**Uncertain emotion discrimination differences between musicians and non-musical persons is determined by fine structure association**

**Abstract**

How an emotion is encoded and what cues are used to discern emotion in sound have not been elucidated. Here, we sought to identify what attributes of sound confer an emotion to music and to determine if non-musically trained individuals have different musical emotion perception than professional musicians. The objective was to determine which cues are used to resolve emotional signals. Happy or Sad classical musical excerpts encoded in fine structure or envelope were decomposed using a Hilbert transform yielding stimuli conveying different degrees of emotional certainty based on the psychophysical response of identification with the original unaltered excerpts. Certainty was determined by identification of stimuli presented during a descending interval forced choice discrimination psychophysical task. Participants were categorized as good and poor performers on the task (n = 32, age 21.17 ± 2.63 SD) and individuals in the first and last year of musical education in a conservatory (n = 32, age 21.97 ± 2.42). We report fine structure encoded emotional information was essential to detecting emotionality in sound. Emotion resolvability curves were differentiated by group. Non-musically educated individuals used less fine structure information to discriminate emotion in music. The present experiments revealed what cues are used to resolve emotional signals and how they differ between musically educated and non-musically educated individuals. Future studies should determine the threshold of emotional sound for a variety of emotions.

**Keywords**: auditory cortex; emotion; psychophysics; modulation; fine structure component; envelope component

# Introduction

The process of resolving emotions has been described as the optimization of an economic choice (Seymour and Dolan, 2008). Experimentally, emotions are selected due to their perceived certainty and robust differentiation (Koelsch, 2014; LeDoux, 2000; Phan et al., 2002; Barret, 1998). Few studies have attempted to analyze uncertain emotion or the psychoacoustic cues which endow sound with emotion. How an emotion is encoded (Phan et al., 2002; Coutinho and Dibben, 2013) at threshold experience where the ability to distinguish emotion is most important (Barrett, 1998; Pfeifer, 1988) and which sound cues are used to discern emotions (Liberman, and Michaels, 1962; Zatorre and Belin, 2011; Moon, et al., 2014) have not been evaluated thoroughly (Juslin and Västfjäll, 2008). The ability to distinguish emotions (Koelsch, 2014; LeDoux, 2000; Phan et al., 2002; Barret, 1998; Juslin and Västfjäll, 2008; Coutinho and Dibben, 2013) will empower research into determining what constitutes an ‘emotion’ from non-emotional sound (Pfeifer, 1988; LeDoux, 2000; Koelsch, 2014; Wildgruber et al., 2005). Here, we sought to identify the sound attributes involved in musical emotion detection and to determine if non-musically trained individuals have different emotion perception than professionally trained musicians. We were interested in the uncertainty of musical emotion perception when the modulation of sound attribute made it difficult to distinguish from non-emotion.

What are the cues in musical emotion? Music is transmitted through temporal fine structure (FIS) and envelope (ENV) modulations, but the exact contributions to resolving emotional information are unknown. In a seminal study, Liberman and Michaels (1962) removed phonentic information from speech and found when pitch was present the identification of emotional content was 44%, while amplitude cues added only 3% more to the identification of emotion in speech. Although the influence of pitch on emotional content of speech has been known for some time (Fairbanks and Pronovost, 1938), its contribution to music has been less clear (Scherer, 1995; Coutinho and Dibben, 2013). Most studies have concentrated on emotion in speech (Fairbanks, 1940; Lieberman, et al., 1964) or what cues (FIS or ENV) confer intelligibility to speech (Licklider, 1946; Miller and Licklider, 1950; Liberman, and Michaels, 1962; Williams and Stevens, 1972; Frick, 1985; Murray and Arnott, 1993; Drullman, et al., 1994; Drullman, 1995; Scherer, 1995; Drullman et al., 1996; Leinonen et a., 1997). For example, Shannon et al., (1995) found speech recognition primarily utilized temporal cues with a few spectral channels. Smith et al., (2002) using a Hilbert transform found the ENV is most important for speech reception, and FIS most important for pitch perception. Several follow-up studies have corroborated the importance of ENV for speech intelligibility up to a certain number of bands including aspects of FIS (Zeng et al., 2004; Zeng et al., 2005; Davidson, et al., 2009; Swaminathan and Heinz, 2012; Fogerty, 2011; Apoux et al., 2013; Shamma and Lorenzi, 2013; Moon et al., 2014; Swaminathan et al., 2014). Although the emotion carried in speech is a similar, but different auditory perception than music, resolving emotion in music will aide our understanding of identifying emotions in sound.

Do certain individuals possess better discrimination of auditory cues that convey emotion? For example, musicians have an enhanced auditory perception for several acoustic features, such as the ability to learn lexical tones (Wong et al., 2007), enhanced audiovisual processing (Musacchia, et al., 2007), better speech-in-noise perception (Başkent and Gaudrain, 2016; Coffey, et al., 2017), better pitch discrimination threshold (Micheyl et al., 2006), and superior frequency discrimination (Mandikal Vasuki, et al., 2016; Liang et al., 2016; Madsen et al., 2017) compared with non-musicians. Musical training and musical experience shape linguistic patterns (Wong et al., 2007) and enhance speech-in-noise discrimination (Parbery-Clark et al, 2009b), altering brainstem and cortical responses to music and non-musical acoustic features (Parbery-Clark et al., 2016a; Kraus, Chandrasekaran, 2010; Strait et al., 2010). Musicians possess different auditory perceptive abilities than non-musicians; hence, a musician’s ability to discriminate emotion in sound when linked to FIS or ENV changes should also differ from their non-musician counterparts. The sound properties that confer emotion to music have been less studied than the emotion in speech, i.e. prosody (Kotz et al., 2006; Witteman et al., 2012; Coutinho, Dibben, 2013), and how musicians compare with non-musicians may give us cues as to the relative importance of how emotion is discriminated from non-emotion.

Our aim for the present experiment was to investigate certain and uncertain emotion in musical sound in a psychoacoustic experiment (Seymour and Dolan, 2008; Coutinho and Dibben, 2013). The motivation was to determine which cues are used to resolve emotional signals (Pfeifer, 1988). Moreover, we sought to determine whether differences exist between individuals whom have musical experience compared with individuals without musical experience. To accomplish these aims, happy and sad musical stimuli were combined in FIS or ENV band number (nb) domains by a band-wise decomposition process (herein, decomposition; Smith et al., 2002; Moon et al., 2014). A *Happy* or *Sad* excerpt would be encoded in FIS or ENV with increasing ‘nb’ indicating greater stimuli decomposition, altering information essential to emotional identification. The decomposition process yielded stimuli conveying different degrees of emotional certainty based on the psychophysical response of identification with the original unaltered excerpts. Certainty was defined as the ability to identify the decomposed stimuli based on its unaltered form. Approximate categorizations were certain, uncertain and chance identification of emotion. These definitions were approximate because identification varied by band decomposition and by stimuli chimerization (mixing by band). Stimuli varied in emotional certainty were presented in a happy/sad descending interval forced choice discrimination psychophysical task. **First**, we expected varying FIS or ENV information essential to detecting emotionality in sound would result in reduced identification by decomposition; revealing which cue (i.e. FIS or ENV) was important for emotion detection. **Second**, we expected segregating participants by their identification with the original excerpt into best and bad performers, based on the classification reported (Dalla Bella et al., 2001; Peretz et al., 1998; Peretz et al., 2001; Khalfa et al., 2008; Gosselin et al., 2007; Gosselin et al., 2011), would result in different emotional resolvability curves for decompositions of descending certainty. **Third**, we expected assessing musicians entering the first compared to those in their last year of study in conservatory would reveal differences in emotional resolvability based on their musical training (Schlaug et al., 1995; Rauschecker et al., 1999; Münte et al., 2002; Wan and Schlaug, 2002; Zatorre, 2013). **Lastly**, comparing non-musically educated individuals to musically educated individuals would reveal differences in emotional resolvability, based on musical experience. Our aim was to understand the cues used to resolve emotional signals at threshold (i.e. the limit of detecting a sound as emotional).

**Methods**

We included both non-musically and musically educated individuals. Subject with musical studies were in a conservatory in the first and last year of musical education.

**Study participants**

Participants were recruited from a local university and final participants included were randomly selected from approximately n=300 assessed (≈20%; with the sample not differing from population group). All individuals were native Spanish speakers. The study consisted of 64 individuals divided equally into 8 separate groups (Table 1. Participant data). For the non-musical group, we selected participants based on identification of the original emotion in the Montreal Emotional Identification Task (Dalla Bella et al., 2001; Peretz et al., 1998; Peretz et al., 2001; Khalfa et al., 2008; Gosselin et al., 2007; Gosselin et al., 2011). We separated non-musical participants into poor and best performers (n = 32, age 21.17 ± 2.63 SD) based on the original emotion identification and separated musically educated participants (n = 32, age 21.97 ± 2.42) based on entering their first year or last year at the conservatory. Musically educated participants were recruited from the wind and string sections of the conservatory. All volunteers gave informed consent and were free of contraindications for psychoacoustic testing. Participants confirmed normal hearing in addition to undergoing brief audiometric testing. Audiometric testing consisted of presenting and confirming the hearing of a series of pure tones from 400 Hz to 8,000Hz, in addition to linear sweeps, log sweeps, and white noise in the same frequency range. The research protocol was approved by the Internal Review Board Ethics Committee in accordance with the Declaration of Helsinki, 1964. Informed consent prior to undertaking the experiment was granted and abided by as set forth in the Ethical Principles of the Acoustical Society of America for Research Involving Human and Non-Human Animals in Research and Publishing and Presentations.

**Acoustic stimuli**

The 32 original acoustic stimuli classified as sad or happy were taken from a previous study [[www.brams.umontreal.ca/peretz](http://www.brams.umontreal.ca/peretz); Table S1; Peretz et al., 1998]. As detailed previously, stimuli were transcribed for piano computer-generated versions created on a microcomputer running a Musical Instrument Digital Interface (MIDI) sequencing program (Sequencer Plus Gold) and controlling a sample playback digital synthesizer (Rolland Sound Canevas SC 50) by entering manually the original scores (Dalla Bella et al., 2001; Peretz et al., 1998; Peretz et al., 2001; Khalfa et al., 2008; Gosselin et al., 2007; Gosselin et al., 2011). Stimuli were drawn from the corpus of classical western music consisting of baroque (e.g. Bach, Albinoni), classical (e.g. Mozart), romantic (e.g. Verdi), and contemporary (e.g. Ravel) periods. Half of the 32 stimuli evoked a sense of happiness consisting of major mode with a median tempo of 138 beats per min (bpm) and the other half evoked a sense of sadness consisting of minor mode with a median tempo of 53 bpm (Peretz et al., 2001). In the excerpt, each original tone or series of tones occupied their values in terms of pitch and duration, keeping intensity and velocity constant. The MIDI generated files of classical music were processed further in MATLAB to curtail length to 3 second duration wav files, restricted in frequency/amplitude range for presentation, and analyzed spectrally for differences in FIS or ENV. Original and altered stimuli were presented with MATLAB (Statistics Toolbox Release 2012b, The MathWorks, Inc., Natick, Massachusetts) using the Psychophysics Toolbox extension [http://psychtoolbox.org/]. Sound level adjustment was performed prior to beginning psychoacoustic testing as described above.

**Acoustic stimuli decomposition to construct uncertain emotion stimuli**

Happy and sad stimuli were decomposed by a Hilbert transform (Smith et al., 2002; Moon et al., 2014). The decomposition process accomplished two tasks: 1) to associate the acoustic aspects of emotion (happy or sad) with FIS or ENV and 2) when recombined, develop stimuli which consist of different mixtures of emotion by bandwidth decomposition, thereby creating emotional uncertainty in the hybrid acoustic stimuli due to change in identification with the unaltered excerpt. We were interested in the process of creating hybrid mixtures of happy and sad emotion to alter perception by equal step bandwidths. Six band decompositions (2nb, 4nb, 8nb, 16nb, 32nb, 64nb) were created with cut-off frequencies 80, 260, 600, 1,240, 2,420, 4,650 and 8,820 Hz. The Hilbert decomposition resulted in a hybrid acoustic sound, happy emotion in FIS and sad emotion in ENV or sad emotion in FIS and happy emotion in ENV (Smith et al., 2002). Emotion was effectively tied to FIS or ENV in equally spaced decreasing bandwidth (Moon e al., 2014; Smith et al., 2002). This process created uncertain acoustic emotional stimuli based on identification with the original unaltered emotional stimuli. For signal decomposition carried out within the present study, the Hilbert transform *y*(t), in the time domain is related to real function *x*(t), by the analytic signal A(t)=*x*(t)+i*y*(t), with i = (-1)1/2. The Hilbert ENV is the magnitude of the analytic signal, *|A(t)| = ( (sr(t))2 + (si(t))2 )1/2* and the Hilbert FIS is cos φ(t), where *φ(t)= arctan( sr(t)/si(t) )* is the phase of the analytic signal (Smith et al., 2002; King, 2009). If the real (r) part pertains to cosine of the frequency contained within the signal and the imaginary (i) part pertains to sine of the frequency contained within the signal, the magnitude of the amplitude is related by the value of the cosine and sine of the signal (King, 2009; Oswald, 1956). The decomposition process has been elaborated on previously for a variety of signal processing purposes (Smith et al., 2002; King, 2009; Oswald, 1956; Moon et al., 2014). Here, the result of decomposing FIS and ENV with happy or sad acoustic stimuli conveying emotion allows reconstruction of hybrid acoustic stimuli with happy emotion contained within FIS and sad emotion contained within ENV in a recombined acoustic stimuli or vice versa, the reconstruction of hybrid acoustic stimuli with sad emotion contained within FIS and happy emotion contained within ENV in a recombined acoustic stimuli. Recombined hybrid stimuli with differing combinations of ENV and FIS from either emotion category where presented in a happy/sad descending two-interval forced choice discrimination task (Fig. 1a and 1b; Binder et al., 2004). For visualization of the Hilbert process we present a simplified example for a single 440Hz tone (Fig. 1c) and a complex series of tones consisting of the 440Hz 4th order harmonic with 440Hz, 880Hz, 1320Hz, and 1760Hz components (Fig. 1e), undergoing FIS and ENV extraction by the Hilbert transformation and recombination into hybrid stimuli by different band decompositions. Figure 1g presents the Hilbert transformation of the simple 440Hz tone in FIS combined with the complex series of tones, 4th order harmonic in ENV. Figure 1h presents the Hilbert transformation of the complex series of tones, 440Hz 4th order harmonic in FIS combined with the simple 440Hz tone in ENV. From left to right for Fig. 1g/1h, band decomposition increases from 2nb to 64nb. The simple tone and the tones in the series were given an amplitude double its predecessor (t) starting with 10 SPL dB, a phase (π/2)/t change from its lower harmonic and separate durations. For the representation, spectrogram plots contained magnitude (dB) on z-axis, normalized frequency (x π rad/sample) on x-axis, and time in milliseconds (ms) on y axis. Spectrograms and acoustic stimuli were normalized across power/frequency (dB/Hz), here amplitude (SPL dB) by fine structure components (Hz).

**Procedure: Acoustic stimuli happy/sad descending two-interval forced choice discrimination task**

Participants undertook a descending two-interval forced choice discrimination task where they were required to respond to the acoustic stimuli indicating the sound as happy or sad for original and band decomposed hybrid stimuli (Binder et al., 2004). All participants responded to all acoustic stimuli. Classification of stimuli was made based on original categorization (Smith et al., 2002). Originally, a Linkert-scale task as described was performed (Peretz et al., 1998), but participants always categorically decided stimuli along previous classifications, i.e. never indicating neutral; therefore, was not included in the present analysis. Participants were handed a sheet consisting of numbered rows, 1 to 32 organized randomly and consisting of 7 columns. Column was organized as original stimuli presentation, followed from left to right by stimuli decompositions (2nb, 4nb, etc). Trial was organized by random presentation followed by descending decompositions. Decompositions in FIS were presented starting with their original unaltered form and continuing through 2nb to 64nb decompositions (Fig. 1a).

**Statistical Analysis**

1. Average discrimination curves by FIS (best/bad performers, male/female, high/low music, music non-music)
2. ANOVA with participant (best/bad performers, male/female, high/low music, music non-music) and classification in FIS by band decomposition
3. Follow-up t-test analysis differences between participants and band decomposition
4. Discriminability Indices d-prime and A-prime
5. Normalized Benefit
6. Canonical discriminant analysis

*Average response calculations* were derived for FIS and ENV for happy and sad musical stimuli. The average discrimination curves were percent identification of the response for greatest response (found in FIS for all original stimuli). Curves were analyzed separately by non-musically educated poor and best performers separated by male and female, in addition to first and last year musically educated participants separated by male and female. Discrimination curves were tested for significance with an ANOVA and follow-up *t*-tests. The discrimination normalized ratio for identification of the stimuli as happy or sad was calculated by determining the percent identification of happy or sad over its opposite stimuli discrimination. Measures of discriminability D′ and the corrected nonparametric measure of discriminability A′ were utilized for determining differences in emotional resolvability (Stanislaw and Todorov, 1999; Verde, Macmillan and Rotello, 2006). These measures provide an estimation of signal from noise and determination of the threshold response for the acoustic emotional stimuli (Green, 1960; Swets, 1961a; Swets, 1961b; Swets, Tanner, and Birdsall, 1961; Swets, 1986). The interest in determining threshold response for an acoustic emotion stimuli is to ascertain when an emotional stimuli becomes non-emotional. When the participant response changed from identifying emotional sound in FIS or ENV was no longer associated with a correct percent identification similar to its original unaltered excerpt, the emotion in sound was deemed un-emotional. Here we group averaged threshold identifications; however, individual threshold values were be calculated by stimuli. The difference between FIS or ENV emotion discrimination was examined as the ratio between the perception of one emotion using FIS or ENV versus the perception of the other emotion using FIS or ENV by calculating the normalized benefit of FIS or ENV to the emotion discrimination curve (originally calculated for visual contribution to speech in noise; Sumby and Pollack, 1954; Meister et al., 2016). The adopted formula utilized was: FIS benefit = (FIS – ENV)/ (1-ENV) or ENV benefit = (ENV– FIS)/ (1-FIS), to compare both the FIS and ENV contributions, respectively on the scale of +1 to -1. The percent difference normalized by each contribution result was a positive value representing benefit to perception and negative value indicating lack of contribution to perception. A canonical discriminant analysis was used to determine the weighted average associated with our variables (nb0, nb2, nb4, etc), which best separate, our different groups (poor performer, best performer, first year musical education, last year musical education). The two generalized canonical discriminant analysis (one for happy and the other for sad) was computed using the multivariate linear model Group~nb0+nb2+nb4+nb8+nb16+nb32+nb64 to obtain the canonical scores and vectors. It represents a transformation of the original variables in the space of maximal differences for the group.

# Results

1. Non-musical educated
   1. Poor performers
      1. Poor Male
      2. Poor female
   2. Best performers
      1. Good Male
      2. Good Female
   3. Group Comparisons Differences
   4. Poor Males with Best Males
   5. Poor females with Best Females
   6. Discriminability A′ and FIS/ENV benefit for poor and best performers
      1. Discriminability A′
         1. Average discriminability A′ for Poor
         2. Average discriminability A′ for best
         3. Difference discriminability A′ between poor and best
      2. Benefit FIS
         1. Average benefit FIS for Poor
         2. Average benefit FIS for best
         3. Difference benefit FIS between poor and best
   7. Canonical discriminant analysis
      1. What discriminates between the groups most
2. Musically educated
   1. First year
      1. Female
      2. Male
   2. Last year
      1. Female
      2. Male
   3. Group Comparisons Differences
   4. First year Males with Last Year Males
   5. First year females with last Females
   6. Discriminability A′ and FIS/ENV benefit for first and last year musically educated
      1. Discriminability A′
         1. Average discriminability A′ for First Year
         2. Average discriminability A′ for Last Year
         3. Difference discriminability A′ between first and last year
      2. Benefit FIS
         1. Average benefit FIS for first year
         2. Average benefit FIS for last year
         3. Difference benefit FIS between first and last year
3. Canonical discriminant analysis
   1. What discriminates between the groups most

# Results – Poor and best performers

Percent identification for poor (df =6, 18, F = 30.76, *p* < 0.0001) and best (df = 6, 18, F = 79.04, p < 0.0001) performers were significantly different by band identification indicating decrease in certainty (Figure 2a and 2b). Both poor and best performers used fine structure for emotion identification and discrimination performance decreased with increasing band (SI Fig. 1a – 1d). The SI Figure 1a and 2b show the average male poor and best identification curves and SI figure 1c and 1d show the average female poor and best identification curves.

While there were no apparent differences for poor performer’s fine structure or envelope based identification for happy or sad emotion, happy FIS (df = 1, 6, F = 2.545, P=0.1617), sad FIS (df = 1, 6, F = 1.494, *p* = 0.2674), happy ENV (df = 1, 6, F = 1.897, *p* = 0.2176), sad ENV (df = 1, 6, F = 2.885, *p* = 0.1403), best performers showed preferences. For best performers, happy emotion in fine structure (df = 1, 6, F = 7.749, *p* = 0.0318) and sad emotion in envelope (df = 1, 6, F = 7.591, *p* = 0.0331) was different between males and females. Sad in FIS was significantly different between poor and best performers (df = 6, 30, F = 3.773, *p* = 0.0091), but happy in FIS did not reach significance (df = 6, 30, F = 1.912, *p* = 0.1218). Happy in ENV was significantly different between poor and best performers (df = 1, 6, F = 3.630, *p* = 0.0110), but sad in ENV did not reach significance (df = 1, 6, F = 1.881, *p* = 0.1273).

Comparing non-musically educated best and poor performers, no significant differences were apparent between discriminating uncertain emotion (df = 15, 90, F = 1.814, *p* = 0.0445). Best and poor performers use fine structure to discriminate happy and sad uncertain emotions, but do so approximately 4.01% ± 3.33% SD and 9.20% ± 6.82% SD, respectively, differently depending on uncertainty (SI Fig 1 A through D and Discrimination curves). Males and females, irrespective of poor or best performer, use fine structure to discriminate happy and sad uncertain emotions, but do so approximately 2.67% ± 1.62% SD and 2.06% ± 1.67% SD, respectively, differently depending on uncertainty (SI Fig 1 A through D, Discrimination curves).

The corrected nonparametric measure of discriminability A′ used for determining differences in emotional resolvability found no significant difference between poor performers sad (p=0.5957) and happy (p=0.6612) compared with the best performers sad (p=0.6712) and happy (p=0.6644). Figure 3a demonstrates grouped discriminability A′ for poor and best performers by uncertain emotion. Figure 3c depicts the group averaged normalized benefit of FIS or ENV to the emotion discrimination curve for poor and best performers. The FIS was beneficial for emotion discriminability for poor and best performers for sad and happy stimuli. Poor performers had 0.2928 and 0.2377 benefit for happy and sad, respectively. Best performers had 0.3295 and 0.3355 benefit for happy and sad, respectively. The ENV was negatively beneficial for emotion discrimination for all stimuli for poor or best performers.

# Results – First and last year musically educated

Percent identification for first year musicians (df = 6, 18, F = 71.69, *p* < 0.0001) and last year musicians (df = 6, 18, F = 45.37, *p* < 0.0001) were significantly different by band identification indicating decrease in certainty (Figure 2a and 2b). Both first year and last year musicians used fine structure for emotion identification and discrimination performance decreased with increasing band (SI Fig. 2a – 2f). First year (df = 6, 18, F = 28.29, *p* < 0.0001) and last year (df = 6, 18, F = 29.04, *p* < 0.0001) musicians resolved happy and sad emotion differently. The SI Figure 2a and 2d show the average group first or last year musician discrimination curves for happy and sad stimuli. The SI Figure 2b and 2c show male and female first year averaged discrimination curves and SI Figure 2e and 2f show male and female last year averaged discrimination curves.

While there were no apparent identification differences for first year educated musicians related to FIS happy (df = 1, 6, F = 3.364, *p* = 0.1163), FIS sad (df = 1, 6, F = 0.1967, *p* = 0.6729), ENV happy (df = 1, 6, F = 0.1047, *p* = 0.7573), or ENV sad emotion resolvability (df = 1, 6, F = 6.058, *p* = 0.0490), last year musicians showed preferences. Last year musicians were significantly different for emotion resolvability for FIS happy (df = 1, 6, F = 14.88, *p* = 0.0084), FIS sad (df = 1, 6, F = 11.91, *p* = 0.0136), ENV happy (df =1, 6, F = 13.66, *p* = 0.0101) and ENV sad (df =1, 6, F = 14.49, *p* = 0.0089). There were significant differences in emotion resolvability between first and last year musicians for FIS happy (df = 6, 18, F = 7.585, *p* = 0.0017), FIS sad (df = 6, 18, F = 4.574, *p* = 0.0150), ENV happy (df = 6, 18, F = 4.966, *p* = 0.0110), and ENV sad (df = 6, 18, F = 7.816, *p* = 0.0015).

Comparing first and last year musically-educated, a significant difference was found discriminating uncertain emotion (df = 15, 90, F = 4.377, *p* < 0.0001). First and last year musicians use fine structure to discriminate happy and sad uncertain emotion, but do so approximately 2.51% ± 1.68% SD and 3.90% ± 2.30% SD, respectively, differently depending on uncertainty (SI Fig 2 a through f; SI Fig 1 a through f – Discrimination curves). Males and females, irrespective of first/last year musical education, use fine structure to discriminate happy and sad uncertain emotion, but do so approximately 12.10% ± 4.08% SD and 4.24% ± 3.16% SD, respectively, differently depending on uncertainty (SI Fig 2 a through f; SI Fig 2 a through f – Discrimination curves).

The corrected nonparametric measure of discriminability A′ used for determining differences in emotional resolvability found a significant difference between for both first year and last year musically educated participants (Fig 3b). First year participants discriminability A′ for sad was 0.579 and last year participants discriminability A′ for sad was 0.5632; however, first and last year participants discriminability A′ for happy was 0.7725 and 0.7805, respectively. This represents a discriminability A′ difference of 24.99% for first year and 27.83% for last year musically educated participants. Musical education, whether first or last year, significantly enabled discriminability A′ for happy excerpts with a difference of only 2.83 percent by end of musical education. Figure 3b demonstrates grouped discriminability A′ for first year and last year musically educated participants by uncertain emotion. Note the drastic difference in missing blue proportion of the curve, for sad, representing lack of discriminability. Figure 3d depicts the group averaged normalized benefit of FIS or ENV to the emotion discrimination curve for musically educated participants. The FIS was beneficial for emotion discriminability for first or last year musical education and for sad and happy stimuli. First year musical education had 0.4813 and 0.3013 benefit for happy and sad, respectively. Last year musical education had 0.4652 and 0.3094 benefit for happy and sad, respectively. The ENV was negatively beneficial for emotion discrimination for all stimuli for musically educated participants.

# Results – Poor and best performers compared with First and last year musically educated

Identification of happy for non-musically educated and musically educated were significantly different from each other (df = 9, 54, F = 15.68, *p* < 0.0001) and by emotional resolvability (df = 6, 54, F = 315.5, *p* < 0.0001; Fig. 4a). Identification of sad for non-musically educated and musically educated was significantly different from each other (df = 9, 54, F = 3.526, *p* = 0.0017) and by emotional resolvability (df = 6, 54, F = 112.1, *p* < 0.0001; Fig. 4b). Figure 4 demonstrates the separation between the different groups. Group separation and differences were based on emotional resolvability and Fig. 4 shows the spread of separation based on the canonical discriminate functions. For sad, the greatest standardized beta coefficients were org, Can1 = 0.6264 and for nb64 was Can1 = -0.7641. For happy, the greatest standardized beta coefficients were org, Can1 = -1.011 and for nb64 was Can1 = 0.4193. For both sad and happy the greatest Standardized beta coefficients were org and nb64, indicating the original and most uncertain stimuli were most discriminable between our groups (Fig. 4). Differences in discriminability A′ between non-musically educated and musically educated participants (Fig. 3a and 3b) for happy were statistically significant (df = 1.518, 9.109, F = 8.796, *p* = 0.0101) as were differences in discriminability A′ as a function of emotional resolvability (df = 6, 42, F = 191.7, *p* < 0.0001). Differences in discriminability A′ between non-musically educated and musically educated participants (Fig. 3a and 3b) for sad were statistically significant (df = 2.934, 17.61, F = 5.086, *p* = 0.0107) as were differences in discriminability A′ as a function of emotional resolvability (df = 6, 42, F = 156.1, *p* < 0.0001). The averaged normalized benefit of happy FIS to the emotion discrimination curve (Fig. 3c and 3d) was statically different between non-musically educated and musically educated (df = 2.383, 14.30, F = 6.922, *p* = 0.0060) as was the difference in benefit of happy FIS to emotional resolvability (df = 6, 42, F = 298.5, *p* < 0.0001). The averaged normalized benefit of sad FIS to the emotion discrimination curve (Fig. 3c and 3d) was not statically different between non-musically educated and musically educated (df = 2.643, 15.86, F = 1.826, p = 0.1870), although the contribution to emotional resolvability was statistically significant (df = 6, 42, F = 333.0, *p* < 0.0001). The averaged normalized benefit of happy ENV to the emotion discrimination curve (Fig. 3c and 3d) was not statically different between non-musically educated and musically educated (df = 2.638, 15.83, F = 1.708, p = 0.2087), although the contribution to emotional resolvability was statistically significant (df = 6, 42, F = 342.1, *p* < 0.0001). The averaged normalized benefit of sad ENV to the emotion discrimination curve (Fig. 3c and 3d) was statically different between non-musically educated and musically educated (df = 2.909, 17.45, F = 10.05, p = 0.0005), as was the difference in benefit of sad ENV to emotional resolvability (df = 6, 42, F = 450.0, *p* < 0.0001).

# Discussion

The objective for the present experiment was to investigate certain and uncertain emotion in musical sounds and determine if non-musically educated and musically educated individuals resolve emotion differently. Here, stimuli varied in emotional certainty were presented in a happy/sad descending interval forced choice discrimination psychophysical task. There are three results of considerable interest: **First**, FIS information was essential to detecting emotionality in sound. **Second**, different emotion resolvability curves were found dependent on being poor or best performs, in addition to year of musical education. **Lastly**, non-musically educated individuals used less FIS and had reduced emotional resolvability curves compared to musicians. The aim for the present experiments was to understand the cues used to resolve emotional signals at threshold and how they differ between musically educated and non-musically educated individuals.

**Resolving emotion using psychoacoustic cues**

Emotion in sound is transmitted through temporal fine structure and envelope modulations. In a groundbreaking study, Liberman and Michaels (1962) found pitch aided in the identification of emotional content by 44% while amplitude cues added only 3% more. The present study found FIS cues essential and most beneficial to detecting and discriminating emotion in musical excerpts, whether individuals were good performers, poor performers, or had musical training. The results indicate fine structure cues are essential toward resolving emotion in sound and individuals differ in their perceptive ability to discriminate these cues. Furthermore, happy emotion was discriminated with higher accuracy than sad emotion for all groups (Fig. 2A, 3; SI-Fig 1 and 2). This is most likely due to individuals using major mode and the fast tempo of tones for discriminating emotion in sound, which are prominent in happy stimuli (Peretz et al., 2001; Dalla-Bella et al., 2001). However, how these cues determine specific emotions and how the cues are perceived by individuals is not completely understood. For example, in a recent study, individuals differed in their tendency to report the co-occurrence of discrete emotions of the same valence; here individuals varied in their extent to which they distinguished between like-valenced discrete emotions or did not distinguish between like-valenced emotions when reporting on their subjective experience (Barret, 1998). The results indicate that individuals are reporting several affective states together, or it may indicate they are not distinguishing between distinct emotional states. The aforementioned manuscript bolstered support both for the theory of discrete emotion where individuals label emotion determining a subjective level of arousal and the dimensional theory of emotion where individuals focus on the subjective emotional experience, dimensionalized by valence, arousal, and intensity of the affective state (Barret, 1998). The present study found emotional resolvability changed as a function of altering the FIS content of the musical excerpt, revealing an essential cue to discriminating emotion is fine structure. Recently, a study investigating the similarities/dissimilarities of emotion in music and speech prosody found the psychoacoustic features implicated were loudness, tempo and speech rate, melodic and prosodic contour, spectral centroid, and sharpness, whereas the distinct features to music and speech were spectral flux and roughness, respectively. Here, the authors indicated emotional cues in sound are encoded as psychoacoustic spatiotemporal patterns, which for music and speech rely heavily on their ‘shared acoustic profiles’ (Coutinho and Dibben, 2013). We implore, research into determining what constitutes an ‘emotion’ from non-emotional sound (Pfeifer, 1988; LeDoux, 2000; Koelsch, 2014; Wildgruber et al., 2005) will enable a more thorough classification of the neurobiology of emotion. Future studies should further explore the psychoacoustic foundations to emotion.

**Musically educated compared with no musical education**

Musically educated individuals discriminate emotion differently, likely due to musical education. For example, in the speech-in-noise and hearing-in-noise test, musicians performed significantly better than the nonmusicians, derived in part from musicians' enhanced working memory and frequency discrimination (Parbery-Clark et al, 2009b). Musicians in the present study used more FIS, and discriminated emotion in sound differently than non-musicians, using more FIS through each of the nb decompositions. Here, a musician’s ability to discriminate emotion in sound is likely enhanced due to musical training. Within the group of musicians, last year musically educated individuals discriminated happy or sad excerpts slightly differently than first year musically educated individuals. Although the greatest difference was in musically educated individuals between males and females, it is the difference between musically educated and non-musically educated which reveals the most as this group was benefited more by FIS components. For example, musicians discriminating happy or sad excerpts utilized more FIS irrespective of whether individuals were first or last year in their musical education. Recent discrimination tasks bolster these results, in a study where participants were tasked to detect frequency changes in quiet and noisy conditions, the acoustic change complex, a type of late auditory evoked potential, was significantly different (larger P2′ amplitude) compared with non-musicians (Liang et al., 2016). Moreover, in a task where target speech and competing speech were presented with either their natural F0 contours or on a monotone F0, and the F0 difference between the target and masker was systematically varied, F0 discrimination was significantly better for musicians (Madsen et al., 2017). Most of these frequency discrimination tasks indicate that musicians are enhanced in their ability to perceive or discriminate FIS or fine structure components. Future studies should expand the range and variety of emotion discrimination paradigms, to explore differences between musicians and non-musicians.

**Study Limitations and future directions**

The present study concentrated on analyzing two emotions elicited by classical musical excerpts. To constrain the variety of emotion to a dichotomous task is difficult, but aides in ascertaining information concerning how the basic components of sound cue emotion. Future studies should concentrate on analyzing the more diverse emotional repertoire which exists in humans. The present study analyzed non-musically educated individuals and individuals with musical education. We chose these groups because we thought based on prior literature (Wong et al., 2007; Musacchia, et al., 2007; Başkent and Gaudrain, 2016; Coffey, et al., 2017; Micheyl et al., 2006; Mandikal Vasuki, et al., 2016; Liang et al., 2016; Madsen et al., 2017; Wong et al., 2007; Parbery-Clark et al, 2009b; Parbery-Clark et al., 2016a; Kraus, Chandrasekaran, 2010; Strait et al., 2010), musical training would influence acoustic perception in the emotional resolvability task. The present manuscript found year in musical education significantly affected emotional resolvability (*F*15,90 = 4.377, *p* < 0.0001), with late musicians using more fine structure to discriminate happy uncertain emotion 2.51% ± 1.68%. Future studies could analyze different musicians (piano versus string), to determine if graduations on emotional resolvability are associated with instrument played.

# References

Apoux, F., Yoho, S. E., Youngdahl, C. L., & Healy, E. W. (2013). Role and relative contribution of temporal envelope and fine structure cues in sentence recognition by normal-hearing listeners. J Acoust Soc Am. 134(3):2205-2212.

Barrett, L. F. (1998). Discrete emotions or dimensions? The role of valence focus and arousal focus. Cogn Emot 12:579-599.

Başkent, D., & Gaudrain, E. (2016). Musician advantage for speech-on-speech perception. J Acoust Soc Am. 139(3):EL51-6.

Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. Nat. Neurosci. 7(3):295-301.

Coffey, E. B. J., Mogilever, N. B., & Zatorre, R. J. (2017). Speech-in-noise perception in musicians: A review. Hear Res. 352:49-69.

Coutinho, E., & Dibben, N. (2013). Psychoacoustic cues to emotion in speech prosody and music. Cogn. Emot. 27:658-84.

Dalla Bella, S., Peretz, I., Rousseau, L., Gosselin, N., Ayotte, J., & Lavoie, A. (2001). Development of the happy-sad distinction in music appreciation. Does tempo emerge earlier than mode? Ann. N. Y. Acad. Sci. 930:436-8.

Davidson, S. A., Gilkey, R. H., Colburn, H. S., & Carney, L. H. (2009). Diotic and dichotic detection with reproducible chimeric stimuli. J Acoust Soc Am. 126(4):1889-905.

Drullman, R. (1995). Temporal envelope and fine structure cues for speech intelligibility. J Acoust Soc Am. 97(1):585-592.

Drullman, R., Festen, J. M. and Houtgast, T. (1996). Effect of temporal modulation reduction on spectral contrasts in speech. J Acoust Soc Am. 99(4):2358-2364.

Drullman, R., Festen, J. M. and Plomp, R. (1994). Effect of temporal envelope smearing on speech reception. J Acoust Soc Am. 95(2):1053-1064.

Fairbanks, G. (1940). Recent experimental investigations of vocal pitch in speech. J Acoust Soc Am. 11:457-466.

Fairbanks, G., & Pronovost, W. (1939). Vocal pitch during simulated emotion. Science. 88(2286): 382-3.

Fogerty D. (2011). Perceptual weighting of individual and concurrent cues for sentence intelligibility: Frequency, envelope, and fine structure. J Acoust Soc Am. 129(4):1889-905.

Frick, R. W. (1985). Communicating emotion: The role of prosodic features. Psychol. Bull. 97(3):412–429.

Gosselin, N., Peretz, I., Hasboun, D., Baulac, M., & Samson, S. (2011). Impaired recognition of musical emotions and facial expressions following anteromedial temporal lobe excision. Cortex. 47:1116-25.

Gosselin, N., Peretz, I., Johnsen, E., & Adolphs, R. (2007). Amygdala damage impairs emotion recognition from music. Neuropsychologia. 45:236-44.

Green, D. A. (1960). Psychoacoustics and detection theory. 32(10):1189-1203.

Juslin PN, & Västfjäll D. (2008). Emotional responses to music: the need to consider underlying mechanisms. Behav Brain Sci 31:559-75.

Khalfa, S., Guye, M., Peretz, I., Chapon, F., Girard, N., Chauvel, P., & Liégeois-Chauvel, C. (2008). Evidence of lateralized anteromedial temporal structures involvement in musical emotion processing. Neuropsychologia. 46:2485-93.

King, F. W. (2009). Hilbert transforms. Volume 1 and 2. Encyclopedia of Mathematics and Its Applications. Cambridge University Press.

Koelsch, S. (2014). Brain correlates of music-evoked emotions. Nat Rev Neurosci 15:170-80.

Kotz, S. A., Meyer, M., & Paulmann, S. (2006) Lateralization of emotional prosody in the brain: an overview and synopsis on the impact of study design. Prog Brain Res 156:285-94.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. Nature Rev Neurosci. 11:599–605.

LeDoux, J. E. (2000). Emotion circuits in the brain. Annu Rev Neurosci. 23:155–184.

Leinonen, L., Hiltunen, T., Linnankoski, I., & Laakso, M. J. (1997). Expression or emotional-motivational connotations with a one-word utterance. J. Acoust. Soc. Am. 102:1853–1863.

Liang, C., Earl, B., Thompson, I., Whitaker, K., Cahn, S., Xiang, J., Fu, Q.J., & Zhang, F. (2016). Musicians are better than non-musicians in frequency change detection: Behavioral and electrophysiological evidence. Front Neurosci. 10:464.

Liberman, P., and Michaels, S. B. (1962). Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. J. Acoust. Soc. Am. 34:922-927.

Liberman, P., Michaels, S. B. & Soron, H. I. (1964). Emotional effects of fundamental-frequency transformations. J Acoust Soc Am. 36:1048.

Licklider, J. C. R. (1946). Effects of amplitude distortion upon the intelligibility of speech. J Acoust Soc Am. 18(2):429-434.

Madsen, S. M. K., Whiteford, K. L., & Oxenham, A. J. (2017). Musicians do not benefit from differences in fundamental frequency when listening to speech in competing speech backgrounds. Sci Rep. 7(1):12624.

Mandikal Vasuki, P. R., Sharma, M., Demuth, K., & Arciuli, J. (2016). Musicians' edge: A comparison of auditory processing, cognitive abilities and statistical learning. Hear Res. 342:112-123.

Meister, H, Fuersen, K., Schreitmueller, S., & Walger, M. (2016). Effect of acoustic fine structure cues on the recognition of auditory-only and audiovisual speech. J Acoust Soc Am. 139(6):3116-3120.

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. Hear Res. 219(1-2):36-47.

Moon, I. J., Won, J. H., Park, M. H., Ives, D. T., Nie, K., Heinz, M. G., Lorenzi, C., & Rubinstein, J. T. (2014). Optimal combination of neural temporal envelope and fine structure cues to explain speech identification in background noise. J Neurosci 34:12145-54.

Münte, T. F., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. Nat Rev Neurosci. 3(6):473-8.

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. Proc Natl Acad Sci U S A. 104(40):15894-8.

Oswald, J. R. V. (1956). The theory of analytic band-limited signals applied to carrier systems. IRE Trans. on Circuit Theory. 3(4):244-251.

Parbery-Clark, A., Skoe, E., & Kraus, N. (2009a). Musical experience limits the degradative effects of background noise on the neural processing of sound. J Neurosci. 29(45):14100-7.

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009b). Musician enhancement for speech-in-noise. Ear Hear. 30(6):653-61.

Peretz, I., Blood, A. J., Penhune, V., & Zatorre, R. J. (2001). Cortical deafness to dissonance. Brain. 124:928-40.

Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: perceptual determinants, immediacy, and isolation after brain damage. Cognition. 68:111-41.

Pfeifer, R. (1988). Artificial intelligence models of emotion. Cognitive Perspectives on Emotion and Motivation, eds Vernon Hamilton, Gordon H. Bower, Nico H. Frijda (Springer Netherlands) pp 287-320.

Phan, K.L., Wager, T., Taylor, S.F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. Neuroimage 16:331-48.

Rauschecker, J. P. (1999). Auditory cortical plasticity: a comparison with other sensory systems. Trends Neurosci. 22(2):74-80.

Scherer, K. R. (1995). Expression of emotion in voice and music. J Voice. 9(3):235-248.

Schlaug, G., Jäncke, L., Huang, Y., Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. Science. 267(5198):699-701.

Seymour, B., & Dolan, R. (2008). Emotion, decision making, and the amygdala. Neuron 58:662-71.

Shamma, S., & Lorenzi, C. (2013). On the balance of envelope and temporal fine structure in the encoding of speech in the early auditory system. J Acoust Soc Am. 133(5):2818-33.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. Science. 270:303–304.

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. Nature. 416(6876): 87–90.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. Behav Res Methods Instrum Comput. 31(1):137-149.

Strait, D. L., Kraus, N., Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. Hear Res. 261(1-2):22-9.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. J Acoust Soc Am. 26(2):212-215.

Swaminathan, J., & Heinz, M. G. (2012). Psychophysiological analyses demonstrate the importance of neural envelope coding for speech perception in noise. J Neurosci. 32(5):1747-56.

Swaminathan, J., Reed, C. M., Desolge, J. G., Braida, L. D., & Delhorne, L. A. (2014). Consonant identification using temporal fine structure and recovered envelope cues. J Acoust Soc Am. 135(4):2078-2090.

Swets, J. A. (1961). Is there a sensory threshold. Science. 134(3473):168-177.

Swets, J. A. (1961). Stimulus vs response uncertainty in recognition. 33(11):1586-1592.

Swets, J. A. (1986). Indices of discrimination or diagnostic accuracy: their ROCs and implied models. Psychol Bull. 99(1):100-17.

Swets, J. A., Tanner, W. P. & Birdsall, T. G. (1961). Decision processes in peception. Psychol Rev. 68:301-40.

Verde, M. E., MacMillan, N. A., & Rotello, C. M. (2006). Measures of sensitivity based on a single hit rate and false alarm rate: the accuracy, precision, and robustness of d', Az, and A'. Percept Psychophys. 68(4):643-54.

Wan, C. Y., & Schlaug, G. (2010). Music Making as a Tool for Promoting Brain Plasticity across the Life Span. Neuroscientist. 16(5): 566–577.

Wildgruber, D., Riecker, A., Hertrich, I., Erb, M., Grodd, W., Ethofer, T., & Ackermann, H. (2005). Identification of emotional intonation evaluated by fMRI. Neuroimage 24:1233-41.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. J. Acoust. Soc. Am. 52:1238–1250.

Witteman, J., Van Heuven, V. J., & Schiller, N. O. (2012) Hearing feelings: A quantitative meta-analysis on the neuroimaging literature of emotional prosody perception. Neuropsychologia 50:2752-63.

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nat Neurosci. 10(4):420-2.

Zatorre, R. J. (2013). Predispositions and Plasticity in Music and Speech Learning. Science. 342(6158): 585-589.

Zatorre, R. J., & Belin, P. (2011). Spectral and temporal processing in human auditory cortex. Cerebral Cortex. 11:946–953.

Zeng, F. G., Nie, K., Liu, S., Stickney, G., Del Rio, E., Kong, Y. Y., & Chen, H. (2004). On the dichotomy in auditory perception between temporal envelope and fine structure cues. J. Acoust. Soc. Am. 116:1351–1354.

Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargave, A., Wei, C., & Cao, K. (2005). Speech recognition with amplitude and frequency modulations. Proc. Natl. Acad. Sci. USA. 102:2293–2298.

**Figures and tables**

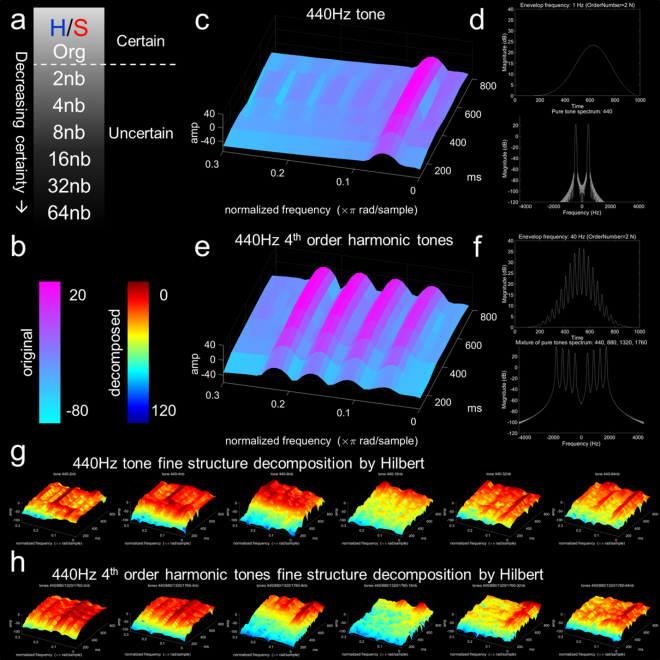
Table 1 | Participant data

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Good performer | | Poor performer | | Early Music Ed | | | Late Music Ed | | |
|  | Male | Female | Male | Female | | Male | Female | | Male | Female |
| Number | n=8 | n=8 | n=8 | n=8 | | n=8 | n=8 | | n=8 | n=8 |
| Age (yr) | 20.15±1.97 | 22.64±3.34 | 21.11±2.12 | 20.74±1.97 | | 20.08±1.62 | 20.56±1.09 | | 22.66±1.75 | 22.50±3.43 |
| Range | 18-26 | 18-25 | 19-26 | 19-29 | | 18-27 | 18-23 | | 19-27 | 19-30 |
| Music education | None | None | None | None | | 2±0 | 2±0 | | 8.00±2. | 7.16±2.48 |
| Correct identification | 97.65%  ±2.07% | 95.31%  ±2.21% | 80.07%  ±2.67% | 80.85%  ±3.96% | | 93.36%  ±2.00% | 92.58%  ±5.52 | | 89.45%  ±3.31% | 90.97%  ±3.96% |
| Total (n=32) | 31.25±0.66 | 30.50±0.71 | 25.63±0.86 | 25.88±1.27 | | 29.88±0.64 | 29.63±1.77 | | 28.63±1.06 | 29.11±1.61 |

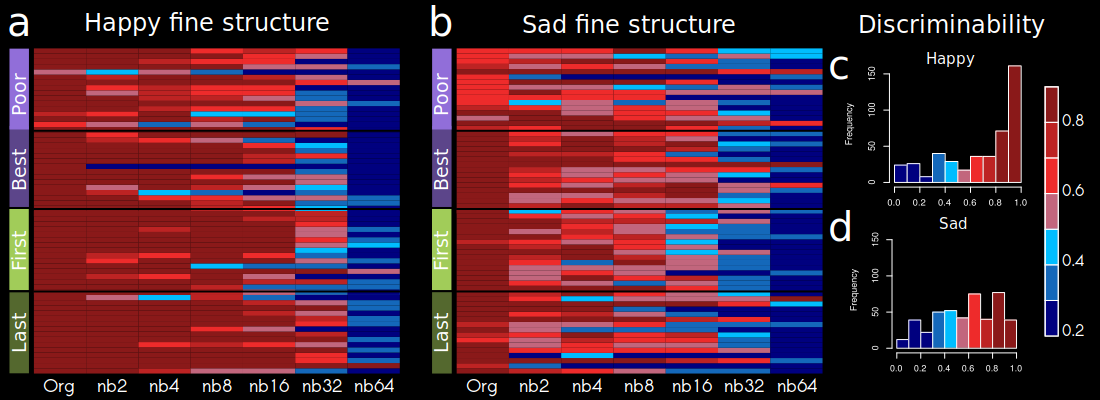
**Table S1 | Acoustic stimuli and their respective decompositions**

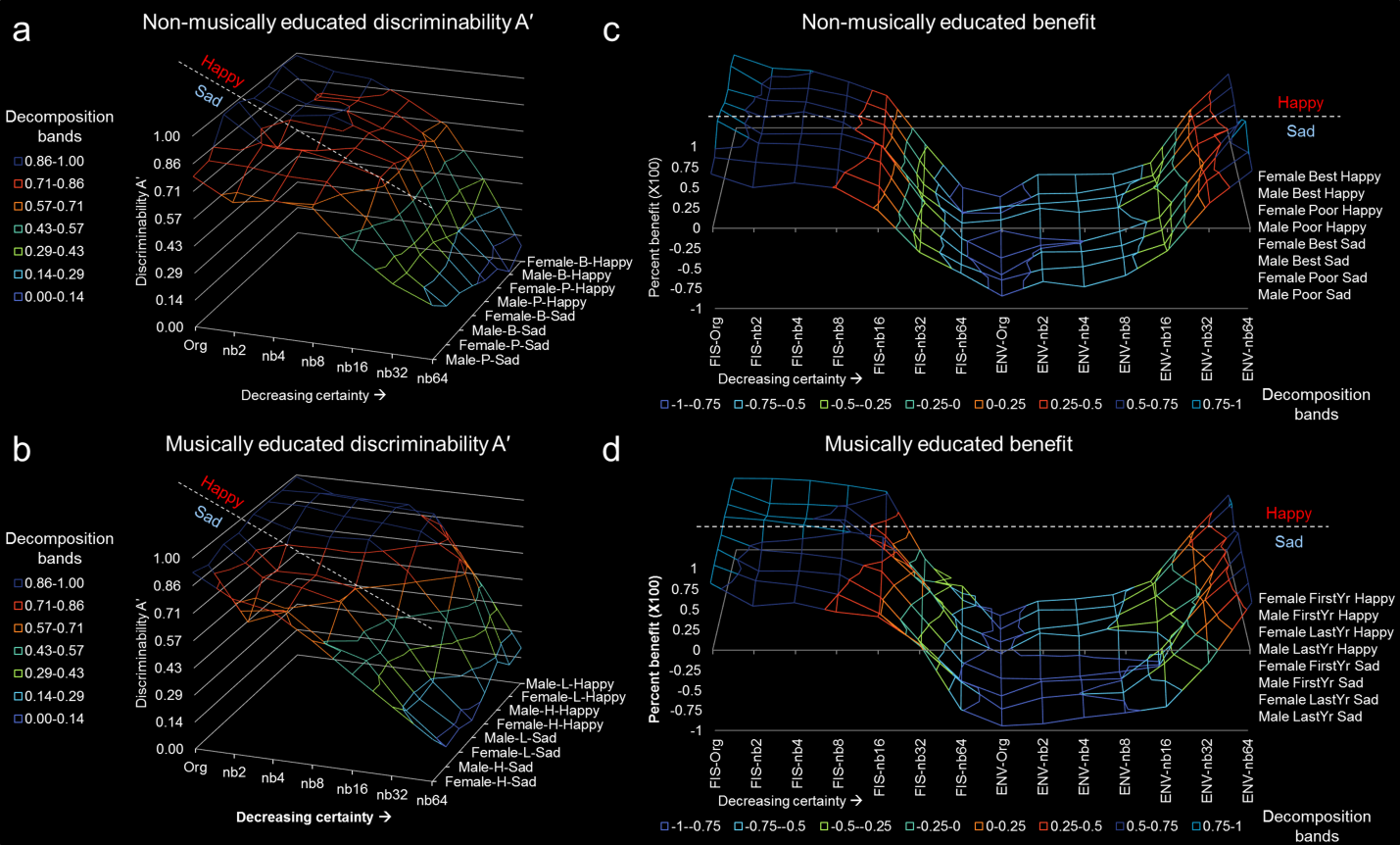
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Composer | Work | Measure(a) | M.M.(b) | Key | Instrumentation | Emotion |
| Beethoven | Piano Concerto no. 4 (3rd mvt) | 191–200(2) | 150 | G Maj | Piano & orchestra | happy |
| Beethoven | Piano Concerto no. 4 (3rd mvt) | 439–452(2) | 150 | G Maj | Piano & orchestra | happy |
| Beethoven | Symphony no. 3 (3rd mvt) | 38–56 | 180b | F Maj | Orchestra | happy |
| Beethoven | Symphony no. 6 (3rd mvt) | 9(3)–16(1) | 240 | D Maj | Orchestra | happy |
| Haendel | Utrecht's Te Deum | 5–14(1) | 112 | D Maj | Orchestra | happy |
| Mozart | Die Zauberflöte (Act 1 no. 2 Papageno's Aria) | 18(2)–24(2) | 80 | G Maj | Orchestra | happy |
| Mozart | Eine kleine nachtmusik (1st mvt) | 5(3)–10(3) | 154 | G Maj | String orchestra | happy |
| Mozart | Piano Concerto no. 23 (3rd mvt) | 1–8 | 255 | A Maj | Piano | happy |
| Mozart | Piano Concerto no. 27 (3rd mvt) | 1–8 | 167 | B flat Maj | Piano & orchestra | happy |
| Ravel | Tombeau de Couperin (Rigaudon) | Bar1–9(2) | 100 | C Maj | Piano | happy |
| Saint-Saëns | Carnaval des Animaux (Finale) | 10–26(4) | 220 | C Maj | Piano & orchestra | happy |
| Saint-Saëns | Carnaval des Animaux (La volière) | 1–9(2) | 88 | F Maj | Piano & orchestra | happy |
| Schumann | Kinderszenen (Op 15 no. 9) | 1–9 | 240 | C Maj | Piano | happy |
| Verdi | La Traviatta (Brindisi) | 1–15(1) | 100 | B flat Maj | Orchestra | happy |
| Verdi | Rigoletto (Act 1 no. 4) | 69–73 | 150 | C Maj | Orchestra | happy |
| Vivaldi | L'Autunno (1st mvt) | 1(2)–4(3) | 126 | F Maj | Orchestra | happy |
| Albinoni | Adagio | 7–14(1) | 48 | G min | Orchestra | sad |
| Bach | Passionsmusik nach dem evangelisten Matthäus | 1–5(2) | 67 | E min | Orchestra | sad |
| Brahms | Piano Concerto no. 1 (2nd mvt) | 21(3)–24(1) | 48 | D Maj | Piano & orchestra | sad |
| Bruch | Kol Nidrei | 9–11(1) | 20 | D min | Double bass & organ | sad |
| Chopin | Nocturne Op 27 no. 1 | 2(2)–6(3) | 72 | C sharp min | Piano | sad |
| Chopin | Nocture Op 48 no. 1 | 1–4(1) | 52 | C min | Piano | sad |
| Chopin | Nocturne Op 9 no. 1 | 0–4(1) | 100 | B flat min | Piano | sad |
| Debussy | Prélude: Des pas sur la Neige | 4–8(1) | 35 | D min | Piano | sad |
| Grieg | Peer Gynt's Suite no. 2 (Solveigs lied) | 13(4)–17(3) | 69 | A min | Orchestra | sad |
| Mahler | Symphony no. 5 (3rd mvt) | 12(4)–16(3) | 54 | A min | Orchestra | sad |
| Mozart | Piano Concerto no. 23 (2nd mvt) | 1–3 | 35 | F sharp min | Piano | sad |
| Rachmaninov | Piano Concerto no. 2 (2nd mvt) | 13(2)–17 | 48 | E Maj | Piano & orchestra | sad |
| Ravel | Concerto in G (2nd mvt) | 1–4(2) | 38 | E Maj | Piano & orchestra | sad |
| Rodrigo | Concerto de Aranjuez (Adagio) | 1–4(4) | 40 | B min | Guitar & orchestra | sad |
| Saint-Saëns | Carnaval des Animaux (Le cygne) | 1–5 | 55 | G Maj | Piano & cello | sad |
| Schubert | String Quartet no. 14 (2nd mvt) | 1–4 | 72 | G min | String quartet | sad |

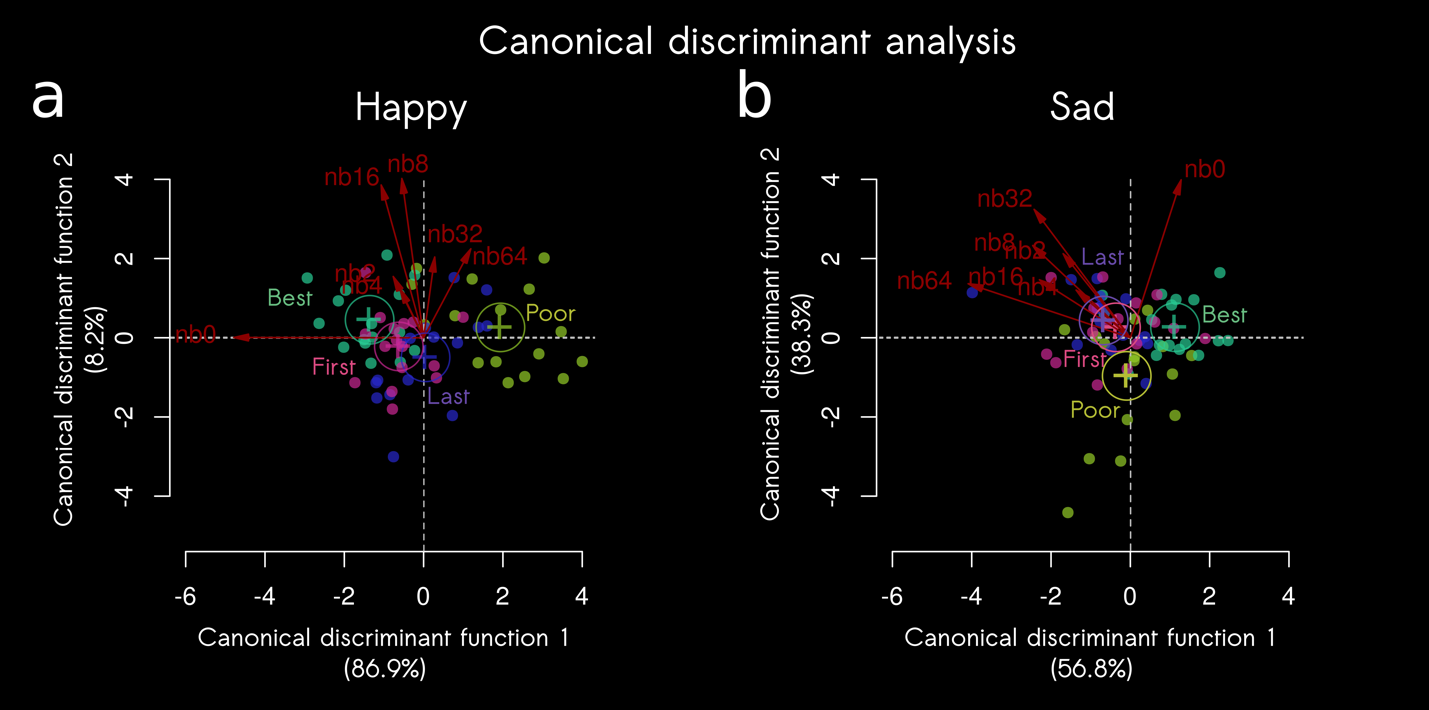
(a)=The number in parentheses indicates beat number in the measure. (b)=This value corresponds to a half note. Table Adopted from Peretz et al., 1998.



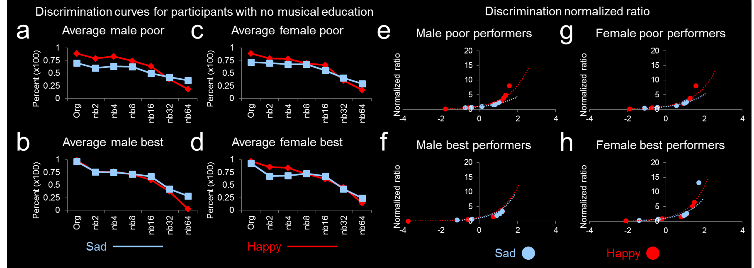
**Figure 1**. Experimental design and simple example decomposition. a) Happy (H)/sad (S) descending interval forced choice discrimination task with decreasing emotion certainty by decomposition. Original excerpts were taken from the Montreal Emotional Identification Task (Dalla Bella et al., 2001; Peretz et al., 1998; Peretz et al., 2001; Khalfa et al., 2008; Gosselin et al., 2007; Gosselin et al., 2011). Decompositions are based on perceptible categorization as happy or sad which pertain to certain (org – original) and increasingly uncertain (2nb, 4nb, 8nb, 16nb, 32nb, 64nb decompositions). b) Colorbar for coding spectrogram in original (cool colorbar) and decomposed (jet colorbar) stimuli (simple stimuli example decomposition). Colorbar represents normalized power/frequency (dB/Hz), or amplitude (sound pressure level dB) by fine structure components (Hz). All spectrogram plots contain magnitude (dB) on z-axis, normalized frequency (x π rad/sample) on x-axis, and time in milliseconds (ms) on y axis. c) Single 440Hz tone. d) Upper panel, single tone with extracted envelope at 1 Hz (order 2N Butterworth filter) with magnitude on the y-axis and time on the x-axis. Lower panel, fast Fourier transform plot of single tone 440 Hz with magnitude on y-axis and frequency (hertz, cycles per second) on x-axis. e) Complex 440Hz 4th order harmonics with 440Hz, 880Hz, 1320Hz, and 1760Hz components. f) Upper panel, complex tone with extracted envelope at 40Hz (order 2N Butterworth filter) with magnitude on the y-axis and time on the x-axis. Lower panel, fast Fourier transform plot of complex 440Hz 4th order harmonic tones with magnitude (dB) on y-axis and frequency hertz cycles per second on x-axis. g) Hilbert decomposition of single 440Hz tone fine structure with complex 4th order harmonics envelope. h) Hilbert decomposition of complex 4th order harmonics fine structure with single 440Hz tone envelope. Progression from left to right for both (g) and (h) represent the Hilbert transformation for this simple example with 2nb, 4nb, 8nb, 16nb, 32nb, and 64nb decompositions. The simple example demonstrates how a complex acoustic stimulus which is deemed emotional can be decomposed.

**Figure 2.** Heatmap accuracy of response by group and uncertainty. Discrimination heatmap for accuracy concerning uncertain (a) happy fine structure accuracy and (b) sad fine structure accuracy. Greater red color represents greater discriminability of emotion and similarity with the original un altered excerpt. On the x-axis, band number is represented with Org unaltered original excerpt to nb64. The discriminability is represented by identification accuracy with the unaltered excerpt. (c) Happy discriminability and (d) sad discriminability. The color bar represents discriminability 20th percentile bands.

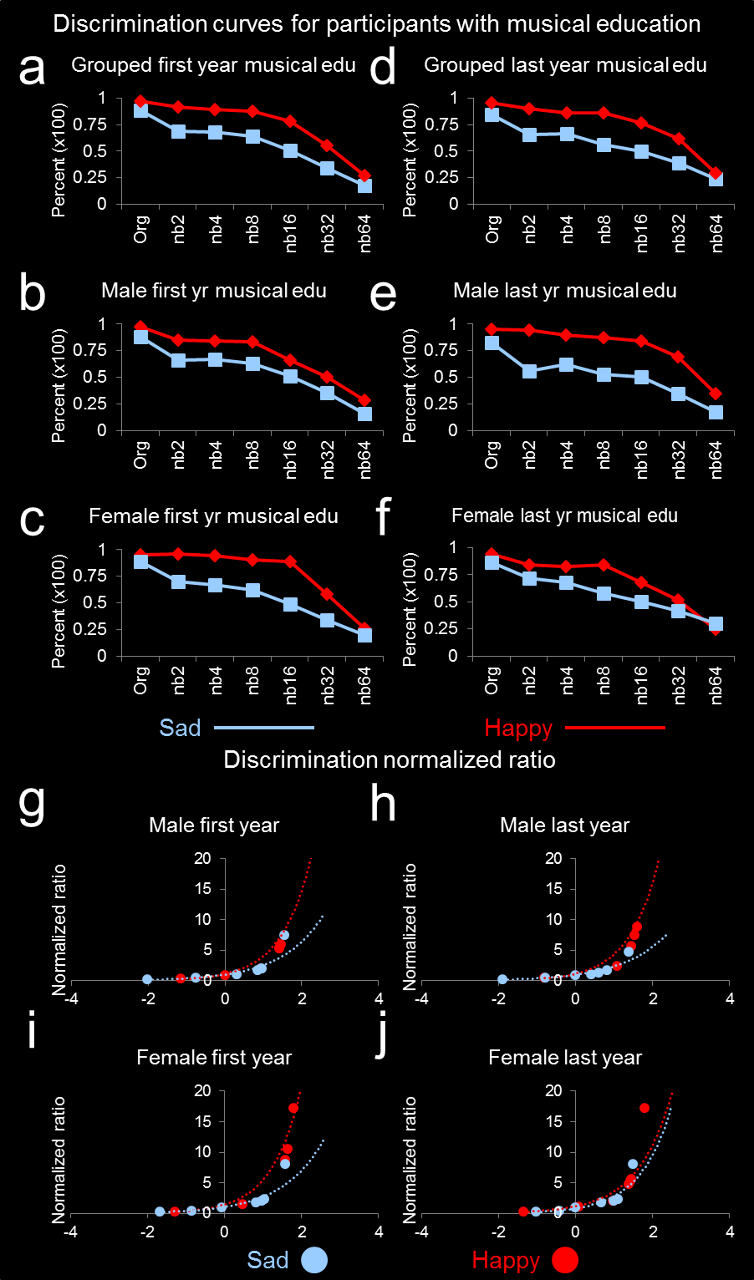
**Figure 3**. Discriminability A′ and FIS/ENV benefit to stimuli perception of uncertain emotion. (a) Non-musically educated discriminability A′ and (b) musically educated discriminability A′. (c) Non-musically educated benefit and (d) musically educated benefit. For all figures, the x-axis represents band number from Org to nb64. The band decompositions were associated in ranges to represent the discriminability. The z-axis represents sex (male or female), type of group (B – best, P – poor, L – first year musician, H – last year musician), and emotion type (happy or sad). Here groups are connected to visualize the trend and pattern more readily. For example (a and b), all groups interpret a stimuli with greater certainty i.e. greater discriminability A′ shows more blue coloring for happy than sad. The dashed line represents the grouping of happy and sad. Note, happy is much more fuller than sad, indicating high A′. For benefit (c and d), note for FIS there is a fuller grouping, representing that the majority of individuals use fine structure to discriminate uncertain emotion. Note the subtle difference between happy and sad, represented by the dashed line. Musically educated individuals benefit from FIS more than non-musically educated individuals.

**Figure 4**. Canonical discriminant analysis for (a) happy and (b) sad uncertain emotion discrimination. Group is sorted by poor performer, best performer, first year musical education and last year musical education with band number from Org to nb64 in red. The plot shows the canonical scores for the groups defined by the term as points and the canonical structure coefficients as vectors from the origin. Standardized beta coefficients are given for each variable in each discriminant (canonical) function, and the larger the standardized coefficient, the greater is the contribution of the respective variable to the discrimination between groups. Here, the discriminant function coefficients denote the unique contribution of each variable to the discriminant function, while the structure coefficients denote the simple correlations between the variables and the functions. For sad, the greatest standardized beta coefficients were org, Can1 = 0.6264 and for nb64 was Can1 = -0.7641. For happy, the greatest standardized beta coefficients were org, Can1 = -1.011 and for nb64 was Can1 = 0.4193.

**Supplementary Information**



**SI Figure 1**. Discrimination curves and discrimination normalized ratio indices for participants with no musical education for uncertain emotion. Discrimination curves of percent identification of happy or sad stimuli. (a) Average male poor performers, (b) average male best performs, (c) average female poor performers, (d) average female best performers. Sad curve represented with blue line and happy curve represented red line. The discrimination normalized ratio of stimuli identification of happy or sad stimuli. (e) Male poor performers, (f) male best performs, (g) female poor performers, (h) female best performers. Sad discrimination ratios are represented in red and happy discrimination ratios are represented in blue.



**SI-Figure 2**. Discrimination curves and discrimination normalized ratio indices for participants with musical education in their first or last year of study. Discrimination curves of percent identification of happy or sad stimuli. (a) Grouped first year (low) and (d) grouped last year musical education (high). Male (b) and female (c) first year and male (e) and female (f) last year musical education. Sad curve represented with blue line and happy curve represented red line.