Emotion Without Words: A Comparison Study of Music and Speech Prosody

L'émotion sans mots : Une étude comparative de la prosodie musicale et de prosodie de la parole

Sarah Faber, BMT(Hons), MA, MTA
Research Assistant, Anglia Ruskin University, United Kingdom

Anna Fiveash, BPsy (Hons), MA
Doctoral Student, Macquarie University, Australia

Abstract

Music and language are two human behaviours that are linked through their innateness, universality, and complexity. Recent research has investigated the communicative similarities between music and language and has found syntactic, semantic, and emotional dimensions in both. Emotional communication is thought to be related to the prosody of language and the dynamics of music. The purpose of this study was to investigate whether language's prosody can successfully communicate a phrase's emotional intent with the lexical elements of speech removed, and whether the results are comparable with a musical phrase of the same perceived emotion. Eighty-five participants ranked a selection of emotional music and prosodic vocalizations on scales of happy and sad. Results showed consistency and correctness in the emotional rankings; however, there was higher variance and lower intensity in the speech examples across all participants and more consistency in music examples among musicians compared to non-musicians. This study suggests that speech prosody can communicate a phrase's emotional content without lexical elements and that the results are comparable, though less intense, than the same emotion conveyed by music. This study has implications for the field of music therapy through support for the accurate identification of emotional information in non-verbal stimuli.

Keywords: music therapy, prosody, emotion, language, music, semantics, dynamics

Résumé

La musique et le langage sont deux comportements humains liés par leur complexité et leur universalité innée. La recherche récente a examiné les similarités se rapportant à la communication entre la musique et le langage et des dimensions syntaxiques, sémantiques et affectives ont été décelées dans les deux cas. Il est usuel de penser que la communication affective est reliée à la prosodie de la parole ainsi qu'aux dynamiques musicales. Le but de cette étude est d'examiner si la prosodie de la parole peut réussir à communiquer une phrase contenant une intention affective sans les éléments lexicaux de la parole et si les résultats sont comparables à une phrase musicale contenant la perception de la même dimension émotive. Quatre-vingt-cinq participants ont classé une sélection de musique émotionnelle et des vocalisations prosodiques sur des échelles allant de « heureux » à « triste ». Les résultats montrent de la cohérence et de l'exactitude dans les catégories d'émotions; cependant il y a de plus grands écarts et une intensité plus faible dans les exemples parlés parmi tous les participants et plus de cohérence dans les exemples musicaux parmi les musiciens comparés aux non-musiciens. Cette étude suggère que la prosodie de la parole peut communiquer le contenu émotionnel d'une phrase sans éléments lexicaux et que les résultats sont comparables, toutefois moins intenses que la même émotion transmise par la musique. Cette étude a des implications pour le domaine de la musicothérapie et ce par l'entremise de l'identification exacte de l'information affective transmise, dans un stimulus non verbal.

Mots clés : musicothérapie, prosodie, émotion, langage, musique, sémantique, dynamiques

Strong connections have been identified between music and language, especially in relation to evolutionary background (Perlovsky, 2012), brain connectivity (Koelsch, Gunter, Wittfoth, & Sammler, 2005), and skill transfer (Besson, Chobert, & Marie, 2011). Some evolutionary theorists have suggested the connection between music and language is in their communicative uses (Cross, 2009; Juslin & Laukka, 2003), and a common communicative use is emotional communication. It has been further suggested that language is a more advanced form of emotional communication derived from music (Mithen, 2009) and that both music and language derive from a pre-linguistic system, which shared elements of music and language for communicative purposes (Masataka, 2009). Such theoretical ideas combined with empirical research on the topic (e.g., Johansson, 2008; Levitin & Menon, 2003; Patel, 2008) show there are many shared similarities between music and language in terms of emotional communication. As emotional expression is considered

paramount in communicating internal feelings and actions to others (Scherer, 1995), and both music and language have been shown to effectively communicate emotions (Steinbeis & Koelsch, 2008), it is useful to compare the similarities and differences in how emotion is communicated through both music and language. These connections will be explored throughout the literature review.

Literature Review

Emotional Communication in Language and Music

Emotional communication in language and music is thought to be related to the prosody of language (Pell, 2006), the dynamics of music (Van der Zwaag, Westerlink, & Van den Broek, 2011), and how they effectively convey meaning and emotion. Prosody in language is defined by intonation, loudness, and tempo (Mitchell, Elliott, Barry, Cruttenden, & Woodruff, 2003) and can be independent of the lexical elements of speech, defined as the word and word-like elements of language (Friederici, Meyer, & von Cramon, 2000). The combination of semantic content and different combinations of prosodic elements lead to emotional communication. The dynamic aspects of music, which help to express emotion, are said to include but are not limited to tempo, mode, harmony, tonality, pitch, rhythm, tension–resolution patterns, and timing (Thompson, 2009). Different combinations of prosodic or dynamic information display differences in the type of emotion communicated both in music and in language. Interestingly, neuroimaging studies have shown a primary emotion pathway activated in response to both musical and spoken emotional content as well as distinct networks activating different areas of the right hemisphere of the brain (Steinbeis & Koelsch, 2008). Music and language also share processing pathways, and the processing of music and language interact and affect each other in the brain (Fiveash & Pammer, 2014). The ability of both speech and music to communicate emotions has been widely researched, and a growing body of evidence is pointing towards strong connections between music and language, particularly in relation to their respective emotional communication abilities.

Communicative Connections Between Music and Language

The extent of the connection between music and speech emotions can be seen when looking at neuropsychological cases where impairment in one domain affects performance in the other. While there are many processing differences between music and speech (Zatorre & Baum, 2012; Zatorre, Belin, & Penhune, 2002), there are also many similarities. For example, Nicholson, Baum, Kilgour, Koh, Munhall, and Cuddy (2003) studied an amusic patient who, following a right hemisphere stroke, was unable to detect differences in music pitch and rhythm or recognize different melodies. After numerous

tests they found that the patient's pitch and rhythm recognition in speech was similarly affected. He was unable to process intonation or discriminate a question from a statement. He was, however, able to perceive speech in other ways and carry on normally in most aspects of speaking life. Such a phenomenon was also observed in two other amusic individuals, who had similar issues with detecting intonation and rhythmic differences across both speech and music stimuli (Patel, Peretz, Tramo, & Labreque, 1998). Such evidence from neuropsychology suggests elements of prosody and musical dynamics are linked in the brain, and therefore theoretical connections between musical dynamics and speech prosody are warranted. Such connections have also been found in other areas of research.

Further links between speech prosody and musical dynamics support the existence of a connection between emotional communication in music and in speech. Patel, Iversen, and Rosenberg (2006) looked at the differences in instrumental music in England and France compared to the respective prosody of English and French languages. They found that the composed music had a number of similarities to the respective languages and that the music reflected differences in prosodic speech. Such research suggests language and music are mutually influential and share similar evolutionary roots. Bowling, Sundararajan, Han, and Purves (2012) suggested there are universal underpinnings of emotional communication in music dynamics and speech prosody. They compared Western and Eastern (South Indian) music and found similarities in tones used to express basic emotions as well as similarities in emotional prosody within the different languages. This led to the conclusion that universal prosodic and musical utterances exist across cultures. Such evidence outlines the evolutionary basis for the connection between music and speech prosody in terms of emotive communication. This can be seen more clearly when looking at transfer effects between music training and emotional identification of speech prosody.

Musical Expertise and Emotional Identification in Speech

Correlations have also been found between musical expertise and a higher ability to correctly identify emotional content in speech prosody (Lima & Castro, 2011; Thompson, Schellenberg, & Husain, 2004). Such transfer effects have been found in musically trained adults of different ages (Lima & Castro, 2011; Thompson et al., 2004) as well as in children who have been taking music lessons for only one year (Thompson et al., 2004), suggesting cross-domain effects of practice between musical expertise and emotion identification in speech. In addition to increased recognition of speech prosody in unintelligible utterances, musicians have also been shown to understand speech prosody in foreign languages better than non-musicians (Thompson et al., 2004). Such transference has been linked to musicians'

greater sensitivity to pitch sounds through training, which in turn influences pitch sensitivity in language (Schön, Magne, & Besson, 2004). Other theories suggest musical training leads to enhanced emotional sensitivity and emotional intelligence (Thompson, 2009), which facilitates a greater ability to detect emotional expression in speech prosody (Trimmer & Cuddy, 2008). Such transfer effects give strong support to the theory that music and speech share similar emotional communication mechanisms, which has implications in the practice of music therapy.

Music Therapy Implications

Therapeutically, the ability to convey and perceive emotions through non-verbal means (whether through non-verbal utterances or instrumental improvisation) is an incredibly important skill in working and communicating with non-verbal, minimally verbal, and communicatively impaired clients (Wigram, 2004). In practice, being aware of the emotive quality of vocal and musical phrases allows clients and music therapists to communicate emotions effectively without engaging in speech (Bruscia, 1998), which may be difficult or impossible for some clients, as seen in those with stroke, aphasia, and autism, for example. Awareness of emotional, behavioural, and musical features can also aid the music therapist in empathic improvisation (see Bruscia, 1987, for a description of Alvin's methods) and in reacting to the utterances of communicatively impaired clients. This influenced the present study.

Purpose of the Study

The significant body of research suggesting connections between musical dynamics and speech prosody in terms of emotional communication led to the main research question of whether the same emotion can be communicated through both music and speech prosody. When the lexical/ verbal elements of language are removed, the listener must rely on linguistic prosody alone. Compared to the dynamics and melody within the music, it was hypothesised that both music and speech prosody would convey similar emotions. It was also hypothesized that musicians would be able to perceive emotions in speech prosody better than non-musicians.

The proposed research questions aimed to investigate whether speech prosody can accurately convey emotion, the extent to which the prosodic speech ratings are comparable to ratings of emotional music, and the differences in ratings between musicians and non-musicians.

Method

Participants

Eighty-five subjects with a mean age of 35.87 years (range = 15 to 72 years) participated in a survey comprising short audio clips of film music and prosodic speech. Prior to beginning the survey, participants were asked their age, native language (85% native English speakers), and other known languages with self-reported level of fluency. They were then asked to rank their musicianship using one of the following: non-musician (11%), music-loving non-musician (41%), amateur musician (26%), semi-professional musician (16%), or professional musician (6%). Participants followed a public link on a social media site to access the survey and could withdraw from the study at any time. Participants remained anonymous (identifying characteristics, such as names and addresses were not collected) and did not receive financial compensation for completion of the survey. Ethical approval was granted from the University of Jyväskylä, Finland, prior to data collection, and participants gave written consent prior to participation.

Design

To test the hypothesis that participants would be able to identify emotion at a similar level in both music and prosodic examples, six film music excerpts (three happy and three sad) and six emotional speech excerpts (three happy and three sad) were selected for the listening examples. The musical excerpts were adapted from Eerola and Vouskoski's (2011) study on emotion in film music, and the emotional speech excerpts were adapted from the Surrey Audio-Visual Expressed Emotion (SAVEE) database, which contains recordings of voice actors reading lines of text in specific emotions as rated by a test panel. Five of the music excerpts were reduced from full score to a single midi instrument line using Finale NotePad software. The sixth excerpt was a single-line piano melody that did not require further reduction. Reductions retained the tempo, dynamics, solo instrumentation, and articulation of the original pieces. The speech excerpts were re-recorded by a voice actor deliberately muffling his voice to retain the prosody of the phrases while obscuring their lexical elements. Excerpts were between 18 and 24 seconds long. The survey was designed using Qualtrics (www.qualtrics.com) and was then distributed via social media networks.

Procedure

Participants were asked to rate each film music excerpt and each speech excerpt on a Likert scale of emotional intensity ranging from *1* for *not at all* to *5* for *very* in response to the questions "How happy did the audio sound?" and "How sad did the audio sound?" Each excerpt was rated for both happiness

and sadness. The music excerpts, besides the happy and sad ratings, had an additional scale of *1* to *5* for familiarity. The excerpts were randomized and participants were able to repeat them as needed. There was no time limit for completion of the survey.

Results

To evaluate the happiness and sadness ratings of each condition (happy music, happy vocals, sad music, sad vocals), the Likert scale ratings were averaged and a final number for each condition was calculated. This resulted in eight averaged responses.

A repeated measures one-way analysis of variance (ANOVA) was run to see if there was a difference in responses between the eight groups. For coding purposes, *HM* = happy music, *HV* = happy vocals, *SM* = sad music, and *SV* = sad vocals. The results for the questions "How happy/sad did the audio sound?" can be seen in Table 1, and are visualised in Figure 1. The ANOVA was run with a Greenhouse Geisser correction as sphericity was not assumed. The result of the ANOVA was significant, $F(7,84) = 156.9$, $p < .001$. A bivariate correlation was run to see whether familiarity scores were correlated with music ratings. A significant positive correlation was found between happiness ratings on happy music and familiarity scores, $r = 0.31$, $p < .01$. All other correlations between familiarity and music ratings were non-significant.

TABLE 1

Mean and Standard Deviation Scores Across Conditions

Condition	Mean	Standard Deviation
HM (Happy)	3.55	0.69
SM (Happy)	1.70	0.65
HM (Sad)	2.09	0.59
SM (Sad)	3.94	0.81
HV (Happy)	3.05	1.06
SV (Happy)	2.04	0.83
HV (Sad)	2.13	0.85
SV (Sad)	2.64	0.99

Note: *HM* = happy music; *SM* = sad music; *HV* = happy vocals; *SV* = sad vocals. The conditions *Happy* and *Sad* refer to the ratings of the question, "How happy/sad did the audio sound?"
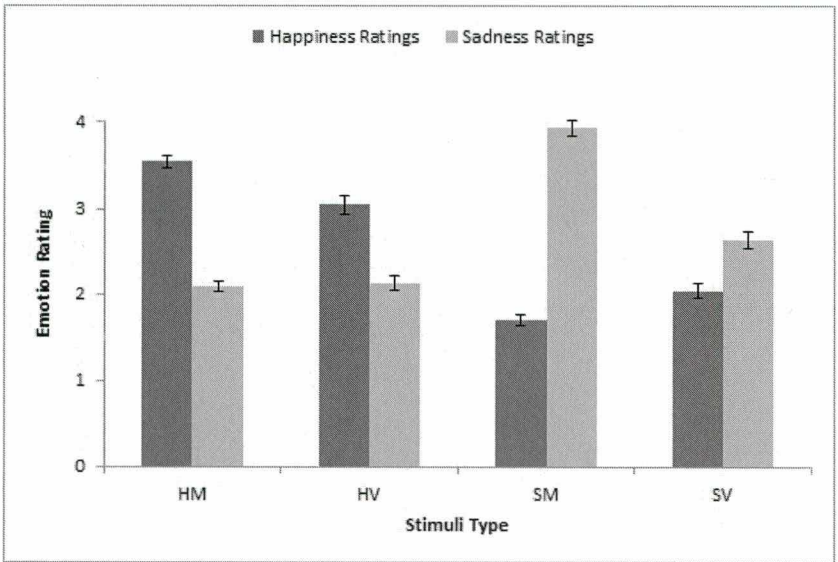
Figure 1. Graph of emotion rating (where *1*= not at all; *2* = not very; *3* = somewhat; *4* = quite; *5*= very) for stimuli type (*HM* = happy music; *HV* = happy vocals; *SM* = sad music; *SV* = sad vocals) to the question, "How happy/sad did the audio sound?" Error bars indicate one standard error either side of the mean.

To determine which differences between groups were significant, post hoc tests using the Bonferroni correction were conducted. For coding purposes, *(happy)* or *(sad)* refers to the happiness or sadness ratings of the questions, "How happy/sad did the audio sound?" Pairwise comparisons were made between the means of HM (happy) and HV (happy), HM (sad) and HV (sad), SM (happy) and SV (happy), SM (sad) and SV (sad), SV (happy) and HV (happy), SV (sad) and HV (sad), SM (happy) and HM (happy), SM (sad) and HM (sad). All comparisons showed significant differences ($p = 0.00$) except for the comparison of HM (sad) and HV (sad), which was $p = 1.00$.

Musical Ability

To assess whether the results differed depending on musical ability, participants were grouped into those who rated themselves as a non-musician (non-musician or music-loving non-musician; $n = 44$) and those who rated themselves as a musician (amateur musician, semi-professional musician, or professional musician; $n = 41$). T-tests were run for the eight groups compared above with a Bonferroni corrected alpha level of $\alpha_i = \alpha/n$, making the required significance level $0.05/8, = 0.006$. Most comparisons remained significant; however, for musicians the comparison between HM (happy) and

HV (happy) was non-significant: $t(40) = 2.527$, $p = .02$. For non-musicians, the comparison between SM (happy) and SV (happy) was non-significant, $t(43) = 1.547$, $p = .13$. Figures 2 and 3 graph the difference in responses between the four levels of rated musicianship and illustrate that sad music was rated as more sad than sad vocals, whereas the difference between happy music and happy vocals was not as profound but still apparent.
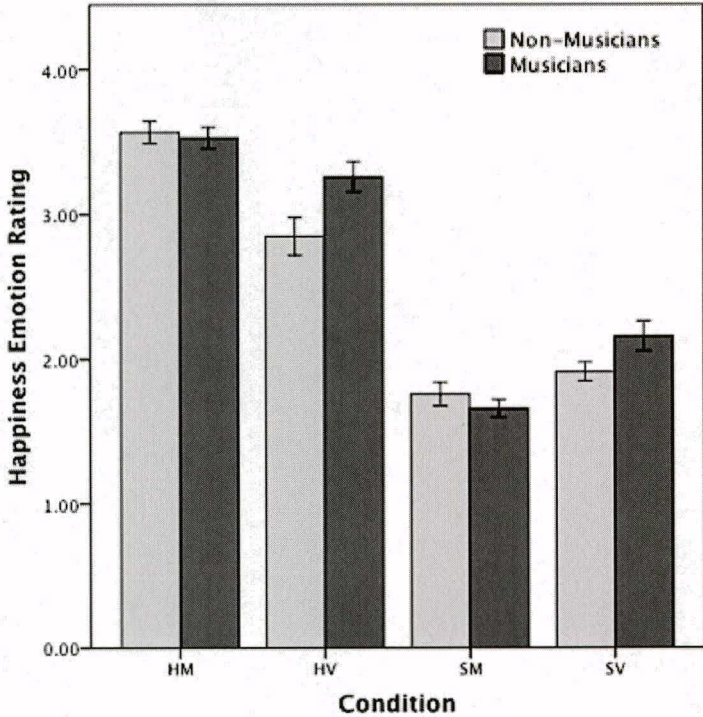

Figure 2. Answers to the question, "How happy did the audio sound?" on a scale of 1 to 5, where 1= not at all and 5 = very, depending on whether the stimuli was happy music (HM), happy vocals (HV), sad music (SM), or sad vocals (SV). Error bars indicate one standard error either side of the mean.
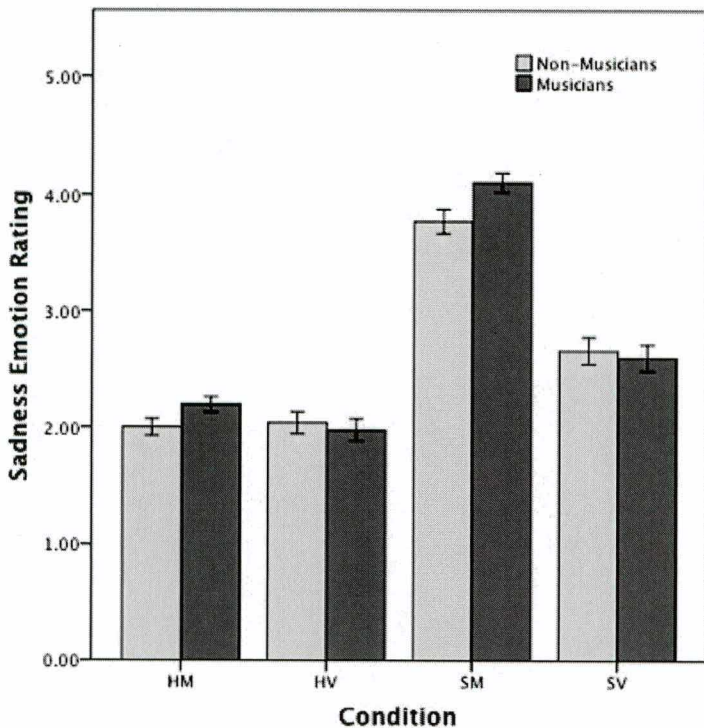
Figure 3. Answers to the question, "How sad did the audio sound?" on a scale of 1 to 5, where 1= not at all and 5 = very, depending on whether the stimuli was happy music (*HM*), happy vocals (*HV*), sad music (*SM*), or sad vocals (*SV*). Error bars indicate one standard error either side of the mean.

Discussion

This study investigated the perceived emotional content of melodic and speech prosodic phrases. Participants were presented with six single-line musical excerpts and six prosodic phrases, three happy and three sad, and were instructed to rate the happiness and sadness of each stimulus.

The initial results showed significant differences in the ratings across music and vocal stimuli. While it was hypothesised that the music and prosodic samples would convey similar levels of emotion and the samples with the same intended emotion would not have significantly different ratings, this only occurred for the happiness ratings of the sad music and sad vocal stimuli. All other comparisons were significantly different in the overall analysis. However, while results were not completely as hypothesised,

participants were able to discriminate intended emotions in both the music and vocal prosody examples (though at different levels), indicating that both vocal prosody without the lexical/verbal elements and musical excerpts can effectively communicate emotion to varying degrees.

Significant differences between ratings of music and prosody with the same intended emotion may be due to the nature of the tasks: participants were not screened prior to the experiment and were able to control the stimuli in terms of repetition, volume, and the equipment used to play the samples. When separated into musician and non-musician categories, the results were non-significant for happiness ratings across happy stimuli in musicians, indicating that happy music and happy prosody were communicating happiness at a similar level for musicians. In relation to the hypothesis, this could suggest that musicians were better able to discriminate happiness in the prosodic excerpts, perhaps due to greater pitch training. Sadness ratings across happy stimuli in non-musicians were also non-significant. While non-musicians did not rate the intended emotion at similar levels, this result shows that they were aware that the sad music and prosody examples were not communicating happiness. This is encouraging in indicating that, across all excerpts, happiness in both music and prosody was rated with little variance, offering some support for the idea that music and prosody have universally communicative potential (Bowling et al., 2012).

A notable difference was observed in the variance between the ratings for music and prosody. Results indicated greater variance in the ratings for vocal prosody compared to music, indicating a more consistent response to music among participants. A possible reason for this could be the context in which participants typically process musical and linguistic information. Humans are generally exposed to music every day, whether intentionally, as with consciously listening to a personal listening device or going to a concert, or unintentionally, as with having a radio or television playing in the background (North, Hargreaves, & Hargreaves, 2004). Music in everyday life is strongly linked to emotional expression and perception (Juslin & Laukka, 2004), and this frequent exposure to music may prime us for perceiving emotion during future exposure to different musical stimuli (Thompson, 2009).

Language, conversely, is most often coupled with lexical elements that express, clearly, the speaker's intentions (Cross et al., 2013). Past studies have found that incongruously paired speech–voice stimuli results in delayed identification of the meaning of the stimulus (Ishii, Reyes, & Kitayama, 2003; Kitayama & Ishii, 2002) and that prosody can affect visual tracking patterns (Rigoulot & Pell, 2012); however, little research exists on emotional accuracy

ratings of speech prosody alone. It can be argued that, when removed from a secondary source of emotional cueing (such as lexical speech, facial expression, or gesture), prosody may be more difficult to classify than music, as indicated by the variance observed. It is also possible that the prosodic excerpts in this experiment, which were specific to the English language, could have been difficult to rank for the participants who were not native English speakers (15% in this study), or were bi- or multilingual, as it has been shown to be easier to understand the prosody inherent in a person's native language (Pell & Skorup, 2008). Further research on emotional perception in speech prosody could be conducted to investigate the role of visual information in emotional identification of prosody as well as cross-cultural studies incorporating languages with differing prosodic patterns.

Another possibility to account for the variance between music and prosody, as well as the intensity of the music ratings, may be found in the structural elements of each stimulus. As per Mitchell et al. (2003), the elements of linguistic prosody are intonation, loudness, and tempo, whereas music also includes mode, pitch, tension, harmony, and other elements (Thompson, 2009). Given the scope of music's dynamic elements and the relative frequency with which music is used to communicate emotion without the benefit of visual or lexical partnering, it would seem music is more emotionally robust than language when the latter is reduced purely to prosody. Greater consistency in results may have been observed with a greater number of excerpts as well.

An interesting pattern is in the consistency in ratings of sad stimuli. No significant discrepancies existed between participants for the sad prosodic and musical stimuli, possibly indicating more universal communicability in sadness. Furthermore, ratings for sad music were rated as more sad than happy music was rated happy despite a positive correlation with familiarity for happiness ratings in the happy music condition. This contradicts past findings on emotional ratings of music with and without lyrics (Ali & Peynircioglu, 2006) and may suggest an emotional memory-based influence on the happy ratings of the happy music stimuli. This may also be due to the original context and cultural specificity of the music. Film music is composed with the intent to augment the atmosphere of a specific scenario and may be more successful at expressing sadness when reduced to a single-line melody. Further research using musical feature analysis software could be employed to investigate this occurrence, as well as controlling the familiarity of the musical stimuli.

Implications for Music Therapy

The use and identification of emotional aspects of music and speech are important in the delivery of music therapy, particularly with non-verbal or communicatively impaired client populations. The ability to identify the intended emotion in a non-verbal utterance or melody is a vital component in a music therapist's understanding of a client's emotional state and music in therapy. Similarly, being able to convey emotion allows the therapist to react to, support, and communicate with that client. This study provides support for the consistent accuracy in which individuals can correctly identify happy and sad emotions in music and prosodic vocalisations (Bowling et al., 2012; Steinbeis & Koelsch, 2008). This knowledge can be used by music therapists to analyze emotional characteristics in a client's instrumental music and vocalization in session and in post-session analysis of musical and emotional data. Future research incorporating additional basic emotions such as fear and anger should be completed to expand what is known about the perception of emotions in music and speech prosody.

Conclusion

Music and spoken language are both advanced cognitive processes that convey emotion, whether through the intentions of the speaker, composer, or musician or the perception of the listener. The aim of this experiment was to investigate whether music and speech prosody could convey comparatively similar emotions and whether the perception of emotions would be greater in musicians compared to non-musicians. It was found that music and speech prosody did appear to communicate the same emotions with some degree of accuracy; the musicians showed greater statistical reliability in their ratings of happiness across happy stimuli and the non-musicians greater statistical reliability in their ratings of happiness across sad stimuli. The results indicated stronger and less varied ratings of music than speech prosody, possibly due to music's dynamic elements, with the highest degree of similarity in ratings of sad stimuli. More research could be done incorporating additional emotions and a greater number of excerpts to further enhance knowledge surrounding music and prosody as related communicative processes.

Acknowledgements

References

Ali, S. O., & Peynircioglu, Z. F. (2006). Songs and emotions: Are lyrics and melodies equal partners? *Psychology of Music, 34*, 511–534.

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of training between music and speech: Common processing, attention, and memory. *Frontiers in Psychology, 2*, 1–12. doi:10.3389/fpsyg.2011.00094

Bowling, D. L, Sundararajan, J., Han, S., & Purves, D. (2012). Expression of emotion in Eastern and Western music mirrors vocalization. *PLoS ONE, 7*(3), e31942. doi:10.1371/journal.pone.0031942

Bruscia, K. E. (1998). *Defining music therapy* (2nd ed.). Gilsum, NH: Barcelona.

Bruscia, K. E. (1987). *Improvisational models of music therapy*. Springfield, IL: Charles C. Thomas.

Cross, I. (2009). Music as a communicative medium. In R. Botha & C. Knight (Eds.), *The Prehistory of Language* (pp. 113–133). Oxford, England: Oxford University Press.

Cross, I., Fitch, W. T., Aboitiz, F., Iriki, A., Jarvis, E. D., Lewis, J., Trehub, S. E. (2013). Culture and evolution. In M. A. Arbib (Ed.), *Language, music and the brain* (pp. 541–562). Cambridge, MA: MIT Press.

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.

Fiveash, A., & Pammer, K. (2014). Music and language: Do they draw on similar syntactic working memory resources? *Psychology of Music, 42*, 190–209. doi:10.1177/0305735612463949

Friederici, A. D., Meyer, M., & von Cramon, D. Y. (2000). Auditory language comprehension: An event-related fMRI study on the processing of syntactic and lexical information. *Brain and Language, 74*, 289–300.

Ishii, K., Reyes, J. A., & Kitayama, S. (2003). Spontaneous attention to word content versus emotional tone: Differences among three cultures. *Psychological Science, 14*, 39–46.

Johansson, B. B. (2008). Language and music: What do they have in common and how do they differ? A neuroscientific approach. *European Review, 16*, 413–427. doi:10.1017/S1062798708000379

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

Juslin, P. N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research, 33*, 217–238.

Kitayama, S., & Ishii, K. (2002). Word and voice: Spontaneous attention to emotional utterances in two languages. *Cognition and Emotion, 16*, 29–60.

Koelsch, S., Gunter, T. C., Wittfoth, M., & Sammler, D. (2005). Interaction between syntax processing in language and in music: An ERP study. *Journal of Cognitive Neuroscience, 17*, 1565–1577. doi:10.1162/089892905774597290

Levitin, D. J., & Menon, V. (2003). Musical structure is processed in "language" areas of the brain: A possible role for Brodmann Area 47 in temporal coherence. *NeuroImage, 20*, 2142–2152. doi:10.1016/j.neuroimage.2003.08.016

Lima, C. F., & Castro, S. L. (2011). Speaking to the trained ear: Musical expertise enhances the recognition of emotions in speech prosody. *Emotion, 11*, 1021–1031. doi:10.1037/a0024521

Masataka, N. (2009). The origins of language and the evolution of music: A comparative approach. *Physics of Life Reviews, 6*, 11–22. doi:10.1016/j.plrev.2008.08.003

Mitchell, R., Elliott, R., Barry, M., Cruttenden, A., & Woodruff, P. (2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia, 41*, 1410–1421. doi:10.1016/S0028-3932(03)00017-4

Mithen, S. (2009). The music instinct: The evolutionary basis of musicality. *Neurosciences and Music III—Disorders and Plasticity, Annual New York Academy of Sciences, 1169*, 3–12. doi:10.1111/j.1749-6632.2009.04590.x

Nicholson, K. G., Baum, S., Kilgour, A., Koh, C. K., Munhall, K. G., & Cuddy, L. L. (2003). Impaired processing of prosodic and musical patterns after right hemisphere damage. *Brain and Cognition, 52*, 382–389. doi:10.1016/S0278-2626(03)00182-9

North, A., Hargreaves, D. J., Hargreaves, J. J. (2004). Uses of music in everyday life. *Music Perception: An Interdisciplinary Journal, 22*, 41–77.

Patel, A. D. (2008). *Music, language and the brain.* New York, NY: Oxford University Press.

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.

Patel, A. D., Peretz, I., Tramo, M., & Labreque, R. (1998). Processing prosodic and musical patterns: A neuropsychological investigation. *Brain and Language, 61*, 123–144.

Pell, M. D. (2006). Judging emotion and attitudes from prosody following brain damage. *Progress in Brain Research, 156*, 303–317.

Pell, M. D., & Skorup, V. (2008). Implicit processing of emotional prosody in a foreign versus native language. *Speech Communication, 50*, 519–530.

Perlovsky, L. (2012). Cognitive function, origin, and evolution of musical emotions. *Musicae Scientiae, 16*(2), 185–199.

Rigoulot, S., & Pell, M. D. (2012). Seeing emotion with your ears: Emotional prosody implicitly guides visual attention to faces. *PLoS ONE, 7*(1), e30740. doi:10.1371/journal.pone.0030740

Scherer, K. R. (1995). Expression of emotion in voice and music. *Journal of Voice, 9*, 235–248.

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates processing in both music and language. *Psychophysiology, 41*, 341–349. doi:10.1111/1469-8986.00172.x

Steinbeis, N., & Koelsch, S. (2008). Comparing the processing of music and language meaning using EEG and fMRI provides evidence for similar and distinct neural representations. *PLoS ONE, 3*, e2226. doi:10.1371/journal.pone.0002226

Thompson, W. F. (2009). *Music, thought and feeling: Understanding the psychology of music.* New York, NY: Oxford University Press.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion, 4*, 46–64. doi:10.1037/1528-3542.4.1.46

Trimmer, C. G., Cuddy, L. L. (2008). Emotional intelligence, not music training, predicts recognition of emotional speech prosody. *Emotion, 8*, 838–849. doi:10.1037/a0014080

Van der Zwaag, M. D., Westerink, J. H. D. M., & van den Broek, E. L. (2011). Emotional and psychophysiological responses to tempo, mode, and percussiveness. *Musicae Scientiae, 15*(2), 250–269.

Wigram, T. (2004). *Improvisation: Methods and techniques for music therapy clinicians, educators and students.* London, England: Jessica Kingsley.

Zatorre, R. J., & Baum, S. R. (2012). Musical melody and speech intonation: Singing a different tune? *PLoS Biol, 10*, e1001372. doi:10.1371/journal.pbio.1001372

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences, 6*, 37–46.