Stream segregation revisited: dynamic listening and influences of emotional context on stream perception and attention

Renee Timmers[*1], Yuko Arthurs[*2] & Harriet Crook[*#3]

[*] Department of Music, The University of Sheffield

[#] Department of Audiology, Royal Hallamshire Hospital

[1] r.timmers@sheffield.ac.uk, [2] y.arthurs@bathspa.ac.uk, [3] harriet.crook@nhs.net

Accepted for publication in *Consciousness & Cognition*

Highlights

- Emotional context influences perception of auditory streams
- Listeners dynamically adapt auditory attention to emotional context
- Emotion and musical engagement are intrinsic to music perception processes

Abstract

A classical experiment of auditory stream segregation is revisited, reconceptualising perceptual ambiguity in terms of affordances and musical engagement. Specifically, three experiments are reported that investigate how listeners' perception of auditory sequences change dynamically depending on emotional context. The experiments show that listeners adapt their attention to higher or lower pitched streams (Experiments 1 and 2) and the degree of auditory stream integration or segregation (Experiment 3) in accordance with the presented emotional context. Participants with and without formal musical training show this influence, although to differing degrees (Experiment 2). Contributing evidence to the literature on interactions between emotion and cognition, these experiments demonstrate how emotion is an intrinsic part of music perception and not merely a product of the listening experience.

Keywords: auditory scene analysis; emotion; music; attention; auditory perception; musical training

1. Introduction

Recent models of perception and cognition emphasise the interactive and dynamic nature of perception and ways in which perception is intertwined with action-goals, environmental affordances, and observers' perception-action histories (e.g. Barsalou, 2003; Chemero, 2003; Bishop & Martin, 2014; O'Regan & Noë, 2001). This dynamic characteristic is also highly relevant for the perception of music, which often affords multiple and diverse ways of engagement and perception (Clarke, 2005; Krueger, 2011, 2014; Myin, 2016; Reybrouk, 2012, 2014; Schiavio, & van der Schyff, 2016). Classically interpreted as structural ambiguity, listeners may for example entrain to music at different temporal levels (Parncutt, 1994; London, 2012; Drake, Jones & Baruch, 2000), they may differ in attention given to melodic voices, perceiving some to be in the foreground and others in the background (Bigand, McAdams, & Forêt, 2000; Sloboda & Edworthy, 1981; Teki, Chait, , Kumar, von Kriegstein, & Griffiths, 2011), and differ in their patterning of melodic material as belonging to a single voice (or stream) or separate voices (e.g. Bendixen, Denham, & Winkler, 2014, Davis, 2011). We argue that differences in perception of music should be investigated in the context of types of engagement with music.

Specifically, we will investigate processes of auditory stream perception and examine the role of emotions in influencing dynamic attending to auditory streams. In the classical auditory stream segregation experiments as used by Van Noorden (1975), a sequence of alternating high and low tones is presented that can be heard as a single stream of a melody that goes up and down, or it can segregate into two separate streams, one consisting of high tones and the other of low tones. Whether the sequence is integrated or segregated is both a function of the time interval and the pitch interval between tones. Other characteristics may also influence stream perception such as differences in intensity or timbre (Bregman, 1990). Our study examines how contextual factors, specifically emotional context, play a role in listeners' dynamic attention, including stream integration or segregation and attention to streams that are higher or lower in pitch. We use the notion of dynamic attention to refer to processes of predictive focus in time and pitch that are common to music perception and does not require explicit attention (Jones & Boltz, 1989; Desain, 1992; Pearce & Wiggins, 2012).

Early work on stream segregation suggested a model of mandatory or automated function that was driven by bottom up processing controlled by the parameters of the incoming stimulus in particular pitch, duration, location and intensity (Beauvois & Meddis, 1996; Hartmann & Johnson, 1991; Van Noorden, 1975). For example, Hartmann and Johnson (1991) employed a "peripheral channelling" model whereby frequency separation by narrow band filtering in the cochlear would allow the streams to segregate. They hypothesized that sounds similar in spectral regions and presented to the same ear would not be likely to segregate. However, others have challenged this model including Bregman (1990) suggesting two mechanisms, a stimulus driven and a schema driven mechanism, supported by experimental data that showed it was possible to stream stimuli with the same central frequency (Moore, Gockel & Hedwig, 2002). In the absence of cues to separate streams (e.g. different timbres), it is still possible to detect a melody in the context of distractor tones that have similar pitch heights, based on top-down expectations, i.e. knowing the melody (Devergie, Grimault, Tillmann, & Berthommier, 2010). Top-down influences are further demonstrated in experiments that vary patterned perception of sequences or prime perception in ways that promote stream segregation or integration (Bendixen et al., 2014; Rogers & Bregman, 1993, respectively). The framework of predictive coding provides a useful dynamic perspective on perception where bottom-up and top-down processes are closely interlinked (Heilbron & Chait, 2018). In a review article, Denham & Winkler (2018) discuss how processes of auditory stream segregation and integration can be usefully interpreted from a predictive coding perspective, but also pose a number of challenges to the framework, such as the relevance of prediction error.

Taking an embodied and enactive perspective (e.g. Thompson, 2007), perception cannot be understood as an isolated phenomenon, separated from our active engagement and interaction with the world. In musical contexts, this means that perception of music is closely intertwined with our emotional engagement with it, given the close relationship between music listening and emotional experiences (Juslin et al., 2008; Randall, Rickard, & Vella-Brodrick, 2014; Zentner, Grandjean, & Scherer, 2008). Music perception and emotion are too often primarily understood as the end-products of static cognitive processes. Instead, they are likely to operate in tandem through mutual influence.

Whilst music has been used as an “emotion inducer” in studies investigating interactions between emotion and cognition (e.g. Boltz, 2001; Niedenthal, Halberstadt, & Innes-Ker, 1999; Marin, Gingras, & Bhattacharya, 2012), only a few studies have looked at such interactions in the context of music listening (Boltz, Ebendorf, & Field, 2009; Houston & Haddock, 2007; Timmers & Crook, 2014). To our knowledge, no previous work has considered emotional influences on the perception of auditory sequential streams. Here, we will examine the hypothesis that emotional context may influence perception of auditory streams in musical contexts: specifically, we argue that changes in perceived emotional context may influence listeners’ musical expectations, which in turn influences musical stream perception. As argued and demonstrated in Timmers & Crook (2014), emotional context may prime listeners to expect musical properties that are closely associated with the perceived emotion. Specifically, to predict how emotional context and dynamic attention to auditory streams may interact, we can consider different *Gestalts* (or sensorimotor patterns) associated with music-emotion categories. For example, and put simply, happy music is associated with positivity and high energy. Correspondingly, happy music frequently has a relatively fast tempo, high pitch, upwards and large pitch intervals. In contrast, sad music is associated with negativity and low energy, corresponding with e.g. slow tempo, low pitch, downwards, and small intervals (Gabrielsson & Lindström, 2010). Applying these associations to the perception of an auditory stream of alternating tones that vary in pitch interval and time interval, we predict that happy contexts will increase listeners' tendency to attend to the higher pitched tones, while sad contexts will increase the tendency to focus attention to the lower pitched tones. Furthermore, high energy levels associated with happy contexts may enhance stream integration, as this corresponds to faster melodic movement and also larger pitch intervals, while sad contexts instead may strengthen stream segregation into two slower paced melodic figures with small pitch intervals (or rather no melodic change – a repeating tone). Figure 1 illustrates these predictions.

Assuming that perception is strongly related to musical engagement, we expect to see differences in perception between participants who differ in musical background and training. Such differences have occasionally been tested in auditory stream segregation studies with varied results (Bigand et al., 2000; Jones, Jagacinski, Yee, Floyd, Klapp, 1995). Here, we include this as one of our objectives, tested in one of the experiments. We predict that musically trained participants have stronger expectations regarding relationships between emotional context and musical patterning, in line with previous work that showed enhanced musical expectations in musically trained participants (Hansen & Pearce, 2014; Park et al. 2018).

Three experiments were run to test the hypotheses of an influence of emotion on attention and on stream-segregation vs. integration. Emotional context was varied by presenting emotional pictures alongside auditory stimuli. This assumes that listeners process multimodal stimuli in an integrated manner: specifically, that they integrate the emotion perceived in auditory and visual domains. This is in line with empirical work that has demonstrated an influence of emotional musical stimuli on the processing of visual information (e.g. in the context of film, for a review see Tan, 2017) and vice versa an influence of emotional visual stimuli on evaluations of the emotion of musical stimuli (e.g. in the context of music performance, Vines et al., 2011). Only a few studies have investigated influences of emotional context on the perception and memorisation of music (Houston & Haddock, 2007; Boltz et al., 2009), which is in our view an important research gap that this work aims to address (see also Timmers & Crook, 2014).

The experiments closely simulate the classical auditory stream segregation experiments in terms of auditory material (e.g. Van Noorden, 1975; Rogers & Bregman, 1993). The main difference with the classical work is the variation of emotional context by presenting happy, neutral or sad pictures along with the auditory stimuli. Furthermore, to increase ecological validity, musical instrument sounds are used (rather than sine tones). Van Noorden (1975) used psychoacoustic determination of listening thresholds, through gradual adjustments of stimuli characteristics. Later experiments employed paradigms with fewer levels of stimuli characteristics, testing the effects of various stimuli properties, primes and contexts (for a review see Moore & Gockel, 2012). As stream perception changes whilst listening[1], studies may require participants to continuously monitor perception and indicate changes therein (e.g. Bendixen et al., 2010). Others asked listeners to focus on the perception at the end of a sequence (e.g. Carlyon, Plack, Fantini, & Cusack, 2003; Micheyl et al., 2013) or to give a summative rating (e.g. Rogers & Bregman, 1993). The three experiments reported in this paper build on the latter and ask participants to indicate a summative response after listening. They are similar to each other, except for variations in details of stimuli and response modes, providing an opportunity for replication and validation across different

[1] This includes a build-up from stream integration towards segregation after the start of a sequence and "flipping" between different percepts, indicating bi-stability of stream integration and segregation.

participant samples and experimental settings. Experiment 1 employs a rating paradigm, asking participants to reflect on the degree of attention paid to the top or bottom stream, and on the degree of stream integration or segregation. It uses musical stimuli that are perceived as a "galloping" pattern when integrated and as two isochronous patterns (one high and one low pitched) if segregated (see Figure 1). Experiments 2 and 3 employ auditory sequences consisting of regularly alternating high and low pitches, in order to balance note density and therefore the speed of tones in each stream. Experiment 2 compares two groups of participants that differ in degree of musical training, and simplifies the response mode by using a forced choice response paradigm, where participants are asked to indicate their predominant perception and their second most dominant perception of the sequence. Finally, Experiment 3 tests perception of the auditory streams without using a reference to "low" or "high" pitch. This is realised by adding a stepwise motion in either the low or the high stream, and by asking participants to indicate whether they heard the sequence as integrated, or whether they focused on the repeated tone or the stepwise motion. The experiments always test the same hypotheses of an effect of emotional context on attention to the low or high pitched stream and on stream integration versus segregation (see Figure 1). Later experiments employ response modes that are more implicit. They include stronger emotional visual stimuli and employ auditory stimuli that are balanced (Experiment 2) or systematically varied (Experiment 3) in pitch material.

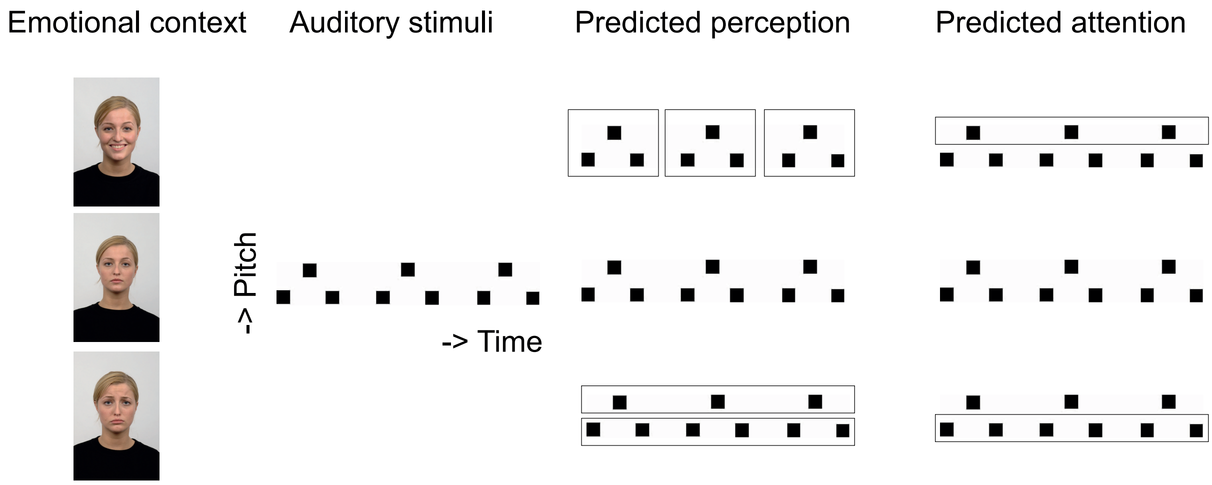

Figure 1: Illustration of auditory stimuli of Experiment 1 and their predicted perception as integrated (top right) or segregated (bottom right) in a happy, neutral or sad context (left)

2. Experiment 1

2.1 Materials and method

2.1.1 Design

A factorial within-participants design was used that varied emotional visual context (3 levels: happy, neutral, sad), time interval between tones (3 levels: 120 ms, 150 ms, 180 ms), and pitch interval between tones (2 levels: 5 or 6 semitones). Two dependent measures were included related to the subjective indication on a seven-point scale of attention to the higher or lower tone and perception of the sequence as integrated or segregated.

2.1.2 Participants

35 adults participated in the experiment (13 M, 22 F, median age = 34). All participants had played music in the past for more than 5 years or were still performing actively. Participants had played an instrument[2] for a median of 14 years, and had taken lessons for a median of 12 years. Sample size was determined on the basis of a previous study, in which we found a strong influence of emotion on musical expectation for pitch

[2] Including voice

register (Timmers & Crook, 2014). Experiment 3 from Timmers & Crook (2014) was closest to the current study in using a rating paradigm, which showed a strong effect size for the interaction between pitch register and emotion (eta_p^2=.346), although it only found a weak effect size for the interaction between interval size and emotion (eta_p^2=.076). Power calculations using G*Power indicated that a sample size of 35 is sufficient to demonstrate a modest effect of eta_p^2 = 0.13 with an alpha level of 0.05 and a power of 0.8 in the context of a repeated measures ANOVA with 3 levels[3]. The sample size of 35 was deemed sufficient for the current experiment, as the experiment tests the association between emotion and pitch register and the association between emotion and tempo and both are known to be strong associations (see e.g. Timmers & Crook, 2014; Gagnon & Peretz, 2003).

2.1.3 Material – emotional context

Photos taken from a database of facial portraits of people portraying emotional expressions were used as emotional material. The photos were portraits from the front taken from the Radboud Affective Faces Database (RAFD, Langner, et al., 2010). Portraits of 20 male and 20 female subjects were included, leading to a total of 120 pictures for three emotions (40 happy, 40 sad and 40 neutral).

2.1.4 Material – auditory sequences

A piano sound was selected to generate auditory sequences using a MIDI sequencer. The sequences consisted of a repeating pattern of low and high tones with a rest following every second low tone (Low-High-Low-Rest, Low-High-Low–Rest). The distance between the tones in a sequence was either 5 or 6 semitones, corresponding to a perfect fourth or an augmented fourth, respectively. Sequences were always 7 seconds in duration. The pitch height of sequences was varied across trials. A total 12 different pitch heights were used, corresponding to 12 different low tones of a sequence (6 for each type of sequence).

[3] These calculations take into account that the partial eta squared effect sizes were derived from SPSS.

The time-interval between successive tones was either 120 ms (fastest tempo), 150 ms (intermediate tempo), or 180 ms (slowest tempo). If a tone was followed by a rest, the interval to the next tone was twice as long. Duration from note onset to offset was always 50 ms.

These time and pitch intervals and the auditory pattern of low-high-low-rest correspond to stimuli properties used in previous research (e.g. Van Noorden, 1975). The duration of 7 seconds was chosen on the basis that listeners tend to only start to segregate pitches of a sequence after a few seconds (for a review, see Moore & Gockel, 2012).

2.1.5 Procedure

Participants saw an emotional picture and were asked to imagine the emotion portrayed in the picture. After 7 seconds, the musical sequence started to play. After the presentation of the picture and musical sequence, participants indicated their perception of the sequence using two seven-point rating scales: First they evaluated the extent to which they had perceived the sequence as one or two streams, where 1 indicated “strongly as one stream” and 7 indicated “strongly as two streams”. A rating of 4 meant “both as one and as two streams”. Secondly, participants evaluated the extent to which they had heard the low or high pitches as being on the foreground, where 1 meant “low notes on the foreground”, 4 meant “low and high notes on the foreground” and 7 meant “high notes on the foreground”.

Participants gave informed consent before participating in the study and practiced the task before starting the experimental trials. In the experiment, each participant received three trials of each condition in random order with randomly selected pitch heights and faces.

2.1.6 Data processing

Data was averaged across the three repetitions of a condition for each participant, resulting in a single data point per condition for each participant. Due to averaging across several trials, distributions of variables were normal or close to normal, indicating the suitability of the use of parametric tests for data analysis[4]. Descriptive statistics (mean and SE) per experimental condition and dependent variable are provided for reference in Appendix 1, Table A1.

2.2. Results

Two repeated measures ANOVAs were run to test the effects of emotion, tempo and pitch interval on each type of attention rating. For the first type of rating of attention to low or high tones, the effect of emotion was the only significant effect (see Table 1). Planned contrasts confirmed a significant linear relationship between emotion (coded as -1 for sad, 0 for neutral, and 1 for happy) and attention ratings (p = .024), as illustrated in Figure 2: attention to higher compared to lower tones increased from sadness to happiness as predicted. Pairwise comparisons confirmed a significant difference between attention ratings in Happy compared to Sad contexts (p = .024). The differences in attention ratings with the Neutral contexts failed to reach significance. The difference in ratings between Neutral and Happy was relatively close to significant (p = .071).

For the other rating of stream integration, only the main effects of tempo and interval were significant. The effect of emotion was not significant, nor were any of the interactions significant (see Table 1). The effect of tempo on stream segregation was as expected: at faster speeds, participants perceived more stream segregation than at slower speeds (T120: M = 4.537, SE = 0.139; T150: M = 3.960, SE = 0.156; T180: M = 3.435, SE = 0.185). Planned contrasts confirmed a significant linear relationship between tempo and stream integration (p < .001). The main effect of pitch interval was also as expected with more stream segregation with a larger pitch interval (M = 4.143, SE = 0.150), than with a smaller pitch interval (M = 3.812, SE = 0.169).

[4] The effects of emotion on stream perception reported in the results section are similar irrespective of whether parametric or non-parametric tests are conducted. For simplicity, only parametric tests are reported.

Table 1: Overview of statistical results of Experiment 1.

			Rating of attention to low-high pitch			Rating of stream integration		
Effect	df model	df error	*F*	*p*	eta_p^2	*F*	*p*	eta_p^2
Emotion	2	68	3.927	.024	.104	0.134	.875	.004
Tempo	2	68	0.853	.431	.024	42.544	<.001	.557
Pitch	1	34	0.852	.363	.024	6.621	.015	.163
Emotion * Tempo	4	136	1.043	.387	.030	0.481	.750	.014
Emotion * Pitch	2	68	0.340	.713	.010	1.918	.155	.053
Tempo * Pitch	2	68	0.621	.541	.018	0.757	.473	.022
Emotion * Tempo * Pitch	4	136	0.474	.754	.014	0.623	.639	.018

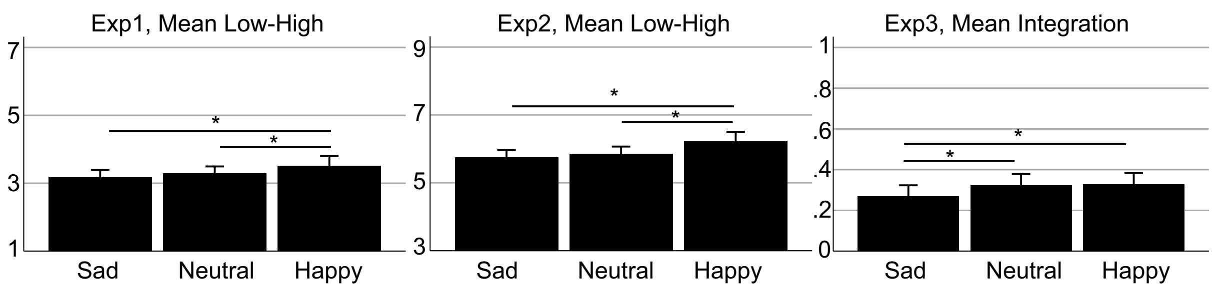

Figure 2. Means and error bars (95% confidence intervals) of dependent variables (attention to low-high stream or degree of stream integration versus segregation) that showed a significant influence of emotion on stream perception. * $p < .05$

2.3 Discussion

The first experiment confirmed that emotional context may influence attention to lower or higher pitches. However, it did not confirm that emotional context influences the tendency to integrate or segregate sequences. Instead tempo influenced stream segregation most strongly, followed by pitch interval. Some participants commented that the lower pitches had more prominence than the high pitches because they occurred more frequently. Participants also commented on the use of an augmented fourth, which has a negative connotation. Furthermore, the tempo of the stimuli influences its affective character, interacting with the presented visual emotional context. These characteristics of the musical material are improved in the second experiment – using only a single tempo, avoiding the augmented fourth and using equally frequent low and high pitches. Additionally, in the second experiment, we aimed to improve emotion induction by presenting several emotional pictures during a musical sequence, redrawing attention to the emotional context while listening, and by presenting stimuli for each emotion in a single block. Emotional pictures of various scenes from the International Affective Pictures Database (IAPS) were used in addition to the photographs of faces with emotional expressions. Finally, response measures were simplified to a forced choice among three options (rather than 7) to make responses as intuitive as possible and facilitate participation in the experiment.

3. Experiment 2

3.1 Method

3.1.1 Design

A factorial mixed design was used with emotion (3 levels: happy, neutral, sad) and pitch pattern (3 levels: simple-fourth, simple-fifth, complex) as within-participant variables and musicianship as between-participant variable. Dynamic attention was measured by asking participants to indicate their predominant perception of the sequence and their second most dominant perception of the sequence.

3.1.2 Participants

42 adults participated in the second experiment. 21 participants had received more than 5 years of musical training (median of 12 years of music performance), were making music regularly and were classified as "musicians" (10 M, 11 F, median age = 22.0), while the remaining participants had received little or no musical training and were classified as "non-musicians" (9 M, 12 F, median age = 46.0). Given the between-participant design, the overall sample size was increased in comparison to Experiment 1. Power calculations using G*Power indicated that a sample size of 42 is sufficient to demonstrate an effect with medium effect size (eta_p^2=.115) with a power of 0.8 and alpha of 0.05, in the context of a mixed ANOVA with two groups and a three-level within-participants variable. A stronger effect size was expected than observed in Experiment 1, given the strengthening of the induction of emotion in this experiment and the reduction of potentially interfering factors (i.e. negative connotation of the pitch interval).

3.1.3 Material – emotional context

Two types of affective pictures were used for this experiment, consisting of photos taken from RAFD (Langner, et al, 2010) and photos from IAPS (Lang, 1995). Brief videos of 12 seconds were created by presenting three affective pictures in a row, each for four seconds. The videos either displayed pictures taken from RAFD or from IAPS. A video always only contained pictures of one type of emotion (either happy, neutral or sad). Per emotion, two sets of photos from each database were used (4 selections) and presented in 6 different orders, generating a total of 72 videos: 6 orders of pictures x 4 selections of photos x 3 emotions.

3.1.4 Material – auditory sequences

A marimba sound was used to generate musical stimuli using a MIDI sequencer. The stimuli consisted of a continuous stream of alternating high and low pitches with 200 ms between successive tones. A simple and complex pattern of alternating tones was used. Patterns 1 and 2 consisted of two tones only – a simple alternation of a low and high pitch a fourth apart (simple-fourth) or a fifth apart (simple-fifth). Pattern 3 was a more complex pattern and consisted of four tones – two high and two low (see Figure 3). Both high and low streams contained a stepwise motion of a whole tone. The distance between the voices was a perfect fourth.

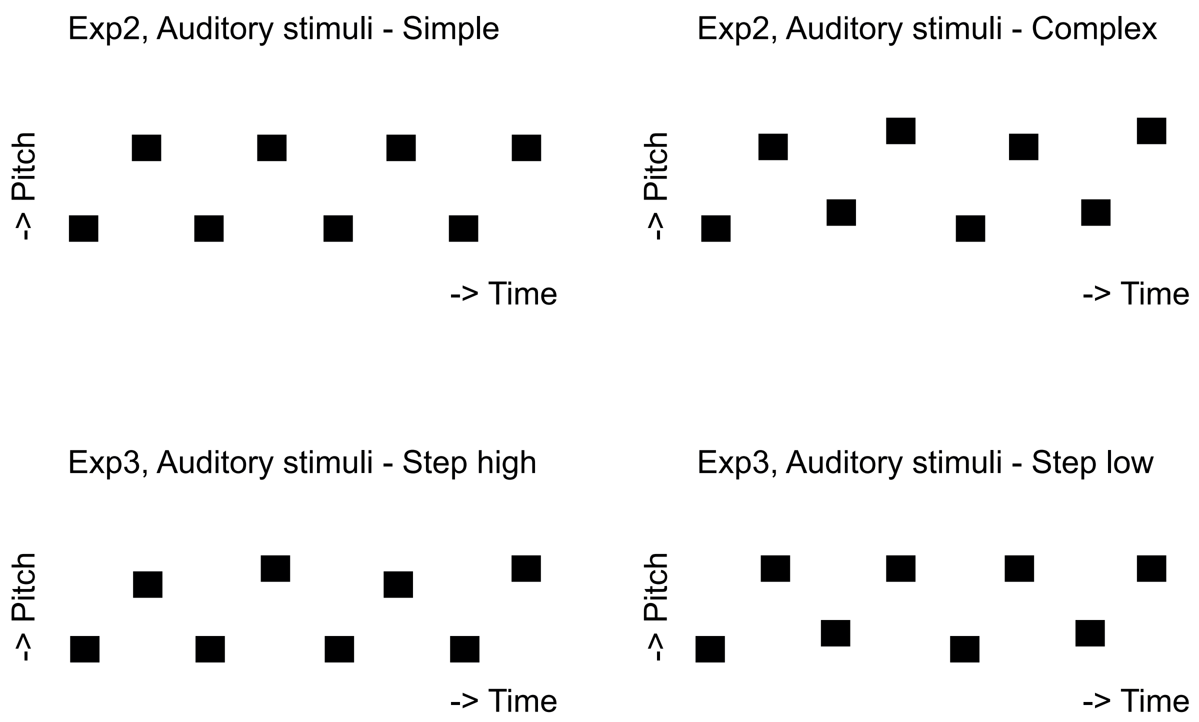

Figure 3. Illustration of auditory material in Experiments 2 (top) and 3 (bottom). Experiment 2 material consisted of simple alternation of high and low pitches, or a more complex alternation that also contains stepwise motion in each stream. Experiment 3 had the stepwise motion in either the high- or low-pitched stream.

All sequences started with a fade in and were 10 seconds long. The fade in was used to make it ambiguous whether sequences started with a high or low tone, which could influence attention allocation to the top or bottom voice. The stepwise motion in each voice in the complex sequence was added to give each voice a melodic presence of its own. Nevertheless, the complex pattern was still easily perceivable as an integrated pattern. The sequences were played at 12 different pitch heights. Half of the sequences started with a high tone and the other half with a low tone.

3.1.5 Procedure

Participants saw an emotional video and were instructed to imagine the emotion portrayed in the video. After 2 seconds an auditory sequence would start. The video and auditory material would end at the same time after 12 seconds from the start of the video, where after participants indicated their predominant perception of the sequence using a forced choice of three options: they could have listened predominantly to the high tones, to the low tones or to high and low tones together as an integrated pattern. Next, they indicated the second predominant perception of the sequence. If they had only heard the sequence in one way, they were asked to respond twice with the same answer.

A blocked procedure was used grouping trials per emotion. The order of emotion blocks was randomized across participants. Each emotion block had 16 trials – 8 complex and 8 simple. Within each emotion block, musical stimuli were presented in a random order. Pitch height of the sequence and type of video (showing photos from IAPS or RAFD) were randomly selected as long as the video represented the correct emotion. Each emotion block had 8 trials: 4 trials with a complex pattern and 4 trials with a simple pattern (two sequences with tones a fourth apart and two sequences with tones a fifth apart). Participants gave informed consent before participating in the study and practiced the task before starting the experimental trials.

3.1.6 Data processing

The responses of one of the musicians was discarded due to missing data. Two dependent variables were calculated related to 1) degree of attention to lower or higher stream, 2) degree of stream integration. Each variable was defined as a weighted sum of the two responses that participants gave, where the first response (predominant response) was weighted double. For the attention measure (1), the sum of attention from low to high was calculated, where attention to low was coded as 1, attention to both as 2, and attention to high as 3. The resulting total weighted sum of responses ranged from 3 (2x 1 + 1) to 9 (2x 3 + 3). For the stream integration measure (2), attention to low or high was coded as 0 (stream segregation) and attention to both streams was coded as 1 (stream integration). Subsequently the weighted sum ranged from 0 (2x 0 + 0) to 3 (2x 1 + 1). Responses were averaged across 2 trials for simple patterns and across 4 trials for complex patterns.

Due to averaging across several trials, distributions of variables were normal or close to normal, indicating the suitability of the use of parametric tests for data analysis[5]. Descriptive statistics (mean and SE) per experimental condition and dependent variable are provided for reference in Appendix 1, Table A2.

3.2 Results

Two univariate ANOVAs were run with emotional context and pattern as within-participant variables, musicianship as between-participant variable and attention to low-high pitch or stream integration as dependent variable. The main statistical results are presented in Table 2.

[5] The effects of emotion on stream perception reported in the results section are similar irrespective of whether parametric or non-parametric tests are conducted. For simplicity, only parametric tests are reported.

Table 2: Overview of statistical results of Experiment 2.

			Attention to low-high			Stream integration		
Effect	df model	df error	*F*	*p*	eta_p^2	*F*	*p*	eta_p^2
Emotion	2	78	4.970	.009	.113	1.805	.171	.044
Pattern	2	78	2.546	.085	.061	6.216	.003	.137
Musician	1	39	4.627	.038	.106	3.159	.083	.075
Emotion * Pattern	4	156	2.226	.069	.054	0.755	.556	.019
Emotion * Musician	2	78	0.312	.733	.008	3.042	.053	.072
Pattern * Musician	2	78	1.067	.349	.027	0.738	.468	.019
Emotion * Pattern * Musician	4	156	2.450	.048	.059	1.101	.358	.027

3.2.1 Attention to the lower or higher stream

For attention to the lower or higher stream, the main effects of emotion and musician were significant as well as the three-way interaction between emotion, pattern and musicianship (see Table 2). No other effect or interaction was significant. For the main effect of emotion, planned contrasts confirmed a significant linear relationship between emotion (coded as -1 for sad, 0 for neutral, and 1 for happy) and attention to the lower stream (p = .011), as illustrated in Figure 2 (middle panel), attention to low-high stream increased from sad to neutral to happy. Pairwise comparisons confirmed a significant difference in attention scores between the Happy and Sad contexts (p = .011) and between the Happy and Neutral contexts (p = .020).

The effect of musicianship was such that non-musicians (*M* = 6.098, *SE* = 0.103) indicated on average more attention to the higher stream than musicians (*M* = 5.779 *SE* = 0.106). Note however that both main effects of emotion and musicianship are mediated by a three-way interaction between emotion, pattern and musicianship.

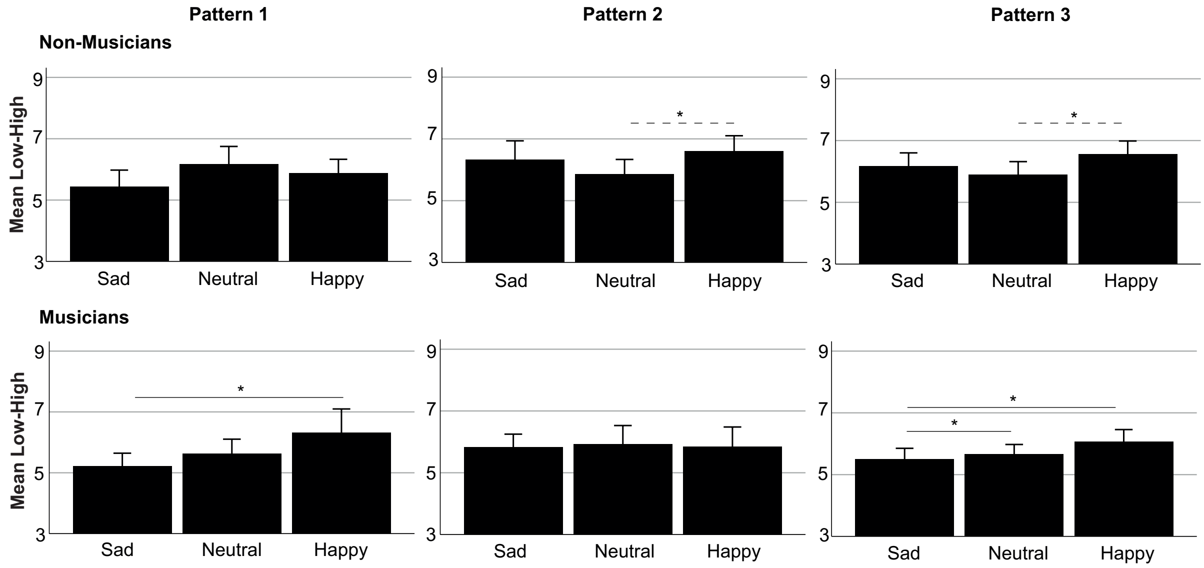

Figure 4. Means and error bars (95% confidence intervals) of indications of attention to low or high streams for each emotion, pattern and participant group of Experiment 2. Pattern 1 and 2 consist of simple alternations of a low and high pitch, a fourth or fifth apart, respectively. Pattern 3 combines stepwise motion with an alternation between high and low tones (complex pattern). Dotted lines indicate differences between conditions for which no overall effect of emotion was found. * $p < .05$

Figure 4 shows the means and confidence intervals for each emotion and pattern for non-musicians (top) and musicians (bottom). Testing the effect of emotion for each pattern and each group showed a linear relationship between emotion (coded as -1 for sad, 0 for neutral, and 1 for happy) and attention to low-high stream for musicians in Patterns 1 (simple-fourth) and 3 (complex) ($p < .05$, $eta_p^2 = .187$; $eta_p^2 = .173$, respectively), but not in

Pattern 2 (simple-fifth). For non-musicians, the effect of emotion did not reach significance for any of the patterns ($p > .06$, $.097 < eta_p^2 > .129$). Exploring pairwise comparisons between emotion conditions showed that only the differences in attention to low-high streams between the Happy and Neutral conditions reached significance for the non-musicians in Patterns 2 (simple-fifth) and 3 (complex) ($p < .05$).

3.2.2 Attention to both streams, stream integration

For stream integration vs segregation responses, the main effect of pattern was the only significant effect (Table 2). Post hoc comparisons confirmed a significant difference in integration responses between the complex pattern and each of the simple patterns ($p < .05$): The complex pattern received a relatively high integration score ($M = 1.271$, $SE = 0.085$), followed by the simple pattern a fourth apart ($M = 1.047$, $SE = 0.090$), and the simple pattern a fifth apart ($M = 0.940$, $SE = 0.080$).

3.3. Discussion

The effect of emotion on attention to high and low streams were as predicted: sad contexts promoted the focus of attention to the lower stream, while a happy context promoted the focus of attention to higher streams. For non-musicians, this effect of emotion on attention was restricted to the promotion of attention to higher streams in happy contexts, while for musicians, distinctions in attention concerned sad as well as happy contexts. These effects of emotion on attention may have had a side effect of promoting stream segregation for both the sad and happy contexts compared to the neutral contexts. This (non-significant) tendency ($p = .053$, $eta_p^2 = .093$) for a quadratic relationship between emotion and stream integration was in contrast to our prediction that happy contexts may promote stream integration through associations with higher energy. It is possible that happy contexts have both a tendency to increase stream integration and to increase stream segregation due to promoting attention to the higher stream, making it relatively hard to demonstrate an overall effect of happy contexts on stream perception experimentally.

The final experiment was designed to again test the influence of emotional context on the perception of streams, but in a way that avoids explicit mention of the words “high” and “low”, as such explicit indication may influence answers through verbal mediation. To operationalise this, one stream consisted of stepwise motion while the other stream consisted of a repeating tone.

4. Experiment 3

4.1. Method

4.1.1 Design

A factorial within-participants design was used with the independent variables of emotion (3 levels: happy, neutral, sad), pitch interval (2 levels: fourth, octave) and register of stepwise motion (2 levels: low or high stream). Dynamic attention was measured by asking participants to indicate their predominant perception of the sequence by pressing one of three buttons – attention to repeated tone, stepwise motion or pattern as a whole. Responses were recoded in two dependent variables: degree of attention to low or high stream and degree of stream segregation or integration.

4.1.2 Participants

44 adults participated in the experiment (16 M, 28 F, median age = 26.5). Participants had a range of music performance experience from no musical performance and lessons (N=7) to life-long experience. Median of music performance experience was 9 years, and 8.5 years of music lessons. Power calculations indicated that a sample size of 44 participants is sufficient to detect an effect with a modest size of $eta_p^2 = 0.105$ and power of 0.8, in the context of a repeated measures ANOVA with 3 levels.

4.1.3 Materials – emotional context

The same video material was used to vary emotional context as in Experiment 2.

4.1.4 Materials – auditory sequences

A marimba sound was used to generate musical stimuli using a MIDI sequencer. The stimuli consisted of a continuous stream of alternating high and low pitches with 150 ms between successive tones. The distance between the high and low pitched streams was either a fourth (5 semitones) or an octave (12 semitones). One of the streams contained only a repeating tone, while the other stream contained stepwise motion (as illustrated in Figure 3). This stepwise motion was present either in the higher or in the lower stream, which is referred to as register of the stepwise motion. Pitch interval and register made up the main experimental manipulations of the musical material. Additional variation in material was added by creating sequences that started with a downward or upward stepwise motion, that either started with a tone of the high stream or with a tone of the low stream, and by using eight pitch heights to present sequences to participants.

4.1.5 Procedure

Participants watched the emotional video and heard the auditory material simultaneously, which lasted for 12 seconds. They were instructed to imagine the emotion portrayed in the video, while listening to the musical sequence. At the end of the video and music, participants indicated their predominant perception of the sequence using one of three response options: integrated perception of the sequence, attentional focus on the repeating tone, or attentional focus on the stepwise motion. Three keys of the computer keyboard were used as response keys that had white stickers with small representations of the tone patterns printed on it.

A blocked procedure was used grouping trials per emotion. The order of emotion blocks was randomized across participants. Each emotion block had 16 trials – 2 intervals x 2 registers x 2 motion directions x 2 starting tones. Within each emotion block, musical stimuli were presented in a random order. Pitch height of the sequence and type of video was randomly selected as long as the video represented the correct emotion. Participants gave informed consent before participating in the study and practiced the task before starting the experimental trials.

4.1.6 Data processing

The indications of focus of attention were used to calculate two dependent variables. The first considered whether the focus of attention was in the lower or higher stream. Specifically, if participants had indicated to focus on the repeating tone or the stepwise motion and this was present in the lower stream, attention to low-high stream was coded as 0. If it was present in the higher stream, attention to low-high stream was coded as 1. Alternatively, if participants had attended to the sequence as a whole, attention to low-high stream was coded as 0.5. The second dependent variable classified the responses in terms of segregation (in case participants had indicated to focus on the repeating tone or the stepwise motion), which was coded as 0, or integration, which was coded as 1. Responses were averaged across multiple trials of the same condition. Due to averaging across several trials, the distribution of the dependent variables was close to normal, suggesting the suitability of the use of parametric tests for data analysis[6]. Descriptive statistics (mean and SE) per experimental condition and dependent variable are provided for reference in Appendix 1, Table A3.

4.2 Results

Two univariate ANOVAs were run with emotional context, interval and register as within-participant variables and the attention to low or high stream and degree of stream integration as dependent variables. The main statistical results are presented in Table 3.

[6] The effects of emotion on stream perception reported in the results section are similar irrespective of whether parametric or non-parametric tests are conducted. For simplicity only parametric tests are reported.

Table 3. Overview of statistical results of Experiment 3

			Attention to low-high stream			Degree of stream integration		
Effect	df model	df error	*F*	*p*	eta_p^2	*F*	*p*	eta_p^2
Emotion	2	86	1.949	.149	.043	4.439	.015	.094
Interval	1	43	28.947	<.001	.402	52.121	<.001	.548
Register	1	43	25.963	<.001	.376	20.352	<.001	.321
Emotion * Interval	2	86	0.088	.915	.002	0.331	.791	.008
Emotion * Register	2	86	0.201	.818	.005	0.480	.620	.011
Interval * Register	1	43	1.286	.263	.029	0.123	.727	.036
Emotion * Interval * Register	2	86	1.817	.169	.041	1.590	.210	.001

4.2.1. Attention to the lower or higher stream

For attention to the lower stream, both, or higher stream indications, the main effects of interval and register were highly significant, but not the effect of emotion, nor any of the interactions (see Table 3). The effect of interval was related to stronger attention to the higher stream for the larger interval of an octave (*M* = .619, *SE* = .027) than the smaller interval of a fourth (*M* = 497, *SE* = .015). The effect of register was as can be expected – more attention to the higher stream when the stepwise motion is in the higher register (*M* = .629, *SE* = .024) than when it is in the lower register (*M* = .487, *SE* = .022).

4.2.2. Stream segregation or integration

For stream segregation or integration indications, all three main effects were significant (Table 3), while none of the interactions between effects were significant. For the main effect of emotion, planned contrasts confirmed a linear relationship between emotion (coded as -1 for sad, 0 for neutral, and 1 for happy) and stream integration (p = .018) with a higher degree of integration for happy than neutral and sad (see Figure 2, right panel). Pairwise comparisons confirmed a significant difference in attention scores between the happy and sad contexts (p = .018) and between the neutral and sad contexts (p = .014).

The main effect of interval was also as expected with more stream integration for the smaller interval of a fourth (M = .439, SE = .035) than the larger interval of an octave (M = .354, SE = .028). The main effect of register was such that conditions with stepwise motion in the lower stream had relatively more stream integration than conditions with stepwise movement in the higher stream (M = .260, SE = .024).

5. General discussion and conclusion

The three experiments provided important first evidence of the role and influence of emotional context on auditory stream segregation in a musical setting. Experiments 1 and 2 showed that (musically-trained) listeners adapted attention to the higher or lower stream in accordance with the emotional context, while Experiment 3 confirmed an effect of emotional context on the degree of stream integration vs. segregation. All significant effects were in the expected direction: relatively more attention was allocated to the lower stream in sad contexts, while, in happy contexts, relatively more attention was allocated to the higher stream. Furthermore, stream segregation was particularly strong for the sad condition in Experiment 3, congruent with a slow pace and limited melodic activity association with sadness. Experiment 2 showed that listeners without extensive musical training may show an effect of emotion on attention, but in a more restricted manner than musically trained participants, as differences were only observed for the happy contexts and overall effects fell short of being significant. This may possibly be due to a limited level of attention to a lower pitch range in non-musicians (as also discussed below).

Not all of our predictions were confirmed. Specifically, it was found to be difficult to demonstrate that happy contexts promote stream integration. Indeed, this may not be the case, if happy contexts promote attention to the higher stream, as that increases stream segregation. In future studies, the effect of emotional or semantic context on stream integration may be better tested using associations that specifically relate to activation and less to high pitch, such as presenting a “lively” or “active” context (Timmers, Schiavio Cowell, 2015).

Furthermore, the effect of emotion on attention was not significant in Experiment 3, while it was significant for the first two experiments. This could be related to the influence of the location of the stepwise movement on attention to streams, which draws attention, but also to the large pitch interval that was used in this experiment (an octave), which may have inadvertently caused differences in salience of the tones (due to e.g. register or perceptual intensity of the tones). Follow-up tests run for exploratory purposes highlighted that the effect of emotion was significant (p = .044) for only one of the four musical conditions, namely in the context of the smaller interval and stepwise motion in the lower stream.

Observed sizes for the effects of emotion on attention and stream segregation were generally small and not always significant. This may have been related to some limitations in the experimental designs that asked participants to integrate quite unrelated auditory and visual material, namely static images of emotional faces and simple auditory sequences. The influence of emotion is likely to be stronger in stimuli where emotional context and music are more closely associated, such as in the context of film music, musical theatre or influences of visual emotional expression of a performer.

Even though the experiments employed traditional cognitive paradigms of testing the effect of variables on perception, the results across experiments are illustrative of the dynamic character of music perception. Subtle variations in musical material afforded listeners different ways of engagement: in Experiment 1, both musical material and visual emotional context were interpreted as emotional. In Experiment 2, the complex pattern afforded stronger stream segregation, possibly through adding a greater sense of activation to the musical material. In Experiment 3, stream segregation was promoted by the sad context, possibly particularly because of a relatively high activity of the pattern as a whole including a large register difference between the two stream and melodic movement in one of the streams. What these results demonstrate is the relevance of musical character, emotion and meaning for the structural perception of music. This parallels findings in language perception of the influence of semantics on syntax perception (e.g. Britt, 1994; Hoeks, Stowe, & Doedens, 2004), but in a domain for which semantic meaning is highly debated (Koelsch et al., 2004; Cross, 2011). While music may not strictly show "semantics" in the way language does (e.g. Reich, 2011), our results do indicate the need for adaptation of models of music perception, that primarily consider syntax and organisation of sound (e.g. Huron, 2006; Krumhansl, 2001; Pearce & Wiggins, 2012).

Experiment 2 further showed an effect of musical training on perception: musically trained participants showed stronger influences of emotional context on stream perception than non-musicians, and showed relatively greater attention to the lower stream. These differences between groups illustrate the relevance of participants' music engagement histories for the ways in which they perceive music. It will be of interest to explore further what about these histories may have contributed to the observed differences in perception and what forms of musical training and engagement are most influential – performance histories or other forms of "musical sophistication" (Müllensiefen, Gingras, Musil, & Stewart, 2014).

For example, musicians may learn to pay relatively more attention to the bass and the harmony, in addition to listening to the higher pitched melody (Loui & Wessel, 2007). Furthermore, musicians may be emotionally involved with music, even if music is as simple as the sequences used in these experiments, and may find it more intuitive to connect visual emotional context with these musical patterns. Through their performance experiences, they may also find it easier to playfully interact with the musical material in perception (Przysinda, Zeng, Maves, Arkin, & Loui, 2017). Whilst these examples relate to the potential influence of performance experience, other forms of musical development such as frequent and intense listening to music for emotional purposes may be similarly powerful in shaping music perception (Honing & Ladinig, 2009). A limitation of the presented experiments was that only one of them tested participants with limited or no formal musical training, and that musical training was only examined in a binary manner (with/without), not allowing for the examination of the extent to which various forms of experience may contribute to variations in perception.

Recent research in music has seen a rise in embodied and enactive perspectives on music cognition, many of which provide theoretical accounts or specifically focus on interrelationships between music performance and perception (e.g. Reybrouk 2012; 2014; Leman, 2008; Schiavio, van der Schyff, Cespedes-Guevara, & Reybrouk, 2017). Our experiments showcase an important aspect of a dynamic relationship between listeners and music: meaning may arise through meaningful interaction with music and this meaningful interaction shapes our perception of music. It provides crucial empirical evidence for claims that perception cannot be purely understood by investigating "cold mechanisms" (as argued in e.g. Schiavio et al., 2017).

To conclude, even though the three experiments of this study employed a classical experimental paradigm, they provided clear insight and evidence of the dynamic nature of music perception where emotional meaning and music engagement histories shape perception. The study highlights the need for the revision of dominant syntax-driven models of music perception, including (emotionally driven) semantics as part of our understanding of music cognition.

Acknowledgments

This study was realised with financial support from the British Academy to RT (SG-54560). We would like to thank Sally Barker, Hannah Beezer, Hannah Dunn, Ayane Fujiwara, M. Isabel Gutierrez Blasco, Xin Link, Joanne Mounsey, Margarida Olive Masdeu, and Dashiell Oatman-Stanford for help with data collection for Experiment 2 and Nicola Dibben and Eduardo Coutinho for their feedback on the studies and the manuscript.

Author contributions

RT – conceptualisation, method, data analysis, write up, YA – method, data collection, data analysis, contributions to write up and conceptualisation, HC – contributions to conceptualisation, literature review, and write up.

References

Barsalou, L. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes*, *18*(5-6), 513-562.

Beauvois, M. W., & Meddis, R. (1996). Computer simulation of auditory stream segregation in alternating-tone sequences. *The Journal of the Acoustical Society of America*, *99*(4), 2270-2280.

Bendixen, A., Denham, S. L., Gyimesi, K., & Winkler, I. (2010). Regular patterns stabilize auditory streams. *The Journal of the Acoustical Society of America*, *128*(6), 3658-3666.

Bendixen, A., Denham, S. L., & Winkler, I. (2014). Feature predictability flexibly supports auditory stream segregation or integration. *Acta Acustica united with Acustica*, *100*(5), 888-899.

Bigand, E., McAdams, S., & Forêt, S. (2000). Divided attention in music. *International Journal of Psychology*, *35*(6), 270-278.

Bishop, J. M., & Martin, A. O. (2014). Contemporary sensorimotor theory: a brief introduction. In J. M. Bishop & A. O. Martin (Eds.), *Contemporary Sensorimotor Theory. Studies in Applied Philosophy, Epistemology and Rational Ethics* (Vol. 15, pp. 1–22). Cham: Springer.

Boltz, M. G. (2001). Musical soundtracks as a schematic influence on the cognitive processing of filmed events. *Music Perception: An Interdisciplinary Journal, 18,* 427-454.

Boltz, M. G., Ebendorf, B., & Field, B. (2009). Audiovisual interactions: The impact of visual information on music perception and memory. *Music Perception: An Interdisciplinary Journal, 27,* 43–59.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA.: MIT press.

Britt, M. A. (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language*, *33*(2), 251-283.

Carlyon, R. P., Plack, C. J., Fantini, D. A., & Cusack, R. (2003). Crossmodal and non-sensory influences on auditory streaming. *Perception, 32,* 1393–1402.

Chemero, A. (2003). An Outline of a Theory of Affordances. *Ecological Psychology*, *15*(2), 181–195.

Clarke, E. F. (2005). *Ways of listening: An ecological approach to the perception of musical meaning*. Oxford University Press.

Cross, I. (2011). The meanings of musical meanings: Comment on “Towards a neural basis of processing musical semantics” by Stefan Koelsch. *Physics of Life Reviews*, *8*(2), 116-119.

Davis, S. (2011). Stream segregation and perceived syncopation: Analyzing the rhythmic effects of implied polyphony in Bach’s unaccompanied string works. *Music Theory Online*, *17*(1), mto.11.17.1.davis

Denham, S.L. and Winkler, I. (2018), Predictive coding in auditory perception: challenges and unresolved questions. *European Journal of Neuroscience.* doi:10.1111/ejn.13802

Desain, P. (1992). A (de) composable theory of rhythm perception. *Music Perception: An Interdisciplinary Journal*, *9*(4), 439-454.

Devergie, A., Grimault, N., Tillmann, B., & Berthommier, F. (2010). Effect of rhythmic attention on the segregation of interleaved melodies. *The Journal of the Acoustical Society of America*, *128*(1), EL1-EL7.

Drake, C., Jones, M. R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition*, *77*(3), 251-288.

Gabrielsson, A., & Lindström, E. (2010). The role of structure in the musical expression of emotions. In P. N. Juslin, & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 367–400). Oxford, UK: Oxford University Press.

Gagnon, L., & Peretz, I. (2003). Mode and tempo relative contributions to “happy-sad” judgements in equitone melodies. *Cognition & Emotion, 17*(1), 25–40.

Hansen, N. C., & Pearce, M. T. (2014). Predictive uncertainty in auditory sequence processing. *Frontiers in Psychology*, *5*, 1052. doi.org/10.3389/fpsyg.2014.01052

Hartmann, W. M., & Johnson, D. (1991). Stream segregation and peripheral channeling. *Music Perception: An Interdisciplinary Journal, 9,* 155–183.

Heilbron, M., & Chait, M. (2018). Great expectations: is there evidence for predictive coding in auditory cortex?. *Neuroscience*, *389*, 54-73.

Hoeks, J. C., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: the interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, *19*(1), 59-73.

Honing, H., & Ladinig, O. (2009). Exposure influences expressive timing judgments in music. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(1), 281-288.

Houston, D., & Haddock, G. (2007). On auditing auditory information: The influence of mood on memory for music. Psychology of Music, 35, 201–212.

Huron, D. B. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MA: MIT press.

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, *96*(3), 459-491.

Jones, M. R., Jagacinski, R. J., Yee, W., Floyd, R. L., & Klapp, S. T. (1995). Tests of attentional flexibility in listening to polyrhythmic patterns. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(2), 293-307.

Juslin, P. N., Liljeström, S., Västfjäll, D., Barradas, G., & Silva, A. (2008). An experience sampling study of emotional reactions to music: listener, music, and situation. *Emotion*, *8*(5), 668-683.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nature Neuroscience*, *7*(3), 302-307.

Krueger, J. W. (2011). Doing things with music. *Phenomenology and the Cognitive Sciences*, *10*(1), 1–22.

Krueger, J. W. (2014). Affordances and the musically extended mind. *Frontiers in Psychology*, *4*, 1003. https://doi.org/10.3389/fpsyg.2013.01003

Krumhansl, C. L. (2001). *Cognitive foundations of musical pitch*. Oxford, New York: Oxford University Press.

Lang, P. J. (1995). The emotion probe: Studies of motivation and attention. *American Psychologist, 50,* 372-385.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion, 24,* 1377-1388.

Leman, M. (2008). *Embodied music cognition and mediation technology*. MIT press.

London, J. (2012). *Hearing in time: Psychological aspects of musical meter*. Oxford University Press.

Marin, M. M., Gingras, B., & Bhattacharya, J. (2012). Crossmodal transfer of arousal, but not pleasantness, from the musical to the visual domain. *Emotion*, *12*(3), 618-631.

Micheyl, C., Hanson, C., Oxenham, A., Demancy, L., & Shamma, S. (2013). Auditory stream segregation for alternating and synchronous tones, *Journal of Experimental Psychology: Human Perception and Performance, 39*(6), 1568–1580.

Moore, B. C. J., Gockel, Hedwig, (2002) Factors influencing sequential stream segregation, *Acta Acoustica United with Acustica, 88*(3), 320-333.

Moore, B. C., & Gockel, H. E. (2012). Properties of auditory stream formation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1591), 919-931.

Myin, E. (2016). Perception as something we do. *Journal of Consciousness Studies*, *23*(5–6), 80–104.

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE, 9*(2), e89642.

Niedenthal, P. M., Halberstadt, J. B., & Innes-Ker, Å. H. (1999). Emotional response categorization. *Psychological Review*, *106*(2), 337-361.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, *24*(5), 939–1031.

Park, J. M., Chung, C. K., Kim, J. S., Lee, K. M., Seol, J., & Yi, S. W. (2018). Musical expectations enhance auditory cortical processing in musicians: A magnetoencephalography study. *Neuroscience*, *369*, 325-335.

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception: An Interdisciplinary Journal*, *11*(4), 409-464.

Pearce, M. T., & Wiggins, G. A. (2012). Auditory expectation: the information dynamics of music perception and cognition. *Topics in Cognitive Science*, *4*(4), 625-652.

Przysinda, E., Zeng, T., Maves, K., Arkin, C., & Loui, P. (2017). Jazz musicians reveal role of expectancy in human creativity. *Brain and Cognition*, *119*, 45-53.

Randall, W. M., Rickard, N. S., & Vella-Brodrick, D. A. (2014). Emotional outcomes of regulation strategies used during personal music listening: A mobile experience sampling study. *Musicae Scientiae*, *18*(3), 275-291.

Reich, U. (2011). The meanings of semantics: Comment on "Towards a neural basis of processing musical semantics" by Stefan Koelsch. *Physics of Life Reviews*, *8*(2), 120-121.

Reybrouck, M. (2012). Musical sense-making and the concept of affordance: An ecosemiotic and experiential approach. *Biosemiotics*, *5*(3), 391–409.

Reybrouck, M. (2014). Music as environment: an ecological and biosemiotic approach. *Behavioral Sciences*, *5*(1), 1–26.

Rogers, W. L., & Bregman, A. S. (1993). An experimental evaluation of three theories of auditory stream segregation. *Perception & psychophysics*, *53*(2), 179-189.

Schiavio, A., & van der Schyff, D. (2016). Beyond musical qualia. Reflecting on the concept of experience. *Psychomusicology: Music, Mind, and Brain*, *26*(4), 366–378.

Schiavio, A., van der Schyff, D., Cespedes-Guevara, J., & Reybrouck, M. (2017). Enacting musical emotions. Sense-making, dynamic systems, and the embodied mind. *Phenomenology and the Cognitive Sciences*, *16*(5), 785-809.

Sloboda, J., & Edworthy, J. (1981). Attending to two melodies at once: the of key relatedness. *Psychology of Music*, *9*(1), 39-43.

Tan, S. L. (2017). Scene and heard: The role of music in shaping interpretations of film. In R. Ashley & R. Timmers (Eds.), *The Routledge companion to music cognition* (pp. 363-376). New York: Routledge.

Teki, S., Chait, M., Kumar, S., von Kriegstein, K., & Griffiths, T. D. (2011). Brain bases for auditory stimulus-driven figure–ground segregation. *Journal of Neuroscience*, *31*(1), 164-171.

Timmers, R., & Crook, H. (2014). Affective priming in music listening: Emotions as a source of musical expectation. *Music Perception: An Interdisciplinary Journal*, *31*(5), 470-484.

Timmers, R., Schiavio, A., & Cowell, H. (2015). Influences of visual activity on memorised tempo of a performance. In J. Ginsborg, A. Lamont, M. Phillips, and S. Bramley (Eds.), *Proceedings of the Ninth Triennial Conference of the European Society for the Cognitive Sciences of Music (ESCOM)* (p. 782). Manchester, UK: Royal Northern College of Music, 17-22 August.

Thompson, E. (2007). *Mind in life*. Cambridge, MA: Harvard University Press.

Van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation, Eindhoven University of Technology, Eindhoven, the Netherlands.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Dalca, I. M., & Levitin, D. J. (2011). Music to my eyes: Cross-modal interactions in the perception of emotions in musical performance. *Cognition*, *118*(2), 157-170.

Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion*, *8*(4), 494-521.

Appendix 1.

Table A1. Mean and standard errors (SE) for each experimental condition and dependent variable of Experiment 1. Conditions varied in interval size between the low and high note (a fifth or augmented fifth apart (5 or A5, respectively) and tempo.

		Emotion	Sad		Neutral		Happy	
Rating	Pitch interval	Time interval (ms)	Mean	SE	Mean	SE	Mean	SE
Low-High	5	120	4.324	0.194	4.286	0.181	4.457	0.235
		150	3.81	0.267	4	0.209	3.771	0.224
		180	3.076	0.241	3.362	0.252	3.219	0.274
	A5	120	4.771	0.186	4.59	0.215	4.79	0.219
		150	4.114	0.148	3.838	0.255	4.229	0.227
		180	3.781	0.224	3.619	0.239	3.552	0.232
Integration	5	120	3.314	0.162	3.305	0.186	3.552	0.216
		150	2.952	0.158	3.267	0.192	3.524	0.197
		180	3.276	0.153	3.524	0.185	3.59	0.147
	A5	120	3.238	0.203	3.105	0.179	3.571	0.253
		150	3.029	0.178	3.381	0.152	3.39	0.199
		180	3.257	0.137	3.181	0.149	3.476	0.188

Table A2. Mean and standard errors (SE) for each experimental condition and dependent variable of Experiment 2. Pattern refers to a simple alternation of a low and high pitch a fourth or fifth apart (1 and 2, respectively) or a more complex pattern that combines a stepwise motion in each voice with an alternation of a low and high pitch.

		Emotion	Sad		Neutral		Happy	
Group	Rating	Pattern	Mean	SE	Mean	SE	Mean	SE
Non-musicians	Low-High	1	5.429	0.263	6.167	0.279	5.881	0.215
		2	6.333	0.289	5.857	0.229	6.595	0.243
		3	6.167	0.208	5.893	0.204	6.56	0.204
	Integration	1	0.952	0.169	1.024	0.184	0.833	0.139
		2	0.952	0.158	0.762	0.149	0.929	0.177
		3	1.095	0.136	1.083	0.117	1.107	0.115
Musicians	Low-High	1	5.225	0.200	5.625	0.229	6.325	0.369
		2	5.825	0.203	5.925	0.289	5.850	0.302
		3	5.500	0.168	5.663	0.150	6.075	0.183
	Integration	1	1.175	0.167	1.275	0.172	1.025	0.152
		2	0.725	0.143	1.275	0.156	1.000	0.191
		3	1.300	0.153	1.538	0.180	1.500	0.148

Table A3. Descriptive statistics for Experiment 3: Mean and standard errors (SE) for each experimental condition and dependent variable. Pitch interval between low and high pitch was a fourth or an octave. Step refers to the voice that contained the stepwise motion, which could be low (0) or high (1).

		Emotion	Sad		Neutral		Happy	
Rating	Interval	Step	Mean	SE	Mean	SE	Mean	SE
Low-High	4	0	0.393	0.025	0.389	0.031	0.463	0.027
	4	1	0.609	0.031	0.547	0.031	0.582	0.038
	8	0	0.579	0.04	0.532	0.042	0.568	0.041
	8	1	0.664	0.043	0.666	0.041	0.707	0.039
Integration	4	0	0.402	0.042	0.528	0.054	0.517	0.048
	4	1	0.379	0.046	0.390	0.047	0.415	0.049
	8	0	0.206	0.042	0.223	0.040	0.250	0.041
	8	1	0.092	0.022	0.152	0.035	0.131	0.031