-intervals, as almost all the instruments included in the study (with the exception of a few percussion instruments) are able to produce the smallest standard interval in the Western music tradition – the minor second. One might ask why the capability to play micro-intervals (or to manipulate intonation) would be important for the portrayal of sadness? Granted, small pitch variability has been associated with sad expression in both music and speech (see e.g., Juslin & Laukka, 2003), but the Western classical music repertoire – sad music included – rarely requires the production of micro-intervals (although vibrato and glissandos are sometimes used as stylistic devices). Furthermore, Huron (2008) found that the average interval size for minor key themes (2.42 semitones) was only slightly smaller than for major key themes (2.46 semitones), thus not giving any indication that unusually small intervals would be characteristic of sad music. In comparison, "pitch-bending" could be seen as characteristic of speech in general – not just sad speech – in the sense that the pitch and pitch changes in speech are often fuzzy.

Somewhat surprisingly, *pitch-bending* turned out to be the only significant predictor of *sadness capacity*. It may be that the variable *pitch-bending* actually captured something essential about an instrument's capacity to convey sadness; something that comprises more than just the ability to produce

micro-intervals. The three instruments (besides voice) most highly rated on *sadness capacity* – the cello, violin, and viola – were also among the highest rated in terms of *pitch-bending* capability. What is characteristic of these bowed string instruments is that in addition to a very precise control of pitch, they also allow the variation of articulation, loudness, and timbre to an extent that resembles – and even exceeds – the capabilities of the human voice. If music and speech indeed utilize similar acoustic features to convey sadness, it wouldn't be surprising if the instruments judged as most capable of conveying sadness would be those that enable similar manipulations of pitch, timbre, loudness, and articulation as the human voice. Huron and colleagues do mention the possibility that "…acoustical attributes that convey a more voice-like sound are important, or even essential, for expressing or conveying sadness" (p. 39, although their main explanation focuses on a different kind of underlying factor.

LOW ENERGY AS THE UNDERLYING FACTOR?

As the four acoustic features with the strongest correlations with *sadness capacity – pitch-bending, mumbling, dark timbre*, and *lowest pitch* – were also strongly correlated with one another, Huron, Anderson, and Shanahan conclude that there may be a fundamental underlying factor that best explains an instrument's capacity to convey sadness. They suggest that this underlying factor may be *low physical energy*. Although this explanation seems logical, it is also somewhat circular: Low energy is – in addition to negative valence – a fundamental underlying factor of sadness itself (see e.g., Russell, 1980). In the context of emotions conveyed by music, it is the energy arousal (rather than valence) dimension that best predicts sadness ratings (e.g., Eerola & Vuoskoski, 2011). As low energy is a core feature of sadness as an emotion, the *low-energy* explanation does little to advance our understanding of the meaningful factors that make an instrument more or less capable of conveying sadness. Furthermore, the fact that many of the acoustic features were intercorrelated does not necessarily suggest a common underlying factor. While some of the features – such as *mumbling, dark timbre,* and *lowest pitch* – may have measured the same underlying construct, it could also be speculated that some features – such as *pitch-bending* and *dark timbre* – were correlated mainly because of reasons relating to Western instrument building tradition (i.e., those instruments that are able to produce very small intervals also happen to have darker timbre on average). Instead of low energy, I suggest that the findings of Huron, Anderson, and Shanahan are best explained by their original hypothesis; that an instrument's ability to convey sadness is related to its capability to manipulate timbre, pitch, loudness and articulation in ways that match and exaggerate the features of sad vocal expression. It appears that only the variable *pitch bending* sufficiently captured this capability in Huron and colleagues' study.

CONCLUSION

In summary, the explanation rate (R^2 = .51) obtained in the study by Huron and colleagues may have been higher if a more representative range of acoustic features with less overlap would have been employed. For example, slow voice onsets have been shown to be a reliable cue of sad vocal expression (for a review, see Juslin & Laukka, 2003), and there is reason to speculate that the variable *"mumbling"* used by Huron and colleagues did not adequately capture this feature in the context of music. Another factor that may have decreased the explanation rate concerns cultural conceptions about which instruments are more or less able to convey sadness. These conceptions are arguably influenced by the repertoire composed for different instruments, and the frequency in which different instruments are used to convey sadness. As mentioned by Huron and colleagues, the frequency in which a given instrument is used to convey sadness is not only dependent on that instrument's ability to convey sadness, but also on availability and cultural conventions. I would argue that this relationship is even more complex, and that – when asked to judge an instrument's capability to convey sadness in a questionnaire – people are inevitably influenced by cultural conceptions and conventions related to that instrument. This speculation is supported by the high correlation between the *sadness capacity* and *frequency* judgments. In other words, all the variance in the ratings of *sadness capacity* – as measured in Huron and colleagues' study – could not be explained even if an endless amount of acoustic features were included.

In conclusion, low energy can be considered a fundamental feature of sadness as an emotion, and thus it is logical that, to be able to successfully convey sadness, an instrument should successfully convey low energy. However, the vocal and musical expression of sadness is made of more than just low energy, albeit that "more" is difficult to dissect. In line with the original hypothesis of Huron, Anderson, and

Shanahan, I suggest that an instrument's capacity to express sadness is best explained by the ability to manipulate timbre, pitch, loudness and articulation in ways that match and exaggerate the features of sad vocal expression (see e.g., Juslin & Laukka, 2003). The variable *pitch-bending* best reflected this ability, although additional features (such as legato articulation, for example) may have provided a more comprehensive picture. The method selected by Huron and colleagues (i.e., questionnaire) may have inflated the effect of cultural conceptions on the judgments of *sadness capacity*. Although the cultural factors contributing to an instrument's judged capacity to convey sadness are an important and interesting phenomenon as such, they become a source of error when the objective is to investigate the acoustic features that contribute to *sadness capacity*. In future, it could be worthwhile to investigate instruments' capacities to convey sadness using controlled sound examples as stimuli, and to extract acoustic features using a computational (music information retrieval) approach. This kind of an investigation could provide a complimentary picture of the acoustic features contributing to *sadness capacity* while avoiding the problem of appropriate operationalization of the relevant acoustic features. Furthermore, by asking participants to rate the perceived sadness conveyed by sound examples, one can also reduce the effect of cultural variables and participants' level of expertise and past experiences on their ratings of *sadness capacity*. Granted, this type of approach is riddled with other types of problems and potential sources of error, but a combination of different methods – questionnaires, listening experiments, and music information retrieval – should provide a more comprehensive and reliable account of the phenomenon.

ACKNOWLEDGEMENTS

NOTES

[1] Correspondence concerning this commentary can be addressed to: Dr Jonna Vuoskoski, Faculty of Music, Oxford University, St Aldate's, Oxford OX1 1DB, UK. E-mail: jonna.vuoskoski@music.ox.ac.uk

REFERENCES

Bresin, R., & Friberg, A. (2011). Emotion rendering in music: range and characteristic values of seven musical variables. *Cortex*, *47*(9), 1068-1081.

Eerola, T., & Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music, 39*(1), 18-49.

Huron, D. (2008). A comparison of average pitch height and interval size in major-and minor-key themes: Evidence consistent with affect-related pitch prosody. *Empirical Musicology Review*, *3*(2), 59-63.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*(5), 770-814.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161-1178.

Schubert, E., & Wolfe, J. (2006). Does timbral brightness scale with frequency and spectral centroid? *Acta Acustica united with Acustica, 92*(5), 820-825.