Do minor thirds characterize the prosody of sad speech?

Andrés Buxó-Lugo (buxolugo@umd.edu) & L. Robert Slevc (slevc@umd.edu)

University of Maryland, College Park

This is an Accepted Manuscript of an article published by Taylor & Francis in Auditory Perception & Cognition on May 23rd, 2021 available online: http://www.tandfonline.com/10.1080/25742442.2021.1930465

Corresponding author:

Andrés Buxó-Lugo

Department of Psychology, University of Maryland, College Park

4094 Campus Dr,

College Park, MD, 20742

buxolugo@umd.edu

Dated: May 10, 2021

Abstract

Pitch can convey information about emotion in both spoken language and in music. Given this, do people use pitch to communicate emotion in similar ways across both domains? To investigate this question we look at intervals between the fundamental frequency (f_0) of adjacent syllables in emotional speech produced by actors. We first investigate whether descending minor third intervals are more prevalent in sad speech compared to other types of emotional speech, as has been reported previously. In these data, we see no evidence for descending minor thirds being characteristic of sad speech. In fact, we find little evidence for any specific musical intervals being associated with specific emotions in these longer sentences. We suggest that speakers might borrow emotional cues from music only when other prosodic options are infeasible.

One fascinating aspect of music is how effective it can be at conveying different emotions to a listener. For example, major chords and songs in major keys are often described as sounding "happy", while minor chords and songs in minor keys are described as sounding "sad" or conveying a "dark" mood (e.g., Cooke, 1959; Gagnon & Peretz, 2003; Kastner & Crowder, 1990). In speech, one of many ways in which speakers convey emotion is via prosody, the rhythmic and intonational patterns of speech. Interestingly, many of the acoustic cues used to convey information via prosody are the same as those that convey musical information (Juslin & Laukka, 2003). For example, pitch, rhythm, duration, and intensity are important cues in both music and speech prosody. In fact, musical prosody has also been found to signal emotional states (Meyer, 1956; Peretz, Gagnon, & Bouchard, 1998), and there is evidence that shared processes are involved in parsing linguistic and musical prosody, especially in musicians (Palmer & Hutchins, 2006; see also Heffner & Slevc, 2015; Lerdahl & Jackendoff, 1983). Given that music and speech can both convey emotional information via similar types of acoustic cues, a natural question is whether these domains convey the same information via the same overall means. That is, are there cues that convey the same emotions in both music and speech?

Some research indeed has found such parallels between emotional expression in music and speech. Juslin & Laukka (2003) investigated what patterns were found in previous studies of speech production and musical performance signaling emotion, and found parallels in the use of speech rate/tempo, voice intensity/sound level, spectral energy, f_0/pitch levels and movement, and voice onset/tone attack dynamics when signaling different emotions. For example, they summarize the cross-modal patterns relevant to happy speech and music as including high f_0/pitch level, much f_0/pitch variability, and rising f_0/pitch contours, while sad speech and music was characterized as including low f_0/pitch level and little f_0/pitch variability. In terms of musical

intervals, Bowling et al. (2010) found that the spectral information associated with excited speech was most similar to spectra found in major musical intervals, while the spectra associated with more subdued speech was similar to that found in some minor musical intervals (cf. Huron, 2008).

Some studies have even singled out specific pitch intervals as signaling emotions similar to how they function in music. For example, Curtis & Bharucha (2010) found that actors producing simple 2-syllable utterances (e.g., "okay") produced more minor third pitch intervals when conveying sadness than when conveying other emotions (happiness, anger, and pleasantness). Because minor third intervals are often described as sounding sad (Cooke, 1959), this finding provides some evidence for a parallel usage of pitch intervals in both music and speech prosody. However, it is worth noting that neutral two-syllable words provide a highly restricted linguistic context in which to convey emotions. Indeed, some studies have claimed that in similarly restrictive contexts, speakers might take advantage of the parallels between music and speech in order to better express the message they are trying to convey. For example, Day-O'Connell (2013) found an abundance of minor third intervals in stylized interjections (e.g., "Dinner!") and suggested that speakers essentially use sung-speech to better convey an intended meaning (see also Day-O'Connell, 2010, for evidence that minor thirds characterize the intonation of knock-knock jokes). This could be an intentional choice, for example, speakers might essentially choose to "break into song" in order to soften an imperative or request. Alternatively, this might reflect a more implicit association between singing and playfulness (see Day-O'Connell, 2013, for discussion). In either case, these findings reveal an intriguing parallel between minor thirds in music and speech, but it remains unclear if speakers use specific intervals to convey emotional information in broader contexts.

Understanding how and in what contexts specific musical pitch intervals occur in emotional speech can provide important information about both the communication of emotion and the connections between music and speech. Additionally, deciphering these patterns can shed light on how speech prosody works in general. Speech prosody is relevant in the communication of lexical, syntactic, semantic, and pragmatic information, but the cues through which prosody signals this information are often hard to single out, in part because of the large degree of variability between different contexts (e.g., Buxó-Lugo, Toscano, & Watson, 2018). If talkers are targeting specific pitch intervals to communicate emotion at the same time as they communicate other information via prosody, this could help us understand an important systemic source of variability in prosodic production, as emotional prosody is not often considered when investigating other applications of prosody.

If the prevalence of specific pitch intervals in speech is linked to specific emotions quite generally, this would support deep links between emotion in music and speech and could have important implications for the evolution of prosodic and musical emotional communication (cf. Curtis & Bharucha, 2010). Alternatively, if these pattens emerge only in specific contexts (e.g., in short two-word utterances), then this would instead suggest a type of creative "borrowing" of cues across domains, such that speakers sometimes rely on musical cues to express emotion in speech (cf. Day-O'Connell, 2013). To investigate these possibilities, we used a corpus of acted emotional speech (Livingstone & Russo, 2018) to explore what pitch interval patterns are found in emotional speech in longer utterances than have been investigated before. We begin by testing whether we find the same pattern that was reported by Curtis & Bharucha (2010) in these longer sentences (specifically, whether we find the descending minor third interval overrepresented in

sad speech). We then explore whether there are any other informative patterns we can find in the usage of pitch intervals that reliably signal specific emotions.

Methods

Data

The present study made use of the Ryerson Audio-Visual Database Speech and Song (RAVDESS; Livingstone & Russo, 2018). This corpus contains audio and video recordings of 24 professional actors speaking (or singing, though singing is not analyzed here) sentences while channeling seven different emotions at both low and high emotional intensities. These productions were of two emotionally-neutral sentences:

1. Kids are talking by the door.
2. Dogs are sitting by the door.

Importantly for our purposes, the structure of these sentences makes it so that speakers are less constrained as to their usage of prosody compared to the two-syllable, one-word utterances used in the studies described above. The fact that both sentences have the same number of words and syllables and the same syntactic structure makes it easy to look for the same patterns in both sentences. Importantly, Livingstone & Russo (2018) found that listeners were highly accurate when judging recordings for the intended emotional content, suggesting that the actors were indeed successful in conveying the desired emotions. For the present study, we made use of only the audio spoken recordings, which resulted in a total of 1436 sentences from 24 actors, each of whom channeled eight emotions twice for each of two sentence frames, at two different emotional intensities each[1].

[1] Because it was not possible to convey a 'neutral' emotion at different emotional intensities, this emotion was only included in the low intensity condition. For other emotions, it is reasonable to think that intensity will have an impact in the patterns. We investigated the impact of dividing these analyses by intensity. While some intervals did

In order to extract the relevant pitch information from the original recordings, the recordings were transcribed at the word level using the Gentle forced-aligner (Ochshorn & Hawkins, 2017) - a program trained on speech to automatically determine the onsets and offsets of words - which output Praat (Boersma & Weenink, 2020) textgrid files with estimated word onsets and offsets. These transcriptions were then manually corrected, and the one bisyllabic word in each sentence ("talking" and "sitting") was manually divided into two syllables. This resulted in a syllable-level transcription of all the relevant recordings. Syllables were chosen as the appropriate scope to investigate because each syllable had pitch values that were fairly consistent (f_0 can only be measured from voiced sounds) and yet varied between each other in relatively discrete ways. A Praat script was then run which made use of these transcriptions along with the recordings to extract the duration and mean f_0 values for each syllable (see Figure 1 for an example).

vary within the same emotion depending on intensity profiles, these changes did not meaningfully impact the conclusions of this paper. As such, we collapse between intensity profiles for simplicity.

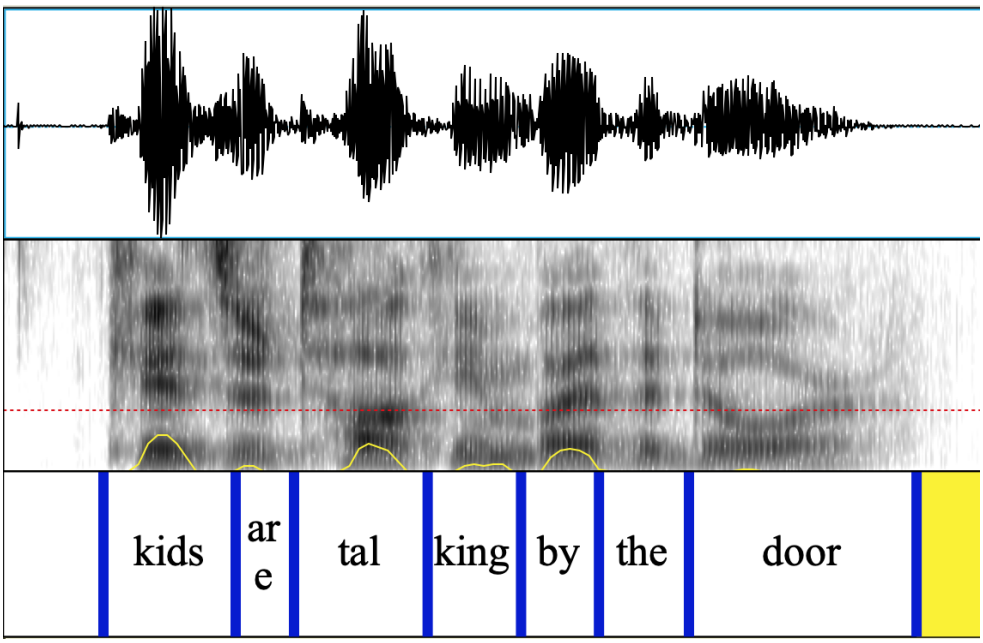

Figure 1: Example transcription of one of the RAVDESS recordings. The top panel shows the waveform of the utterance. The middle panel shows the spectrogram. The bottom panel shows the transcription with the marked starting and ending points for each syllable.

When calculating pitch intervals, we chose to explore intervals between adjacent syllables as a starting point, as this most closely mirrors previous studies. f_0 intervals between adjacent syllables were converted to semitones by using the following formula:

$$1200 * log_2(\frac{syllable\ f0}{preceding\ syllable\ f0})/100$$

The conversion of raw f_0 to semitones is important because semitones are directly analogous to musical pitch intervals (e.g., 3 semitones is equivalent to a minor third interval, 4 semitones to a major third, etc.). In order to finalize the conversion of raw f_0 values to musical intervals, the calculated semitone intervals were rounded to the nearest whole (with .5 rounded up). Lastly, intervals larger than an octave (12 semitones) were excluded from analysis, as were mean f_0 values below 50 or above 500Hz, as these were rare and likely due to measurement error. The acoustic measurements used for this study can be found at https://osf.io/ystgw/.

Minor thirds in emotional speech

We began by investigating whether we find a similar pattern to that found in Curtis & Bharucha (2010): that descending minor third patterns were noticeably more common in sad speech than other types of emotional speech. As a first test, we followed the original study in looking at the distributions of intervals used throughout the utterance under each type of emotional speech. As can be seen in Figure 2, it was not the case that sad speech contained mostly descending minor third intervals between syllables. In fact, descending minor thirds made for only 8.0% of intervals in sad speech, making them less common than descending minor seconds (22.0% of intervals), no movement (19.1%), descending major seconds (14.8%), and ascending minor seconds (10.7%). A logistic regression model with minor third as a dependent variable and emotion and intensity as independent variables (dummy coded with Neutral emotion as the comparison group) confirms these observations (Table 1).

Table 1: Summary of logistic regression model predicting the prevalence of minor third intervals between adjacent syllables as a function of emotion and intensity.

	Estimate	Std. Error	Z value	p value
Intercept	-2.375	0.149	-15.905	$p < .001$
Calm	-0.224	0.193	-1.162	$p = 0.245$
Happy	0.061	0.185	0.328	$p = 0.743$
Sad	-0.127	0.191	-0.666	$p = 0.506$
Angry	0.231	0.182	1.271	$p = 0.204$
Fearful	-0.246	0.193	-1.272	$p = 0.203$
Disgust	0.263	0.182	1.447	$p = 0.148$
Surprised	-0.054	0.188	-0.288	$p = 0.773$
Intensity	0.055	0.080	0.697	$p = 0.486$

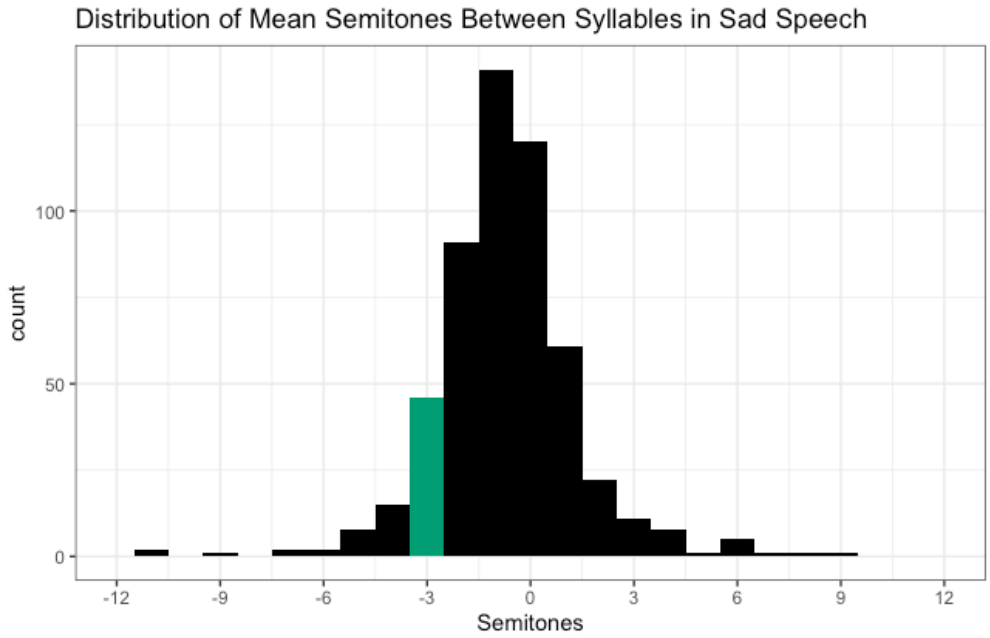

Figure 2: Histogram of semitone intervals found in sad speech recordings from the RAVDESS speech database. Descending minor third intervals (-3 semitones) are highlighted.

Of course, it is unlikely that speakers would use a descending minor third between all (or even most) syllables in a sentence. In addition, the hypothesis is not that sad speech contains *mostly* minor third intervals, but rather that the minor third interval is more likely to occur in sad speech than other types of emotional speech. As such, we tested whether there was an abundance of descending minor third intervals in sad speech. However, as can be seen in Figure 3, it does not appear to be the case that minor thirds were especially characteristic of sad speech. In fact, minor thirds were more common in happy, angry, and disgusted speech than they were in sad speech (this numerical pattern can also be seen in the model results in Table 1). Instead, we find that descending minor *seconds* a to be the most common interval found in sad speech (see Figure 2, above)[2].

[2] Given this finding, an additional logistic regression model was run to test what emotions best predict the likelihood of descending minor *second* intervals. With Neutral emotion used as the comparison group, the model found that happy, angry, disgusted, and surprised speech were significantly less likely to contain descending minor second intervals than neutral speech, whereas sad, fearful, and calm speech did not differ from neutral speech.

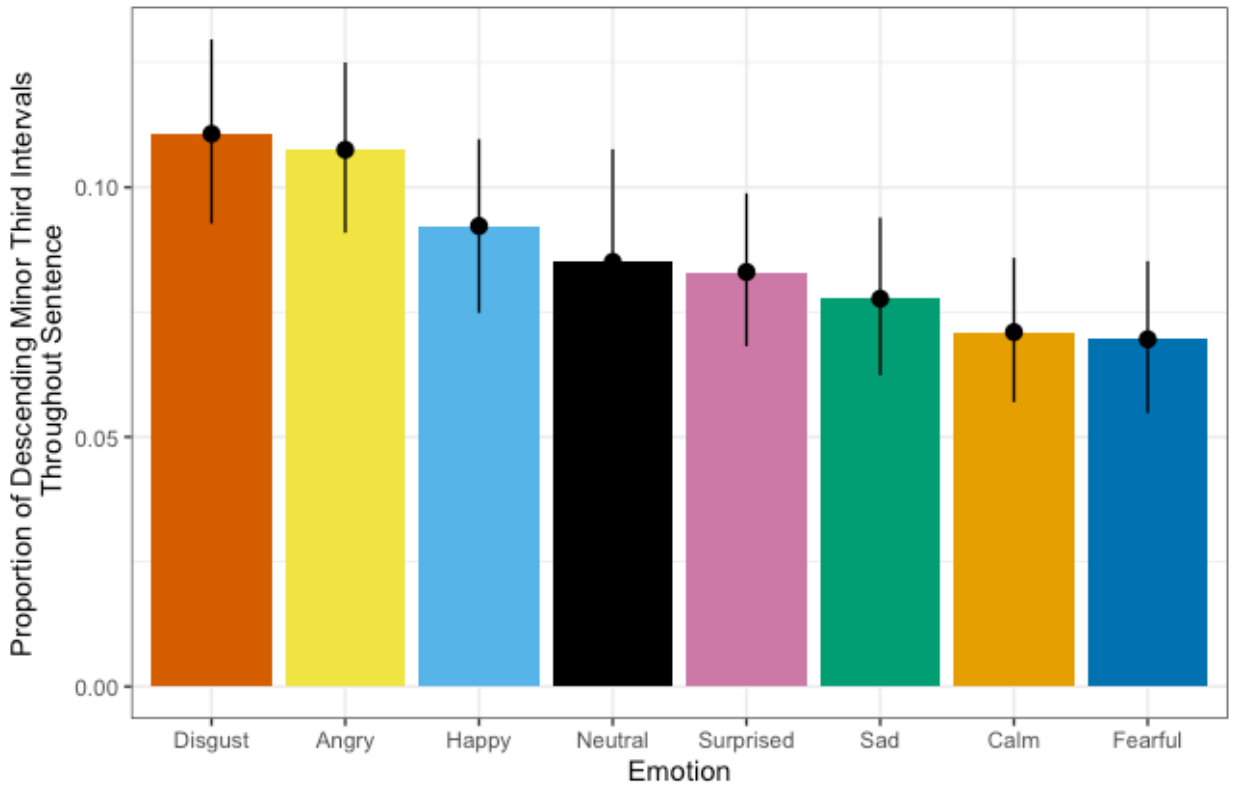

Figure 3: Proportion of descending minor third intervals throughout sentences for different emotions. Note that sentences were produced equally often in each emotional condition.

One likely reason for the difference between these findings and those in Curtis & Bharucha (2010) is the linguistic context in which speakers could use pitch to express emotion. More specifically, the actors in Curtis & Bharucha's (2010) study only had a single two-syllable word to convey emotional information, whereas the actors in the RAVDESS materials had multiple words and syllables where they could modulate pitch as well as other cues (pauses, durations, intensity, etc.). Nevertheless, it is important to think about why a minor-third / sadness relationship might occur only in bisyllabic contexts, since it is certainly still possible to use minor third intervals in longer sentence contexts (and in fact, speakers have more opportunities to do so). We speculate on reasons for these differences between findings in the Discussion.

Broader explorations of musical intervals in emotional speech

Although minor thirds do not seem to characterize sadness in these materials, a broader question is whether any patterns of musical pitch intervals are informative as to different types of emotional speech. Specifically, do we find *any* pitch intervals that specifically signal specific emotions, and are there any especially informative areas throughout the sentence for these intervals to appear?

To investigate this, we used the same data as before to train decision tree classifiers (Therneau & Atkinson, 2019) that take pitch intervals between each syllable as an input and partition the data based on information content to provide a best guess as to what type of emotion the sentence was supposed to convey. This can be thought of as modeling what questions would be best to ask to win a game of *Guess Who?*, but for guessing emotions given only pitch intervals throughout a sentence.

An illustrative decision tree is shown in Figure 4. The most obvious pattern we see from this analysis is that there are no obvious patterns; that is, there do not seem to be particular pitch intervals associated with any of the emotions investigated in this data set. Rather, the best decision trees make use of complex interactions between different locations and interval sizes.

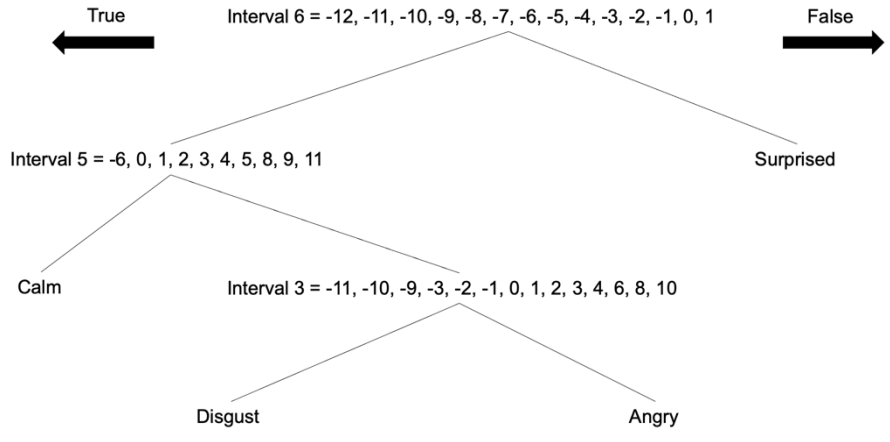

Figure 4: Schematic of the decision tree structured that emerged after training on our data. At each split, the tree divides the data based on specific values for one factor. "Interval" refers to syllable intervals (of six total intervals in

each sentence), and their values are specific semitones. Thus the first decision the model makes is based on the final (6^{th}) interval: if that is any ascending value greater than 1 semitone, the model guesses the sentence was expressing surprise. If not, it moves to the second decision, which is based on the fifth interval, and so on.

One exception to this pattern is that the presence of abnormally large pitch intervals at the end of a sentence strongly signaled that the sentence was meant to convey surprise. In fact, the first and most informative split performed by our decision tree is over whether the last interval in the utterance ascends by two or more semitones. If it does, the model classifies the utterance as being surprised speech. Otherwise, the model performs additional splits checking for esoteric semitone patterns at different intervals. In other words, there seems to be enough of a pattern between semitone intervals and surprised speech to make a decent "surprised speech detector" based on semitone intervals, although crucially, the pattern did not emerge from specific, meaningful relationships between surprised speech and music. Rather, it appears that a strong cue for surprised speech is a sharp rise in pitch towards the end of an utterance, and this is not dependent on the specific interval of that rise.

Thus, overall there does not appear to be a simple pattern between usage of specific pitch intervals and emotions. In fact, even considering the complex interactions represented in the decision trees, the classifiers were not particularly impressive in their performance. A model that always guesses the same emotion yields an 86.6% error rate. By comparison, the first split in our decision tree results in an error rate of 89.7%, somewhat worse than a brute force method. Subsequent splits yield increasingly smaller gains, resulting in an error rate of 82.9% before the model reaches the point where any benefit from further splits are too small to justify. While the first split is interpretable, it is likely that the subsequent splits are due to overfitting and finding marginal gains in noise. Given that listeners from the original RAVDESS study could accurately identify the emotion under which these utterances were produced, it seems likely that pitch intervals between syllables were not the cue through which they reached their conclusions.

Discussion

This study began with an observation on the parallels between musical and linguistic cues to emotion. Of special interest was the possibility of speakers using pitch intervals that are associated with specific emotions when used in music. While some studies have found this pattern, they have investigated only highly constrained linguistic contexts (in particular, just two-syllable words). Here, when assessing pitch patterns in full sentence contexts, we found no evidence for an association between minor thirds and sadness (contra Curtis & Bharucha, 2010). In fact, we found little evidence for relationships between any pitch intervals in speech with any particular emotions.

The contrast between these results and previous evidence for minor thirds in sad speech might mean that descending minor thirds do not actually signal sadness by default, but rather, in highly constrained contexts, they might be the most effective way in which to convey the emotion. That is, there may be relatively few ways to effectively express sadness (or any other particular emotion) when only producing a neutral word like "okay". Thus, a speaker who wants to express an emotion while saying "okay" might (perhaps implicitly) co-opt musical cues to emotion by essentially singing the word, thereby relying on an association between minor thirds and sadness *in music*. This possibility is similar to Day-O'Connell's (2013) proposal that speakers make use of musical cues in language to better convey a message in certain contexts.

A related possibility is that speakers do not use descending minor thirds in longer sentences of sad speech because it goes against other more general patterns that are associated with sad speech. For example, multiple studies have found that sad speech was partially characterized by diminished f_0 variations (Banse & Scherer, 1996; Breitenstein, Van Lancker, & Daum, 2010; Stolarski, 2015; Juslin & Laukka, 2003). In fact, our own data show that the most

common intervals found in sad speech were a descending minor second (1 semitone) and no movement at all, with the likelihood of interval presence decreasing as size increases (see Fig. 2). A reliance on minor third intervals would go against this overall pattern (as pitch variations in speech tend to be relatively small anyway: the most common interval in this dataset was a 1 semitone decrease, followed by no change at all), and so might actually be less effective at conveying sadness than more monotone speech throughout a sentence. Indeed, in these data, we do find that intervals found in sad speech were overall less wide than those found in other types of emotional speech, such as Anger, Disgust, and Surprise (see Figure 5 below).

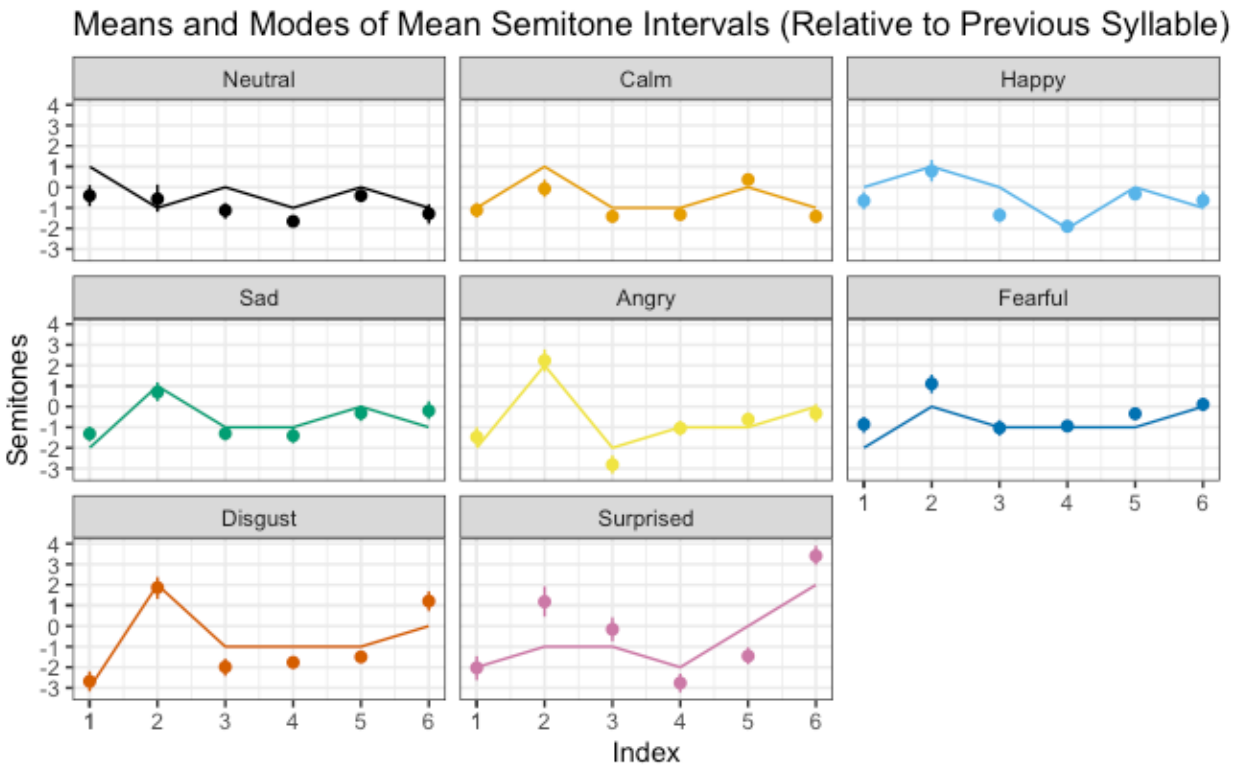

Figure 5: Means (points) and modes (lines) of mean semitone intervals throughout the sentence (note that results are similar when examining median values). The Index on the X axis represents the numbered interval or the syllable following the first syllable. Each of these intervals is calculated relative to the previous syllable of the sentence. Error bars represent bootstrapped 95% Confidence Intervals.

An additional reason for why specific pitch intervals might not be used for communicating emotion in more general contexts is that pitch is already a contested and

notoriously variable cue that is modulated to signal other types of linguistic information. Because of this, listeners sometimes rely on non-pitch cues (ignoring pitch) even when parsing information that is reliably communicated via pitch (e.g., Cole, Mo, & Baek, 2010; Buxó-Lugo & Watson, 2016). Additionally, information that often is reliably conveyed via pitch is done via pitch *accents* which are thought to be defined by overall pitch movement, rather than by targeting specific intervals. For example, an assumption in the ToBI framework for prosody transcription (Beckman & Pierrehumbert, 1986) is that there are pitch targets broadly defined as High or Low, but not specified at the interval level. Targeting specific intervals might then tend to interfere with other aspects of speech prosody. Lastly, even if one were to use pitch intervals communicatively in speech (and to use them reliably, which might be unlikely; e.g., Pfordresher & Brown, 2017), it is not obvious that listeners regularly *attend* to specific pitch intervals in speech, thus it might not be particularly effective. For example, in the speech-to-song illusion, clear musical patterns become obvious in listened speech only after multiple repetitions of a recording (e.g., Deutsch, Henthorn, & Lapidis, 2011; Falk, Rathcke, & Dalla Bella, 2014). This suggests that listeners are not usually particularly attuned to the *specific* musical pitches or intervals represented in speech (perhaps instead intending primarily to prosodically-relevant information about pitch contour). Rather, listeners might be able to start attending to musically-relevant information over repetitions, somewhat analogous to how listeners can start to extract speech-relevant information in sine-wave speech (Remez et al., 1981). For these reasons, relying on specific pitch intervals might not be useful outside of some very specific scenarios.

All together, these data do not support the existence of significant cross-domain mappings of pitch intervals between music and speech. For the reasons described in this discussion, we believe it is unlikely to find such patterns in relatively unconstrained contexts of

speech. Nevertheless, it is clear that talkers vary their pitch in interesting ways when talking under different emotional states (cf. the consistently large rising interval associated with surprise in the data reported above). Although relative interval size or direction may be too coarse a cue to reliably distinguish between emotions, they may combine with other speech cues such as duration and intensity to support listeners' reliable identification of emotions from these speech samples. Future research should study these patterns to better understand how paralinguistic factors interact with linguistic factors that make use of similar cues.

While these data do not support deep cross-domain mappings between pitch intervals in music and speech, previous work suggests that musical pitch intervals *can* emerge in speech when other prosodic cues are infeasible (e.g., when one only has two syllables to work with; Curtis & Bharucha, 2010; Day-O'Connell, 2013). This might suggest that, rather than reflecting inherent cross-domain mappings, these patterns reflect a sort of cross-domain "borrowing": speakers might be able to draw on various types of cues – including musical regularities – to express emotion in speech. While this sort of opportunistic cross-domain borrowing is not particularly consistent with deep evolutionary links between emotional expression in language and music, it is certainly consistent with the creative and flexible use of our powerful linguistic and musical abilities to express and communicate emotion.

Acknowledgments

We would like to thank Mattson Ogg, Peter Pfordresher, Jan Philipp Röer, Michael Hall, and an anonymous reviewer for helpful feedback and comments.

References

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70, 614–636.

Bowling, D. L., Gill, K., Choi, J. D., Prinz J., & Purves, D. (2010). Major and minor music compared to excited and subdued speech. *The Journal of the Acoustic Society of America, 127(1),* 491-503.

Beckman, M. E., & J. B. Pierrehumbert. (1986) Intonational structure in Japanese and English, *Phonology Yearbook 3*, 255-309.

Boersma, P. & Weenink, D. (2021). Praat: doing phonetics by computer [Computer program]. Version 6.1.38, retrieved 2 January 2021 from http://www.praat.org/

Breitenstein, C., Van Lancker, D., & Daum, I. (2001). The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample, *Cognition and Emotion, 15,* 57-79,

Buxó-Lugo, A., Toscano, J. C., & Watson, D. G. (2018). Effects of participant engagement on prosodic prominence. *Discourse Processes, 55(3),* 305-323.

Buxó-Lugo, A. & Watson, D. G. (2016). Evidence for the influence of syntax on prosodic parsing. *Journal of Memory and Language, 90,* 1-13.

Cole, J., Mo, Y., & Baek, S. (2010). The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech, *Language and Cognitive Processes, 25: 7*, 1141-1177.

Cooke, D. (1959). *The Language of Music.* London: Oxford University Press.

Curtis, M. E., & Bharucha, J. J. (2010). The minor third communicates sadness in speech, mirroring its use in music. *Emotion, 10*(3), 335–348.

Day-O'Connell, J. (2013). Speech, song, and the minor third: An acoustic study of the stylized interjection. *Music Perception: An Interdisciplinary Journal, 30,* 441-462.

Day-O'Connell, J. (2010). "Minor third who?": The intonation of the knock-knock joke. In *Speech Prosody 2010,* paper 990.

Deutsch, D., Henthorn, T., & Lapidis, R. (2011). Illusory transformation from speech to song. *Journal of the Acoustical Society of America, 129,* 2245-2252.

Falk, S., Rathcke, T., & Dalla Bella, S. (2014). When speech sounds like music. *Journal of Experimental Psychology: Human Perception and Performance.*

Gagnon, L. & Peretz, I. (2003) Mode and tempo relative contributions to "happy - sad" judgments in equitone melodies. *Cognition and Emotion , 17,* 25-40.

Heffner, C. & Slevc, L. R. (2015). Prosodic structure as a parallel to musical structure. *Frontiers in Psychology. 6.*

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*, Vol. 3, 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.

Kastner, M. P., & Crowder, R. G. (1990). Perception of the major/minor distinction: IV. Emotional connotations in young children. *Music Perception, 8,* 189–202.

Lerdahl, F., & Jackendoff, R. (1983). A Generative Theory of Tonal Music, 1st Edition. Cambridge, MA: MIT Press.

Livingstone S. R. & Russo F. A. (2018). The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PLoS ONE 13(5).*

Meyer, L. B. (1956). *Emotion and meaning in music.* Chicago: Chicago University Press.

Oschorn, R., & Hawkins, M. (2017). Gentle [Computer software]. Retrieved from https://lowerquality.com/gentle/

Palmer, C., and Hutchins, S. (2006). What is musical prosody? *Psychology of Learning and Motivation, 46,* 245-278.

Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: Perceptual determinants, immediacy, and isolation after brain damage. *Cognition, 68,* 111–141.

Pfordresher, P. Q., & Brown, S. (2017). Vocal mistuning reveals the origin of musical scales. *Journal of Cognitive Psychology, 29*(1), 35–52.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 22,* 947-949.

Stolarski, Ł. (2015). Pitch patterns in vocal expression of “happiness” and “sadness” in the reading aloud of prose on the basis of selected audiobooks. *Research in Language, 13,* 141-162.

Therneau, T. & Atkinson, B. (2019). rpart: Recursive Partitioning and Regression Trees. R package version 4.1-15. https://CRAN.R-project.org/package=rpart