Proceedings of the 3rd International Conference on Music & Emotion (ICME3), Jyväskylä, Finland, 11th - 15th June 2013. Geoff Luck & Olivier Brabant (Eds.)

MELODIC AND RHYTHMIC CONTRASTS IN EMOTIONAL SPEECH AND MUSIC

Lena Quinto, William Forde Thompson, Felicity Louise Keating

Psychology, Macquarie University, Australia
lena.quinto@mq.edu.au

Abstract

Many cues convey emotion similarly in speech and music. Researchers have established that acoustic cues such as pitch height, tempo, and intensity carry important emotional information in both domains. In this investigation, we examined the emotional significance of melodic and rhythmic contrasts between successive syllables or tones in speech and music, referred to as Melodic Interval Variability (MIV) and the normalized Pairwise Variability Index (nPVI). The spoken stimuli were 96 tokens expressing the emotions of irritation, fear, happiness, sadness, tenderness and no emotion. The music stimuli were 96 phrases, played with or without performance expression and composed with the intention of communicating similar emotions (anger, fear, happiness, sadness, tenderness and no emotion). The results showed that speech, but not music, was characterized by changes in MIV as a function of intended emotion. However, both speech and music, showed similar changes in nPVI. Emotional portrayals, both spoken and musical, had higher nPVI values than stimuli conveying "no emotion". The results suggest that MIV may function differently in speech and music, but that there may be similarities between the two domains with respect to nPVI.

Keywords: acoustic cues, prosody

1. Introduction

The communication of emotion in speech and music occurs through similar changes in acoustic cues such as intensity, tempo, timing and pitch (e.g., Juslin & Laukka). Both speakers and musicians change these cues to match the intended emotional signal. However, two broad sets of cues that have been difficult to compare in the domains of speech and music have been pitch and rhythmic patterns.

Pitch and timing form two very important aspects of music but do not appear to serve similar functions in speech. Pitch relationships are organized hierarchally in music. The perception and processing of pitch relations are central aspects of music cognition, including the sense key (e.g., Krumhansl, 1990). In speech, there are no similar analogues to the complex pitch relationships that are observed in music. Recent studies have tried to make comparisons. In music, the interval of a minor third may be associated with sadness and the interval of a major third may be associated with happiness. It has been observed that these same intervals may be mirrored in sad and happy speech prosody as signals of emotion (Bowling, Gill, Choi, Prinz, & Purves, 2010; Curtis & Bharucha, 2010). These studies suggest that there may be commonalities in the emotional expression of speech and music through pitch relationships. However, this work has been constrained by the use of a few intervals and does not extend to other pitch relationships.

A second important aspect of music is rhythm. Like pitch, rhythm is also organized hierarchally through meter. Changes in timing can act as a cue to emotion and are important in conveying musical interpretation and expressivity (Kendall & Carterette, 1990). In

speech, rhythm may not serve this same function as there may not be an analogue to expressivity in speech. However, rhythmic differences have been observed between languages and linguists have attempted to quantify and compare these differences (Low, Grabe, & Nolan, 2000).

Two new cues have recently been developed to quantify pitch and rhythmic contrasts in music and speech. Melodic Interval Variability (MIV) and the normalized Pairwise Variability Index (nPVI) assess the degree of contrast between successive pitches or durations. A low MIV or low nPVI indicates that successive changes are relatively uniform as compared to a high MIV or nPVI value.

MIV is calculated as the coefficient of variation (CV) of interval size (CV = SD/M). This requires determining the average interval size between successive tones or syllables and the standard deviation associated with these intervals. NPVI is calculated as the absolute value of the duration of a tone or syllable subtracted from the duration of the subsequent tone or syllable. This value is then divided by the average duration of these two values combined. The sum of these values is then multiplied by 100, divided by the number of elements (syllables or tones) less one.

$$\mathrm{nPVI} = \frac{100}{m-1} \times \sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{\frac{d_k + d_{k+1}}{2}} \right|,$$

These variables were initially used to assess pitch and rhythmic changes in speech. However, Patel, Iverson and Rosenberg (2006) have extended the use of these variables to compare changes in the music and speech of composers whose native languages differed. Patel et al. found that French speech has lower MIV and nPVI values than English speech and that music written by French composers has lower MIV and nPVI values than music written by English composers. This suggests melodic and rhythmic patterns found in speech are reflected in music.

In this study, we sought to determine whether MIV and nPVI would show a similar pattern of changes in *emotional* music and speech. Previous studies have shown that other cues change similarly in emotional speech and music, such as intensity, tempo, and articulation. The work of Patel et al. (2006) showed that MIV and nPVI change similarly in music and speech of different languages. Therefore, it might be possible that MIV and nPVI can also carry emotional information in both domains. Importantly, this could provide a novel way of comparing complex pitch and rhythmic cues that may communicate emotion.

2. Method

Spoken stimuli - Six male and seven female speakers provided samples of emotional speech. Speakers said seven different semantically neutral phrases, such as "The broom is in the cupboard and the lamp is on the table" with the intention of expressing the emotions of irritation, fear, happiness, sadness, tenderness and no expression.

Thirty-five first year university students judged the emotion that they believed was being conveyed. This allowed us to reduce the large number of stimuli down to the 96 most clearly decoded stimuli, balanced for speaker sex.

Music stimuli – 8 musicians (4 violinists and 4 vocalists) composed music expressing the same emotions as the speakers, with the exception that irritation (mild emotion) was replaced with anger (strong emotion). The compositions were limited to 7-9 notes in length. Musicians were recorded performing these stimuli. The stimuli were also digitally recorded on MIDI with the timbres of violin or voice. This removed performance expression or deviations introduced by the musician in pitch and timing. The MIDI version instead retained the relative pitch and timing changes indicated by the musicians through notation.

Acoustic analysis – Each syllable and every tone in the sentence and musical phrase was indicated with text grids (Praat, 2010). The text grids annotated the start and end time of each tone and allowed us to calculate the average pitch for each tone or syllable. Highly unstable pitches (glides) were ignored. MIV was calculated by first finding the musical interval dis-

Table 1: Shown are the mean MIV and nPVI values for each of the six emotions. Standard errors are shown in parentheses.

	Emotion					
	Irritation	*Fear*	*Happiness*	*Sadness*	*Tenderness*	*No emotion*
MIV speech	*86.25 (6.16)*	*76.22 (5.94)*	*63.63 (6.29)*	*97.07 (5.82)*	*84.52 (6.13)*	*83.82 (5.77)*
Interval Size Speech	*2.42 (0.12)*	*1.64 (0.15)*	*4.46 (0.29)*	*2.37 (0.46)*	*2.98 (0.32)*	*2.06 (0.52)*
MIV music	*74.49 (5.96)*	*71.32 (9.06)*	*74.52 (7.63)*	*67.58 (3.47)*	*63.57 (5.35)*	*66.38 (6.63)*
nPVI speech	*59.05 (2.04)*	*52.55 (1.97)*	*53.45 (2.09)*	*58.65 (1.93)*	*63.04 (2.03)*	*51.98 (1.91)*
nPVI music	*56.22 (6.61)*	*46.39 (6.28)*	*52.01 (3.72)*	*56.59 (3.91)*	*46.44 (4.56)*	*31.19 (4.59)*

tance *in semitone*s between successive syllables or tones. The mean interval distance and standard deviation was calculated for each phrase. The CV or MIV for each phrase was obtained by dividing the standard deviation by the mean interval distance. See Quinto, Thompson, Taylor (in press) and Thompson, Marin, Stewart (2012) for more details of the music and speech stimuli respectively.

3. Results

The data were treated separately for each domain (speech and music) and each dependent variable (MIV and nPVI). The spoken stimuli were subjected to a mixed linear effects model with speaker, sentence and emotion as fixed factors. Speaker and sentence were entered as factors because they were a source of variability. The musical stimuli were also subjected to a mixed linear effects model but with emotion, musician and mode of presentation as fixed factors. All the melodies varied with each musician.

Speech stimuli – The mixed linear effects model with emotion, sentence and speaker as fixed factors and MIV as the dependent variable, revealed a significant effect of emotion, $F(5,74) = 3.594$, $p = .006$. As Table 1 shows, the mean MIV for happiness was lower than for all the other emotions. This suggests that happiness was conveyed with the most consistent changes in pitch whereas the emotion of sadness had the least consistent changes in pitch. The main effect of sentence $F(6,74) = 1.211$, $p = .310$, and of speaker were not significant, $F(10,74) = 0.808$, $p = .621$.

The average interval size for each emotion was compared to MIV. The mixed linear effects model showed a main effect of emotion, $F(5,74) = 8.731$, $p < 0.001$. The emotions of fear and no emotional expression were associated with lower average interval size than the emotions of happiness and tenderness. This demonstrates that MIV and average interval size provide unique pitch based information. For average interval size, there was also a main effect of speaker, $F(10,74) = 2.834$, $p = 0.005$, demonstrating that individuals differed in the extent to which they varied the average interval size. There was no significant effect of sentence.

The mixed linear effects model with emotion, sentence and speaker as fixed factors and nPVI as the dependent variable, revealed a significant main effect of emotion, $F(5,74) = 4.178$, $p = 0.001$. As Table 1 shows, the mean nPVI for "no emotion" expressions was significantly lower as compared to all the other emotions. There was also a main effect of speaker, $F(10,74) = 3.337$, $p = 0.001$, demonstrating that different individuals used varying levels of durational contrasts in their speech. Across speakers, the average range of nPVI values varied from 48.67 to 60.95. There was also a main effect of sentence, $F(6,74) = 31.016$, $p < 0.001$ demonstrating that different sentences strongly influenced the durational contrasts

that speakers used. The average nPVI for these sentences was 55.52, *SD* = 12.14 (range = 44.76 – 69.88). Unfortunately, interactions between these variables could not be tested because these factors were not fully crossed.

Music stimuli - The mixed linear effects model with emotion, musician and mode of presentation as fixed factors and MIV as the dependent variable, revealed no significant effect of emotion, $F(5,74) = 0.634$, $p = .675$. While not significant, there does appear to be a weak trend showing that higher arousal emotions such as anger, fear and happiness had higher MIV values than low arousal emotions such as sadness and tenderness. At this point it is speculation, but it might be possible that in music, but not speech, MIV is associated with arousal. There was a significant main effect of musician, $F(7,74) = 5.454$, $p < .001$. This finding suggests that while musicians did not use MIV as a cue to emotion, there were differences in the consistency of pitch changes in the in musicians' compositions. Some musicians had greater changes in MIV than others. This ranged from average MIV values of 42.68 to 91.93 ($M = 69.42$, $SD = 25.99$). There was no significant effect of mode of presentation, $F(1,74) = 0.151$, $p = 0.698$. This is not surprising given that the same pitch information was used in both conditions.

The mixed linear effects model with emotion, musician and mode of presentation as fixed factors and nPVI as the dependent variable, revealed a significant effect of emotion, $F(5,74) = 4.24$, $p = .002$. As table 1 shows, "no emotion" expressions were associated with considerably lower nPVI values than the emotional expressions. This mirrors the finding observed for the speech stimuli. There was also a significant main effect of musician, $F(7,74) = 2.61$, $p = .017$. This finding suggests that there was considerable variability in how musicians used nPVI. The average nPVI values for musicians ranged from 36.82 to 59.20 ($M = 48.14$, $SD = 21.54$). The main effect of condition was also significant, $F(1,74) = 9.90$, $p = .002$. The mean nPVI was significantly higher in the live condition ($M = 54.03$, $SD = 22.48$) than in the deadpan condition ($M = 42.26$, $SD = 19.03$). Musicians introduced increased rhythmic contrasts when they were performing their own pieces relative to the notated durations. This suggests that nPVI might be associated with the expressiveness that is introduced by performers.

4. Discussion

The results of this study suggest that MIV may contribute to emotional communication in speech but not in music. The results also suggest that nPVI may act to differentiate emotional from non-emotional stimuli in both speech and music. Finally, our findings show that there is a high degree of variation in the use of both these cues by musicians and speakers. This last finding highlights the need to test multiple speakers and musicians.

MIV appeared to change with emotional intentions in speech but not in music. In speech, low MIV was associated with happiness and high MIV was associated with sadness. When communicating happiness, speakers alternated between more similar intervals than when expressing other emotions. The consistent and large interval changes may signal a greater amount of physiological control over the pitches that are produced. It is possible that such cues may be used to differentiate emotions that are otherwise similar with respect to pitch cues like interval size.

In music, MIV did not appear to signal emotions. One potential reason for this difference between music and speech is that the expression of emotion in music might be independent of consistency in pitch changes and guided by constraints in pitch relationships. Important pitch based cues may include mode (Hevner, 1935) which may function independently of MIV. It seems that musicians were able to write music expressing various emotions without necessarily showing similarities in the use of MIV in between emotions. Since there were differences in MIV values between musicians, this also suggests that there were other factors that influenced the manner in which they communicated the same emotions which were independent of MIV.

The findings for nPVI suggest that this cue could be important in differentiating emotional from non-emotional stimuli. Changing stimuli may encourage perceivers to attend to in-

formation whereas information that does not change might encourage habituation (Jones & Macken, 1993). By its vary nature, emotion acts to signal information to others. In this way, the rhythmic contrasts associated with nPVI in emotional stimuli may attempt to engage the listener.

In studies that have examined expressiveness and emotional communication through performance expression, a common finding is that the duration of sounded notes is lengthened or shortened relative to the notated values (Gabrielsson & Juslin; 1996; Thompson, Sundberg, Friberg & Frydén, 1989). That is, performers may not play tones evenly or exactly with the duration that has been indicated in the score. NPVI may be used as a measure to quantify these durational deviations and contrasts between tones.

Future studies may assess the perception of MIV and nPVI. Listeners can differentiate high and low nPVI (Hannon, 2009) but not much more is known. The extent to which listeners use these cues in differentiating emotional speech and non-emotional stimuli is not clear. For example, does low MIV in speech *signal* happiness to listeners? Does this occur in other languages?

Our work provides some evidence that MIV and nPVI may be involved in emotional decoding. It is possible that MIV can assist in emotional decoding in speech. NPVI appears to be a promising new cue to differentiate emotional versus non-emotional stimuli in both speech and music. NPVI may also be used to quantify the rhythmic changes that are associated with expressive performances.

References

Boersma, P., & Weenink, D. (2010). Praat: doing phonetics by computer (Version 5.1.28) [Computer program]. Retrieved 11 March 2010 from http://www.praat.org

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

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

Gabrielsson, A., & Juslin, P. N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music*, *24*, 68-91. doi:10.1177/0305735696241007

Hannon, E. E. (2009). Perceiving speech rhythm in music: Listeners classify origins of songs according to language of origin. *Cognition, 11*, 403-409.

Hevner, K. (1935). The affective character of the ma- jor and minor modes in music. *The American Journal of Psychology*, *47*, 103–118.

Jones D. M., & Macken, W. J. (1993). Irrelevant tones produce an 'irrelevant speech effect': Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 19*, 369-381.

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

Kendall, R. A., & Carterette, E. C. (1990). The communication of musical expression. *Music Perception, 8*(2), 129-164.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch.* New York, NY: Oxford University Press.

Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech, 43*, 377- 401.

Patel, A. D., Iversen, J. R., & Rosenberg, J. C. (2006). Comparing the rhythm and melody of speech and music: The case of British English and French. *Journal of the Acoustical Society of America, 119*, 3034-3047.

Quinto, L., Thompson, W. F., & Taylor, A. (in press). The contributions of compositional structure and performance expression to the communication of emotion in music. *Psychology of Music.*

Thompson, W.F., Marin. M.M. & Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences.* Advance online publication.

Thompson, W.F., Sundberg, J., Fryden, L., & Friberg, A. (1989). Rules for expression in the performance of melodies. *Psychology of Music, 17*, 63-82.

This work was recently published in Frontiers in Psychology. The citation for the full article is Quinto, L.R., Thompson, W.F., & Keating, F.L. (2013).

Emotional Communication in Speech and Music: The Role of Melodic and Rhythmic Contrasts. *Frontiers in Psychology: Emotion Science*, 4:184. doi: 10.3389/fpsyg.2013.00184