The Perception of Time in Music

Sandra Quinn

Department of Psychology
University of Stirling

Submitted for the degree of Doctorate of Philosophy
Submitted 30/09/05

THESIS CONTAINS CD/DVD

ACKNOWLEDGEMENTS

There are so many people to thank that I don't even know if I will ever have enough room on this page to get you all in. Firstly I would like to thank one's supervisor (Roger) for his boundless enthusiasm and encouragement in everything I have achieved over the last 3 years. I am indebted to him for putting a bet on me and taking a chance. Secondly I'd like to thank the department for giving me a departmental studentship to enable me to do this work. Many thanks to the technicians for fixing the endless issues I had with PCs. The support of the patient administrative staff here in the department of psychology is boundless, and they made sure I got paid, had paper and indeed kept sane.

Next I would like to thank all of my friends here in Stirling whose support ensured that I remained rational and also brought a great deal of happiness to my life here in Stirling. In alphabetical order I should like to acknowledge, Anna, Binder, Cindy, Debbie, Fhionna, Hayley, Helen, Kay, Lesley, Robbie, Sinead, Sue, and Susan. Had it not been for you guys I would have lost the plot a long time ago!

I now turn to the Dundee crew who have been there for me since I started at University. I'd like to thank in alphabetical order, Angela, Cat, Dawn, Denise, Karen, Paula, and Tracey. You kept me smiling when all was lost and made sure I had a few drinks to keep me that way.

Finally, my last thanks go to my Mum and Dad who have supported me for over 2 decades in my quest in music. Had they not pushed me and told me I could be anything that I wanted to be, I don't think I could have ever applied to do any course at Uni. I love you both. Also many thanks go to my brother Brian, Victoria and Paige. You have brought many happy times to my life. Thank you.

I dedicate this thesis to Brian Quinn senior, Catherine Quinn, Brian Quinn Jnr., Victoria, and Paige. I love you all.

ABSTRACT

This thesis is concerned with the perception of time in music with emphasis on tempo, emotion and time perception in music.

Three studies were conducted to assess whether listeners were able to make consistent judgements about tempo that varied from piece to piece. Listeners heard short extracts of Scottish music played at a range of tempi and were asked to make a two alternative forced choice of 'too fast' or 'too slow' for each extract. The responses for each study were plotted as proportion too fast responses as a function of tempo for each piece, and cumulative normal curves were fitted to each data set. The point where these curves cross 0.5 is the tempo at which the music sounds right to the listeners, referred to as the optimal tempo. The results from each study show that listeners are capable of making consistent tempo judgements and that the optimal tempo varies across extracts. The results also revealed that rhythm plays a role, but not the only role in making temporal judgements.

In the previous studies, it is possible that listeners might be using an average tempo from previously heard extracts to make every subsequent response. We wanted to assess this by presenting a single stimulus per participant and therefore remove any effects of the context on participant's responses. Using this technique we shall show that listeners can make 'too fast' and 'too slow' responses that are independent of previously heard extracts. In addition the data reveal similar results to those found in the first experimental chapter.

The 3rd chapter deals with the effect of changes in the tempo of music on the perception of happy and sadness. Listeners heard short extracts of music that varied in tempo and were asked to make a 2AFC of happy or sad for each extract. Separate psychometric functions were obtained for each extract of music, and the points where these crossed 83% and 17% happy were calculated, and treated as happy tempo and sad tempo respectively. The results show that most extracts can be perceived as both happy and sad just by varying the tempo. However, the tempo at which extracts become happy or sad varies widely from extract to extract. We show that the sad and happy tempi are related to the size of the intervals (pitch changes) in the extract.

In considering what might be involved in the perception of time in music we wanted to assess what effect small changes to a stimulus would have on perceived duration. We presented 2 auditory stimuli and show that the perceived duration of the test stimulus with a change in pitch increased as the size of the pitch change increased. The results are explained in terms of event strength where strong events cause perceived duration to increase whilst weak events are perceived to be shorter by comparison.

CONTENTS

CHAPTER 1: PRIMAL SKETCH OF MUSIC

The perceptual processing of music begins when auditory sequences made up of tones are presented to us. Three main processes are involved in the perception of these sequences that entail what information is encoded, how it is represented and the way it is interpreted. Encoding involves the transferral of information from the stimulus outside in the world to the individual through the sensory system. The process of interpretation follows on from the encoding stage and produces a number of attributes about the information being processed that are based upon our knowledge. Finally, the representational stage provides a description of the stimulus by gathering together the information from the previous 2 stages. At this point all of the possible information about the stimulus is collected and can be used to make a response.

One model that has been influential in visual perception is Marr's model. This model contains 2 relevant sub-systems: the raw primal sketch (the description of the primitive features of the world) and the full primal sketch (the process of integrating those features into perceptual structures). Both of these systems involve all 3 types of processing mentioned above (encoding, interpretation and representation). This approach is useful because it breaks down perception into understandable stages. Before going any further it is important to appreciate that being in the mere presence of a stimulus does not necessarily lead to perceptual processing. Marr did not consider or recognise this, but it is vital to acknowledge that perceptual processing can only come about if an individual actively engages in their environment.

Let's now consider what the primal sketch for music involves. The raw primal sketch in music provides a description of the basic features of the notes in a piece of music and could contain information regarding the duration, pitch, and intensity of each note. The outcome of this process is due to an enormous amount of sensory processing, but at this stage only produces a note-by-note representation. Initially, the raw primal sketch encodes the stimulus in terms of its acoustic waveforms and then interprets them as discrete notes. These notes are processed so that their onset and offset times as well as their duration and intensity are represented in some way. This last part of the process also provides the order in which the notes are presented.

It is only when the characteristics of notes are pulled together into larger groupings that their musical significance can be established. This is initially made possible by encoding the representation produced from the raw primal sketch. The information gathered here is processed further by interpreting the raw primal sketch of each note and converting it into a musical unit such that: note durations become multiples of an underlying beat; note pitches become fixed on a semitone step; and note intensities become forte or piano etc. These notes are then pulled together into musical groupings where note durations collectively become rhythmic patterns, note pitches collectively become a melody, and note intensities collectively become dynamics. The representation stage of the full primal sketch creates a representation that is similar to a piano score.

A final musical representation follows these 2 perceptual processes and can include a large range of possible musical conceptualisations. The listener might produce an optimal tempo, or emotional representation that is the result of the assimilated features contained within the full primal sketch. In order to do so the listener has to pull upon previously learned musical knowledge and decide whether this is comparable with the information gathered within the full primal sketch. Once a match has been obtained the listener can make a response to the music they have heard and decide for example whether it is too slow or happy. The musical representation is the final stage in the perceptual process.

Each stage in the perceptual process may not be necessary and will depend upon what the listener is trying to achieve and the demands of the task. For example, emotional responses to a piece of music would require all 3 stages whilst tasks involving recognising a pitch in a piece of music might only require a raw primal sketch.

As it is not possible to process all of the incoming information we are presented with it is likely that we might filter out some of the noise in order to focus on what is most important. It is possible that we decide early on what information we should deal with and what to ignore. In arguing this I suggest that being in the mere presence of a stimulus does not mean that actual perceptual processing will occur. By thinking of it in this way where we choose what to engage with we can begin to see how perception is not a passive process, but an active one.

Pieces of music contain elements that are more important than others and engage the listener more than others. For example, some notes are more significant than other notes in that they may be louder or be a central pitch in a sequence or simply last longer than other notes that surround them. If notes contain properties that cause them to be more or less influential then we would expect them to have a fundamental effect on our experience of music. Where some notes are more important it would be expected that these would engage the listener more than other notes and have a powerful role to play in shaping the response a listener makes. In this way the perceptual process becomes active because it can only come about if the listener is prepared to recognise the properties of notes that are important in the first place.

The amount of influence notes have is determined by the properties they contain and the context in which they are placed. I argue that notes have varying degrees of strength that are determined by their physical properties and by their relationship with the notes that surround them. The amount of strength a piece of music contains will determine the degree to which it is happy rather than sad or how fast or slow it ought to be played. It is therefore the purpose of this thesis to begin to show how the strength of musical events shape the perceptual processing that listeners experience.

THE ROLE OF TIMING IN MUSIC

Timing is manipulated by musicians in a number of different ways. Musicians use both a metronome marking (Maelzeal Metronome) or a beats per minute (bpm) designation written onto the musical score to guide their overall speed. However, departures from these numerical values involve micro deviations in timing from note to note, or more detectable timing variations such as slowing down at the end of a piece of music (ritardando). The latter of these and other related expressions are dictated by the musical score through the use of Italian, French and German terms. These temporal features are classified as global or local tempo. Global tempo refers to the overall speed of a piece of music (the numerical value) whereas local tempo refers to small departures used in expressive performances (i.e. ritardando).

Musicians are very skilled at using these timing features and can converge upon a single global tempo during group performances where they stream together as a cohesive whole. Without a global tempo to guide them the musicians would lack synchronicity and any performance they give would not be interpretable. Global tempo therefore ensures that multiple musical parts are held together in a way that produces a successful performance and allows the listeners to comprehend what it is expressing. This is not to say that the individual musician's performances do not employ expressive timings, but their overall success also requires a global tempo to be used.

Some researchers have suggested that musicians have listening strategies that are different from non-musicians (Smith, 1987) where they can detect deviations away from global tempi because of their exposure to playing and listening to music (Repp, 1992a). However, musicians often feel that their musical skills are being tested when they take part in a music study. Although this seems obvious this can often affect the responses they make in that any results may be due to their desire to show the experimenter how skilled they are, rather than provide a response that is uncontaminated by the presence of the experimenter. Using non-musicians removes this problem and allows the experimenter the opportunity to assess perceptual phenomena of the population much more effectively.

Later evidence suggested that the ability to detect changes in tempo was due to musical structure rather than musical ability (Repp, 1999a). Although non-musicians are not able to provide the Italian, French and German terms used by musicians to define the temporal effects employed, this does not mean they are not capable of perceiving them. The sheer volume of music that listeners are presented with on a daily basis allows them to become well versed in the nuances of their musical culture. As such the general listening population should be capable of detecting changes in tempo in different pieces of music.

Much of the research in the perception of tempo has been focused on the expressive timings found in music performance (Timmers, Ashley, Desain and Heijink, 2000; Repp, Windsor, and Desain 2002; Repp 1998). Some studies have focused on the effects of pedal timing and showed that these vary as a result of expressive variations in piano performances (Repp, 1997a). Other studies have investigated the role of synchronisation with expressive performances of classical music through finger tapping (Repp, 1999a, 1999b, 1999c, 1997b). The preferences for particular expressive timings have also been assessed and show that musicians' preferred patterns are similar, but vary for non-musicians (Repp, 1995, 1994, 1992a, 1990). However, this may be due to the type of music used in that musically trained participants should know the pieces and genre of music well enough to predict a similar pattern of expressive timing variations. Studies using unfamiliar pieces of music might have shown more variation for the musically trained listeners. More specifically studies have assessed the types of expressive timings used by performances by concert pianists and graduate students (Repp, 1997c, 1992b).

Although this is an interesting area of research it seems that global tempo has been paid little attention. Very few studies have assessed whether listeners are capable of determining the right speed for a piece of music (the chapter that follows shall deal with this in more detail and provides a review of the literature in the area of global tempo). Since expressive timings are based upon deviations away from global tempo it would seem that this would be an important issue to deal with. If a piece of music is played at a tempo that is too slow for a piece this would affect any subsequent temporal deviations that performers use to heighten its expressiveness. In addition, if the optimal tempo for a piece of music varies from piece to piece then surely there must be some structural feature that is responsible.

THE PERCEPTION OF EMOTION IN MUSIC

Research into emotion in music has provided a wealth of literature that has attempted to pin down those musical features that allow one piece of music to be happy whilst another is perceived to be sad. Explanations for the way this is achieved and what musical features drive a listener to perceive it in such a way have still not been found. Most research has been disappointing in that it simply lists the features that are involved rather than showing the dynamic relationship between them and the emotions being perceived. Part of the problem stems from the methods used in that these often do little more than allow the researcher to describe effects rather than show how they come about. Before going on to look at these studies we shall briefly introduce the methods they use.

There are several techniques that a researcher can use to investigate emotions in music. Two types of emotional response can be measured namely an affective or perceptual response. Affective responses occur when an individual's internal state is altered having listened to a piece of music and are measured through physiological changes and self-reports. In order to substantiate whether the listener has had an affective response to a piece of music their heart rate or galvanic skin response is measured. It is believed that changes to these physiological states indicate a level of arousal which has been linked with emotion in music (Rickard, 2004). Although this is an interesting area of research it is not the purpose of this thesis to deal with the issue of affective responses to music. In mentioning it we simply want to acknowledge the difference between this and a perceptual response.

Perceived emotion differs from this and involves the listener's recognition of the emotion a piece of music is expressing without having to consider what they feel. A number of techniques have been used to assess this and include rating scales (Gagnon and Peretz, 2003; Scherer and Oshinsky, 1977; Gabrielsson and Juslin, 1996; Behrens and Green, 1993; Dalla Bella, Peretz, Rousseau and Gosselin, 2001), and adjective checklists (Hevner, 1936, 1935; Rigg, 1937; Gundlach, 1935; Farnsworth, 1954; Hampton, 1945). These are very simple and easy ways of obtaining data, but they are often riddled with problems due to the methods employed that we shall deal with during the course of this section.

Early attempts to assess the emotions perceived in music used adjective checklists. The most widely known of these used 67 adjectives arranged into 8 clusters where the words in each cluster were thought to contain an almost identical character (Hevner, 1936, 1935). Three studies were conducted to assess listener's emotional descriptions of pieces of music. The first study investigated the ability to choose adjectives and use the checklist effectively (Hevner, 1936). Having listened to 5 pieces of classical music the listeners were required to check as many of the adjectives as they liked. The results suggested that listeners were capable of choosing adjectives from the list and from every cluster. The author also suggests that the use of different adjectives could account for the listeners contrasting judgements for the same piece of music.

A second study assessed this issue further by asking participants to check the adjectives as soon as they heard a section of the music that suggested an emotion to them. The experimenter indicated to the participants which section of the music they were hearing so that they could indicate on the checklist those adjectives that they would choose to describe that section of the music. The results showed that listeners made similar selections overall to the first study, but these varied across each musical section. The second study suggests that listeners were responding to variations in the musical structure of the pieces because they chose a variety of different descriptors for the same piece of music.

A final study was conducted to establish whether this was the case by manipulating features of the music. Two different versions of the same pieces of music were presented to 2 groups who were asked to select adjectives from the checklist as they had in study 1. Each version was either manipulated in terms of its rhythm, melodic contour, harmony or mode. Thirty-one pieces of music were chosen and were re-written so that 9 contained an ascending or descending structure, 6 were written with a dissonant or consonant structure and a further 6 used arpeggios or chords. Ten additional pieces were re-written in either the major or minor mode to create a total of 62 variations of the original 31 pieces. The experimental design was set up so that the listeners would only hear one version of each piece with one manipulation.

The results suggested that the major mode is highly associated with positive emotions such as happiness, gaiety, playfulness, and sprightliness (sic) whereas the minor mode is most often associated with sadness. Consonant harmonies were most likely to involve happy, graceful, serene and lyrical selections whilst dissonant harmonies included exciting, agitated, vigorous and to a lesser extent sad choices. Ascending melodies were only partially linked to descriptions such as dignified and serene whilst descending melodies were associated to some degree with exciting, graceful and vigorous adjectives. Pieces of music where the rhythm was made up of chords were most likely to be described as dignified, vigorous and to a much lesser degree exciting. Those pieces with a rhythm containing arpeggios were described as happy, graceful, dreamy, and to a lesser degree serene. However, Hevner suggests that the only musical feature with a clear emotional character was the mode of the music that was specifically associated with happy or sad responses.

Other studies have assessed perceived emotion using smaller checklists to try and avoid the listeners choosing adjectives with similar meanings. Rigg (1937) used both a checklist and self-report method to test the assumption that particular musical phrases would cause listeners to perceive 4 emotions including joy, lamentation, longing, and love. He found that most listeners chose descriptors that were the same or similar to those highlighted by a musicologist, but there was more agreement for pieces of music that were suggested to display joy than the other 3 emotions. This might be due to the use of these words in general to describe music. However, as this was conducted nearly 80 years ago it seems unlikely that lamentation and longing would be used by most people today to describe the emotional character of a piece of music.

There are problems associated with using this type of method to explore the emotions listeners perceive when they hear a piece of music. Over extended listening periods it would be expected that the same piece of music would use a number of different musical techniques in order to continue to engage the listener. As we saw in the Hevner studies, the listeners chose several descriptors with different meanings because the music expressed several different emotions in sequence. One alternative is to use shorter extracts, but individual participants could still use very different adjectives to describe the emotion they perceive with short extracts. Using lists of adjectives can be problematic because they allow participants too much freedom in choosing descriptors that could contain very different meanings. This also makes it difficult to identify the musical features that are involved in the emotions selected. Where the same adjectives are used to describe the different musical features the experimenter could perhaps believe they behave in the same way. Gundlach (1935) showed that the same pieces of music may contain the same musical features, but not be defined using the same emotional terms. Arguably, the need to assess emotional responses to shorter extracts using fewer adjectives and ensuring the extent to which the responses are reliable is important.

More recently researchers have used rating scales to assess the extent to which one emotion is perceived over others. Some studies have examined musician's abilities to identify the emotions expressed in improvisations through the use of rating scales (Behrens and Green, 1993). They showed that the ability to do so was partially influenced by the musical instruments used and by the emotions being expressed. However, Gabrielsson and Juslin (1996) found that the ability to decode the performers' intended emotional expression was not affected by the use of different instruments. Others have shown that ratings of emotion can be made by patients with deficits in music processing after brain damage (Peretz, Gagnon, and Bouchard, 1998). Rating scales have also been used to assess listener's abilities to identify the emotions being expressed in music from different cultures (Adachi, Trehub, and Abe, 2004; Balkwill, Thompson, and Matsunaga, 2004; Balkwill and Thompson, 1999).

Like the final study by Hevner, much of the research that has been conducted in recent years has manipulated musical features in a bid to understand their relationship with the emotions being perceived. These studies have used rating scales to assess perceived emotion in music and to investigate the extent to which particular emotions are perceived over others for specific musical features.

Scherer and Oshinsky (1977) conducted a study where listeners were required to rate 3 types of tone sequences on three 10-point semantic differential scales. The scales contained the words pleasantness to unpleasantness, activity to passivity, and potency to weakness. They were then asked to rate whether each sequence could express any or all of the following words: anger, fear, boredom, surprise, happiness, sadness, and disgust.

The tones sequences were manipulated so that they produced 3 types of stimuli. Type 1 involved manipulating the amplitude (small and large), pitch level (high and low), tempo (fast and slow), pitch contour (upwards and downwards), envelope (attack) and filtered cut-off levels (intermediate and high). The type 2 stimuli consisted of tones sequences where the filtration levels (high, intermediate, and low) and tonality (major and minor) were also manipulated. Type 3 stimuli used a Beethoven melody where the filtered levels, tonality, rhythm and tempo were manipulated. Both the filtered levels and the tonality manipulations were the same as those in the type 2 versions, but the rhythm contained an even and uneven rhythm and a slow and fast tempo.

Four groups were assigned a ¼ of all extracts from each stimulus type to avoid any effects of fatigue and were presented these at random. A small number of each type of sequence was presented more than once in order to assess the reliability of the responses produced. For the purpose of this chapter we shall focus on the musical features that are of relevance to the areas we shall investigate and report in the subsequent chapters. Fast tempi were highly associated with activity, whilst surprise, happiness, pleasantness, potency, fear and anger were gradually less important. Slow tempi were associated more with sadness and less so with boredom and disgust. Small pitch variations were highly associated with disgust, whilst anger, fear and boredom were less highly rated. Large pitch variations were highly associated with happiness, whilst pleasantness, activity and surprise were gradually less important. By comparison to the use of adjective checklists, the present method is better because it can assess the reliability with which the listener's responses are made.

Additionally, the results suggested that there was a significant correlation between listener's ratings on these scales suggesting that they were capable of making similar responses overall. There were also no differences between the activity and potency ratings given for the repeated sequences. However, the second rating of pleasantness was judged much higher on the second rating. They suggest that this was due to the participants becoming more positive having heard 50 or so sequences. However, it was not clear why the listeners would do so over the course of the experiment.

If they were becoming more positive as they went through the course of the study one might argue that this would bias other ratings that were associated with this affective response. The only response a participant could make that is not contaminated by previous responses is the 1st response. A study using a single response per participant and per sequence could assess whether this was really the case and allow the experimenter the opportunity to assess the reliability of their responses more carefully. However, this would require a large population to be available, would be very time consuming and costly.

These studies indicate some of the issues regarding the use of rating scales in the perception of emotion in music. They are an improvement by comparison to adjective checklists because the experimenter can assess the extent to which an adjective would be used to describe the emotion being perceived, rather than relying singularly upon the frequency with which they are chosen (adjective checklists). Additionally they have the advantage of being quick and easy to administer, but also encounter problems because they allow the listener to have control over the response they make rather than the experimenter. This technique is troublesome because the experimenter cannot be assured that the listener is making the effort to respond appropriately. For example, a listener could give a neutral response for all of the extracts of music they hear, but the experimenter would not be able to tell if this was an actual reflection of their response or simply a random response. The results from these sorts of studies would not necessarily suggest whether this had occurred or not. Therefore rating scales are not reliable techniques for assessing perceived emotion.

During the course of listening to a piece of music a range of words to describe the emotion being perceived could be activated. Studies that involve listening to longer passages of music containing different musical features could activate a large range of words with related and unrelated meanings that would vary from moment to moment. Using adjective checklists or a number of different rating scales on different emotional dimensions would increase the difficulty of identifying those musical features involved.

To resolve these issues both the length of the extracts of music presented would have to be much shorter and the number of dimensions used to define the emotions being perceived would have to be reduced to a single dimension. In this way the possibility of identifying the actual relationship between the perception of specific emotions and particular musical features involved becomes more likely.

Studies assessing the relationship between tempo and emotion in music have used small variations in tempi that include only one fast and one slow variation (Hevner, 1935a; 1936; 1937, Scherer and Oshinsky, 1977). Gabrielsson and Lindström (2001) point out that this is not a useful way of establishing the role that particular features of music play in the perception of emotion. They also suggest that any emotional response must involve interactions between musical features rather than one single feature being completely responsible. By using a larger range of global tempi we might begin to see what its relationship is with perceived 'happy' and 'sad' emotions. In doing so we may also determine whether other features contained in these pieces interact with global tempo to produce any given response.

TIME PERCEPTION IN MUSIC

There has been a debate about how we experience time and a number of different theories have been suggested. One way in which we experience time is in the duration of events, but typically we are not very good at making accurate judgements. It has been argued that duration judgements rely upon internal mechanisms such as internal clocks (Treisman, 1963) or oscillator based systems (Church and Broadbent, 1990). They suppose that the poor ability to perceive time and the constant errors recorded in studies involving duration judgements are due to the variability in the operation of these internal systems. Alternatively, it has been suggested that we are not capable of perceiving time and what we actually experience instead are a series of events (Gibson, 1975). We suppose that the perception of time relies on features of these musical events and that these have relative strengths and weaknesses that deliver the constant errors that listeners produce rather than any internal system.

Fraisse (1975) makes an important observation that is useful for understanding the notion of time and events. He suggests that the difficulty in explaining time is a matter of a distinction. According to Fraisse there are two concepts of time; the concept of succession and the concept of duration. The concept of succession supposes that two or more events can be understood as separate from one another whilst being perceived sequentially. Distinct from this is the notion of duration which he describes as the interval between two successive events. Events are therefore sequential and distinguishable by the durations between them. However, this only determines how we separate one event from another and does not define what an event is. In addition, some events overlap where there is no marked gap between them, a notion that Fraisse does not recognise or describe.

Proffitt and Kaiser (1995) proposed that events are incidents that change over time. This seems a rather obvious description and it does not explain what an event is. Similarly, Brown (1995) and Poynter (1989) argue that the experience of time involves the perception of events and that the amount of change in a visual stimulus can lengthen perceived duration. Gibson (1975) takes a similar view and suggests 3 types of changes that visual events could encounter that include re-positioning objects, re-shaping a surface, and creating or destroying a surface. Interestingly, Poynter suggests that the main type of change an event can encounter is in its spatial location. The idea is that moving stimuli inhabit a sequence of different spatial locations and therefore provide the individual with a "dynamic pattern of changing events", (Brown, 1995, p105). Although these views were made in reference to vision the same is true of musical events.

Firstly, music contains notes that change over time in terms of their volume, duration and pitch. Secondly, there are various ways in which musical features can be repositioned, re-shaped, destroyed and created during the course of a piece of music. In terms of the re-positioning of objects it could be argued that the same musical note duration (a crotchet) can played on middle C and played on another note (D above middle C). The duration of the note is the same, but it has been re-positioned elsewhere on the diatonic scale (notes pitch). Re-shaping a surface in music would involve changing the volume of a note (notes intensity). In this case the object is the same, but its surface has increased or decreased in volume. The creation and annihilation of a surface would be the onset of a note or the offset of a note (notes duration). Creating the surface begins when the note onsets and is destroyed when the note discontinues.

It is clear that there are parallels between what is described above and the raw primal sketch for music. To recap we argued that the outcome of the raw primal sketch is due to an enormous amount of sensory processing, but at that stage only produces a note-by-note representation. We suggested that the process of perceiving a musical event requires the basic features of the notes to be established. Any temporal judgement will be based on these features and the way they change over time.

A musical event can also involve collections of notes that are pulled together into musical phrases and they too involve the same 3 types of change described above. A phrase can be re-positioned at other points during a piece of music by repeating it throughout a composition. This is what is known as repetition in music and is often used to emphasis a musical theme. Re-shaping the surface of a musical phrase would involve changing the volume of the notes so that they might increase for one phrase and decrease across a second presentation of that same phrase elsewhere. In this case each phrase is the same, but there surface has changed. The creation and annihilation of a musical surface would be the start of a musical phrase and the end of a musical phrase. These might be characterised by a rest (a pause) in the musical score.

The description provided also seems to resemble the full primal sketch of music outlined at the beginning of this chapter. It was suggested that we can only make sense of a musical passage once we pull notes together into larger musical groupings. Brown (1995) suggested that moving stimuli create a dynamic pattern of changing events. Musical stimuli involve the same effects and as the music progresses it changes in numerous ways in terms of its pitch, volume and duration. Each musical event we experience is therefore characterised in these 3 ways and they are fundamental to our experience of time. The representational outcome of a duration judgement is made on the basis of the processes underlying the raw primal sketch, but on some occasions will require a full primal sketch. Determining the amount of perceptual processing required either at the raw primal sketch or full primal sketch stage will be constrained by the requirements of the task.

There have been a number of different explanations for the way we assess the passage of time and the way we experience the duration of events. There are two types of task a participant can be asked to do in studies on timing namely a prospective and retrospective timing task (Zakay and Block, 1997). In a prospective timing task the participant is aware from the outset that they will be asked to judge the duration of a stimulus whereas retrospective timing tasks involve judgements based upon the participant becoming aware of the task once the stimulus has finished. In this thesis any timing tasks listeners are asked to do involve prospective timing.

There have been a number of mechanisms that have been used to explain our experience of time. In particular, internal devices that include internal clocks (Treisman, 1963) or oscillators (Church and Broadbent, 1990) and the amount of cognitive processing involved have been used to explain this experience (Ornstein, 1969; Casini and Macar, 1999).

Ornstein (1969) supposed that the perception of time is due to the amount of space in memory necessary to encode and store stimuli. The more complex a stimulus is the more likely it is to require a larger store as well as more processing time to encode a stimulus, causing perceived time to be longer. Similar assumptions have been made by models of timing involving temporal and non-temporal information processors (Thomas and Weaver, 1975), where the judged duration of a stimulus will also depend upon the amount of information encoded.

Buffardi (1971) conducted a study to assess this where a filled duration contained up to 5 beeps and un-filled duration contained 2 beeps. Listeners were required to judge whether the first or second duration was longer using a 2AFC. The results showed that the filled duration was perceived as longer than the unfilled duration even where they were of equal physical duration. It was suggested that this was due to increases in information processing and an ability to count the number of elements within a filled duration. If this is the case then it would be expected that as the number of elements increase beyond the listener's ability to count then the task might require even more informational processing to occur in order to decipher the possible number of elements present.

As the number of elements increase the expected outcome would be a continued increase in perceived duration. However, a number of more recent studies have shown that beyond some point the perceived duration of the filled duration seems to decrease as the number of elements increase (Thomas and Brown, 1974). As we often use chunking to reduce the amount of workload involved in processing complex information, the size of storage space required to sustain a chunked stimulus would be reduced. Arguably, this is what caused the judged duration of the filled duration to become shorter rather than longer because the participants were chunking the information together into more manageable units.

Treisman (1963) suggested a model of timing that involved an internal clock called scalar timing that includes a pacemaker, a switch, an accumulator, and a memory store. A similar model has also been suggested by Block and Zakay (1996) called the Attentional-Gate Model. Both of these suggest that in the presence of a stimulus a pacemaker becomes activated and starts to generate a series of pulses. These pulses pass through a switch to an accumulator that acts like a container. Information is then passed through to a memory store for a later comparison with another incoming stimulus. Any response made is based upon the number of pulses the pacemaker generated and the quantity passed through the switch to the accumulator. It has been suggested that the reason why fluctuations in perceived duration occur is due to the variable rate of the pulses produced by the pacemaker and the latency with which the switch closes and opens.

Several studies have shown support for the pacemaker-accumulator model. Wearden, Edwards, Fakhri and Percival (1998) conducted 3 studies to examine the finding that sounds are judged longer than lights that has been documented in previous studies (Goldstone and Lhamon, 1974; Goldstone and Goldfarb, 1972; Walker and Scott, 1981). The first study attempted to replicate earlier findings and involved 4 conditions that included an auditory/auditory presentation, a visual/visual presentation, a visual/auditory presentation, and an auditory/visual presentation, where the visual stimulus was a blue square and the auditory stimulus was a tone. Participants were required to decide whether the second presentation (a test stimulus) was the same length as a standard (first presentation). Both of the within-modality data showed a similar pattern overall, but the visual presentation was more variable than the auditory condition. More interestingly the between-modality conditions showed that the visual stimuli were perceived as shorter than the auditory stimuli when they were they were in fact the same physical duration.

The second study investigated this further and suggested that the results from study 1 might be due to the pacemaker-accumulator clock operating differently for visual and auditory stimuli. They hypothesised that the rate of the pacemaker might run more quickly for the auditory stimuli and would therefore accumulate more pulses as a result. By generating more pulses, the pacemaker would then increase the subjective duration of the auditory stimulus and account for the results they obtained. However, they also suggested that the latency of the switch might close and open at different rates for different stimuli in these modalities. They argue that this would also account for the differences in the variability for the verbal estimates the participants produced. A second study was conducted to assess this more fully.

In the second study the auditory and visual stimuli were the same as study 1, but in this case each trial only involved a single stimulus that were presented at a range of physical durations. The participants were required to make an estimate of the length of the stimulus in milliseconds and were told that all durations were between 50 and 1,500 milliseconds in length. The results revealed that both the verbal estimates for the auditory and visual stimuli increased in a linear fashion, but the auditory stimuli were perceived to be longer. These were most visible at durations where the stimuli were presented for longer. Therefore, the second study supported the initial findings from study 1 that auditory stimuli are perceived to be longer than visual stimuli and that the pacemaker must be running more quickly for auditory stimuli. When an assessment of the variability of estimates for the 2 conditions was conducted they tended to decline with increases in length and were most noticeable for visual stimuli. They suggest that this was due to the variable latency of the switch mechanism within the pacemaker-accumulator clock.

A further study was conducted to assess whether pacemaker speed and variability differences could be manipulated independently. In the third study, participants were presented with the same visual and auditory stimuli as before except that on half of the trials they were preceded by a series of clicks. The idea behind this was taken from an earlier study by Penton-Voak, Edwards, Percival, and Wearden (1996) who showed that auditory and visual stimuli preceded by click trains produced an increase in their subjective duration. They argued that this was due to an increase in pacemaker speed.

The pacemaker is believed to be a Poisson emitter that generates pulses at random such that the duration between pulses are variable, but over longer time spans the variability is averaged out. Wearden et al. (1998) supports this notion, but also suggests that variability in the verbal estimates of stimuli might be due to the variable latency of the switch. They suggest that if the pacemaker is a Poisson emitter then we should not expect this to account for any differences in variable errors found in the same modality.

The condition where the visual and auditory stimuli were not preceded by clicks replicated the results from the study 2. When the visual and auditory stimuli were preceded by clicks these were judged to be much longer than the without-clicks condition and were shown to be greatest for longer stimulus durations. Since both the visual and auditory estimates were increased equally when they were preceded by clicks they suggest that this was due to an increase in the speed of the pacemaker.

In addition they found that visual stimuli produced more variable estimates than auditory stimuli. These were not however, dependent upon the presence or absence of clicks. They suggest that this evidence supports the notion of the variable latency in the operation of the switch. In addition they suggest that this shows that the pacemaker and switch are independent from one another. As the estimates are different for visual and auditory stimuli then the pacemaker has to be running at a different rate for different modalities. Since the variability is greater for the auditory stimuli then they suppose that this must be independent of the pacemaker and more likely to be due to some other feature of the model (in this case the switch). As we shall show in the 4th experimental chapter variable and constant errors can be accounted for by both the pacemaker and switch mechanism and this questions the reliability of the assumptions made by Wearden et al. (1998).

The idea that there are 2 independent pacemaker-accumulator systems seems entirely without foundation based upon the results from study 2. If the visual pacemaker is affected by auditory stimuli (the clicks presented before the stimulus to be judged) this would actually argue against 2 systems. They suggest that there are two independent systems that deal specifically with auditory and visual information. However, as an auditory stimulus can effect the perceived duration of a visual stimulus and therefore cross modalities it would seem that this would fall in favour of a single system not a dual system. One might also argue that a visual flash could effect the perceived duration of an auditory stimulus. Even in the absence of any data to support this last statement it would seem that the results point to a single pacemaker-accumulator model that deals with visual and auditory information.

One important point that seems to have been missed by the previous studies involving lights and sounds is the conflict between the types of stimulus chosen and the marked increases in perceived duration. It seems difficult to imagine how one might choose a visual and auditory stimulus that would contain equivalent features in both domains. A blue square and a tone are very different stimuli that have no features in common to allow a true comparison. By using stimuli that contain common features an explanation for their effects on time can be gained. In light of this it seems that the notion that these 2 stimuli would be perceived equally seems unlikely. If they do not contain similar sorts of features then it should not be too surprising that their durations were estimated differently.

Goldstone and Lhamon (1972) argue that auditory stimuli transmit more information than visual stimuli. This has been supported further by Walker and Scott (1981) who showed that reducing the intensity of an auditory stimulus caused the perceived duration of a 500 millisecond light to be longer than a 500 millisecond tone. By simply changing some feature of the auditory stimulus the modality difference disappears. The results described in the Wearden et al. (1998) study might be due instead to the intensity of the tone over the light. Changing the intensity of a stimulus might provide crucial information about the way we process time. By focusing on the stimulus, we might explain how we experience time rather than concentrating on the variability of an internal clock to explain the process.

Tse, Intriligator, Rivest and Cavanagh (2004) recognised not only the importance of the stimulus for perceived duration, but the effects of changing their characteristics. In a series of studies the researchers assessed the effects of changes within a stimulus and its effects on perceived duration. Their studies involved the use of 4 types of psychophysical techniques to assess a prospective timing task: the method of constant stimuli, the method of magnitude estimation, the method of stimulus duration and the method of single stimuli. Since the studies in this thesis are concerned with the method of constant stimuli we shall focus on the results from this technique only.

Participants were presented with a standard and 'oddball' stimulus and asked to make a 2AFC of longer or shorter for an 'oddball' stimulus. The standard was a black circle with a fixed radius of 363.6 arcmin, whilst the oddball was an expanding disk that varied from a radius of 63.6 to 211.7 arcmin. The physical duration of the oddball stimulus was also varied from 450 to 1,050 milliseconds and the standard lasted 1,050 milliseconds. The PSE values for all 3 participants showed that they were overestimating the oddball stimulus by comparison to the standard. The results overall showed that the oddball stimulus at 675 milliseconds was perceived to be as long as the standard at 1,050 milliseconds. In a second they used exactly the same method as above except that the standard stimulus was also varied at 7 different physical durations from 75 to 2,100 milliseconds and the duration of each oddball stimulus was subsequently varied around each standard. The results showed that at 75 milliseconds the oddball was underestimated. As the standard increased in duration the oddball stimulus was increasingly overestimated by the participants.

They then went on to explore the possibility of changes to the characteristics of the oddball other than size and duration using the same method as the first study. Six different oddball stimuli were chosen that involved; a red stationary disc presented within a set of 6 black circles of the same size, a round circle placed within a series of 6 squares of an equivalent size, a square among 6 circles of equivalent size, and a large circle within a set of 6 smaller circles. In every case the oddball stimuli were overestimated by comparison to the standard. A comparison between the above conditions and the second study mentioned above showed that the strongest effect with the smallest variability was the second study.

Interestingly, the authors go on to look at these in the context of auditory perception. The second study we mention was repeated, but the standard in this case was a pure sinusoidal tone of middle C whilst the oddball was a rising tone from 20 semitones below middle C to 30 semitones above middle C that used a glissando effect that rises smoothly. The method employed here was a replication of the second study. They found that the overall pattern of results were similar to the visual case. However, the visual version produced a greater time expansion effect by comparison to the auditory version at durations beyond 500 milliseconds. The error bars also showed that these were significantly different from chance beyond 500 milliseconds. This indicates that auditory information is overestimated more than visual information.

In explaining there results the authors turn to an information processing account that is similar to those outlined earlier. They suppose that as the amount of processing demands increase so too does the perceived duration of a stimulus. They also suggest that a low-probability stimulus would orient more attentional processing that a high-probability stimulus. These would then cause perceived duration to fluctuate. Although this might be the case the authors have overlooked something essential. It is possible that the offset of the auditory stimulus on a high pitch could have caused the bias observed. It could be argued that an increase in pitch causes listeners to perceive the stimulus as longer because it creates a sense of motion. The same is true of the other stimuli they used where the stimulus getting larger produces the effect of the stimulus moving towards them. If they are perceiving motion when they are presented with these stimuli, then the results could be explained in these simple terms. Had they created a stimulus that smoothly rose and descended back to the same note this may have been perceived as changing its location rather than getting larger. Future studies should assess this feature to establish if this is the case.

Although much of the literature uses assumptions that lead us to believe that internal devices are responsible we shall show that this is not a reasonable direction to take. Additionally, the empirical research that has led to these conclusions, is based on assessing the way that the internal devices behave rather than the role of the stimulus. Those who have supported the clock stance have used static fixed stimuli that do not reflect the constantly fluctuating nature of the real world.

We argue that it is the characteristics of an event that provide us with enough information to make temporal judgements. It is not that we measure them or use an internal device to do so it is simply that we assess their strength on the bases of the characteristics of the stimulus being presented. As such, some events are more likely to be perceived than others and as a result we make duration judgements about them based upon their strength in comparison to other events in their locality. We suggest that events have varying degrees of strength where weak events cause perceived duration to be shorter because they are common events. Comparatively, strong events are perceived to be longer because they are uncommon events. The extent to which an event is common and uncommon is based upon an individual's previous knowledge about a given stimulus.

EXPERIMENTAL CHAPTER 1

We argue that the optimal tempo for a piece of music is driven by features contained within a piece of music and that the relative strength of these features determine the tempi a piece of music should be played at. The first experimental chapter will deal with this and show that there are tempi that listeners perceive to be optimal for pieces of music and that rhythm plays an important role.

EXPERIMENTAL CHAPTER 2

In the previous studies, it is possible that listeners might be using an average tempo from previously heard extracts to make every subsequent response. We wanted to assess this by presenting a single stimulus per participant and therefore remove any effects of the context on participant's responses. Using this technique we shall show that listeners can make 'too fast' and 'too slow' responses that are independent of previously heard extracts. In addition the data reveal similar results to those found in the first experimental chapter.

EXPERIMENTAL CHAPTER 3

The purpose of the 3rd chapter is to assess the perception of emotion in music using the method of constant stimuli and asking listeners to make a 2AFC of 'happy' or 'sad' to musical extracts. This shall show that the perception of emotion in music is driven by the size of pitch changes contained within the music. It shall show that tempo and the size of pitch changes used can change the perceived emotion of a piece of music. We argue that the uncommon pitch changes are strong events and common pitch changes are weak events and that the combination of weak and strong events deliver the perceived emotion of a piece of music.

EXPERIMENTAL CHAPTER 4

In considering what might be involved in the perception of time in music we wanted to assess what effect small changes to a stimulus would have on perceived duration. We presented 2 auditory stimuli and show that the perceived duration of the test stimulus with a change in pitch increased as the size of the pitch change increased. The results are explained in terms of event strength where strong events cause perceived duration to increase whilst weak events are perceived to be shorter by comparison.

CHAPTER 2: METHODS

A number of different methods can be used to measure the Point of Subjective Equality (PSE). Three main methods are commonly used in perception; the method of adjustment, method of constant stimuli and the method of limits. Each of these has benefits and limitations and during the course of the next section we shall discuss what these are and why the method of constant stimuli is preferred.

The method of adjustment is the easiest and quickest method to use in psychophysics. A variable stimulus is initially set to a value far above or far below the PSE. Participants are required to adjust the value of a variable stimulus until it matches a standard stimulus. This process is repeated a number of times and the average of these responses is used to provide a value of the PSE. Clearly this is a very quick way of assessing the participants' perception, but there are limitations in conducting studies in this way. The initial setting of the stimulus often dominates the responses in that stimulus values far above the PSE usually lead to responses close to this value and likewise stimulus values far below the PSE usually provoke responses close to the value presented. Additionally, this method causes problems because the diet of stimuli the participant hears is under their control rather than manipulated by the experimenter. This causes biases in the participants' responses and leads to questions over the reliability of the PSE value gained. The reason why this is the case is because the participants' perception of each stimulus is influenced by recent stimuli. In this method, all of the recent stimuli are drawn from just one part of the stimulus range. Any response that is calculated from stimuli that are drawn from just part of the range is unlikely to be uncontaminated by the procedure.

Alternatively, the method of limits can be used in perceptual studies. In this method the experimenter has more control over the stimulus, in that they present a stimulus at a value well below or above the PSE and ask the participant to make a 2 Alternative Forced Choice (2AFC). If the participant makes a response in one direction the experimenter would gradually increase the level of the stimulus until the participant makes the alternative response. This provides one PSE value. Once the response changes the experimenter restarts the measurement process again starting at a stimulus level well above the PSE. The participant is then asked to make a 2AFC and once the experimenter decreased the stimulus to a level where their response changes again this would provide the experimenter with another PSE value. This procedure continues a number of times and the average of the points where the participant switched from one response to the other is used as a measure of their PSE.

The only difference between this procedure and the one used in the method of adjustment is that the stimulus level is manipulated by the experimenter and not the participant. Again this is problematic for the same reasons we described above for the method of adjustment.

The most effective psychophysical method is the method of constant stimuli. In this method the experimenter chooses a range of stimulus levels that extend well above and below the PSE prior to starting the experiment. Each stimulus level is presented a number of times during a run and in a fully randomized sequence. Depending upon what the experimenter wishes to assess they might present either a single stimulus or paired stimuli. Participants are then required to make a 2 alternative forced choice (2AFC) for each trial. Both the paired and single stimulus trials require participants to discriminate between 2 stimuli (paired stimuli) or the single stimulus and an internal representation (single stimuli). The PSE is measured at the point where the participants are 50/50, so that on 50% of the trials they make one response and on the other 50% of trials they make an alternative response.

The data is plotted as the proportion of times the participant responds in one direction as a function of each stimulus level. This produces a psychometric function which rises from 0 to 1. The PSE is the point where the curve crosses 0.5 and the sensitivity to changes in the stimulus is assessed by the slope of the curve. Sensitivity to the stimulus is important since it gives a measurement of how reliably the participants are responding. If the slope is steep we can be more assured that the PSE is reliable. However, if the slope of the curve is much flatter the PSE becomes more variable and less reliable.

The only limitation of using this method is that the study takes a much longer time to conduct. The benefit of using this method is that control over the experimental procedure is in the hands of the experimenter. All biases are removed by using this method because the participant is not controlling the stimulus in any way and is getting a random but balanced diet of stimuli that remove the problems experienced in the previous 2 methods.

The analysis of the data is derived from the bootstrap technique that produces 95% confidence limits for the PSE. Each psychometric function is analysed by finding the least squares fitting cumulative normal function using a simplex minimization routine (this is equivalent to the pre-computer technique of probit analysis). The curve is then represented by 2 parameters, the PSE (the point at which the function crosses 50%) and the SD (inversely proportional to the slope of the curve). The SD is the increase in stimulus level required to raise performance from 0.5 to 0.83, and is regarded as a measure of sensitivity to stimulus difference. The best fit of the curve provides estimates of the PSE and SD, which are related to the actual PSE and SD but are subject to sampling errors. It is important to know what the size of the sampling errors are so that different PSE values can be compared. A confidence limit approach is adopted throughout that establishes the range within which the population value is likely to fall 95% of the time. The bootstrap technique is used to determine the range of possible PSE values that might produce the given psychometric function. If the 95% confidence limits overlap the PSE values could be the same, however if they do not overlap this suggests that they are different.

The Bootstrap technique is commonly used in psychophysics instead of more traditional statistics because it is more adaptable to psychophysical techniques such as 2AFC. Raw data, as collected from participants, is re-sampled with replacement to make a simulated replication of the experiment a large number of times. The distribution of outcomes from the large set of replications is then a measure of the range of possible outcomes given the population. The only assumption required is that the sample is a fair representation of the population.

So, a psychometric function with 10 responses at each of 7 stimulus levels will yield a measured PSE and SD. From the same data, a simulated psychometric function can be created by sampling 10 times with replacement at each of the 7 stimulus levels, the corresponding 10 real responses. The new set of 70 responses is analyzed to establish its PSE and SD. The whole procedure is then repeated 1,000,000 times, giving 1,000,000 estimates of the PSE and the SD.

From these sets of simulated values for PSE and SD, it is possible to obtain confidence limits and also when appropriate conduct a null hypothesis test:

i) The 95% confidence limits for the real data are obtained by sorting the simulated estimates by numerical value, removing the extreme 2.5% of values at either end and taking the limits at either end that are left.

ii) Where a null hypothesis approach is more natural, the basic procedure is the same. The probability that the measured PSE is greater than 0, for example, is given by the proportion of simulated PSE values greater than 0. If 12% of the simulated PSE values are greater than 0, then the null hypothesis that the PSE is not greater than zero cannot be rejected, because $p=0.12$ (and therefore $p>0.05$).

Having collected a set of data from participants we can establish the mean value of the data. The idea behind the bootstrap technique is to generate a set of virtual data by resampling the original data set with replacement. In resampling the data a million times we can assemble a million possible mean values. The means are then put into numerical order and 2½ percent at each end is removed leaving us with 95% of the data we have resampled. The true mean can then be said to lie within the remaining range with 95% confidence. This technique is used throughout the course of the following experimental chapters and is the most reliable for establishing the slope and PSE values we obtained in the studies.

CHAPTER 3: THE PERCEPTION OF TEMPO IN MUSIC

ABSTRACT

Tempo is one factor that is frequently associated with the expressive nature of a piece of music. Composers often indicate the tempo of a piece of music through the use of numerical markings (beats per minute) and subjective terms (adagio, allegro). Three studies were conducted to assess whether listeners were able to make consistent judgements about tempo that varied from piece to piece. Listeners heard short extracts of Scottish music played at a range of tempi and were asked to make a two alternative forced choice of 'too fast' or 'too slow' for each extract. The responses for each study were plotted as proportion too fast responses as a function of tempo for each piece, and cumulative normal curves were fitted to each data set. The point where these curves cross 0.5 is the tempo at which the music sounds right to the listeners, referred to as the optimal tempo. The results from each study show that listeners are capable of making consistent tempo judgements and that the optimal tempo varies across extracts. The results also revealed that rhythm plays a role, but not the only role in making temporal judgements.

INTRODUCTION

There have been a number of attempts to investigate tempo in music. Some studies have concentrated on the variations in expressiveness of performances of pieces of music (Timmers, et al., 2000). Others have focused on the features of expressive performances such as rhythm (Repp, et al., 2002), meter (London, 2002), dynamics and mode (Kamentsky, Hill and Trehub 1997). Some have shown that infants as young as 7 to 9 months are capable of discriminating between simple tonal rhythmic patterns that varied in tempo (Trehub and Thorpe 1989). Whilst these investigations offer an important insight into musical abilities and the temporal deviations within performance, little is known about the processes involved in the perception of global tempo and whether there are optimal tempi which listeners will perceive to be the most appropriate and perhaps expressive for pieces of music. The purpose of this chapter will be to investigate this more fully. The material in this chapter has been published as Quinn and Watt (2006).

The main focus of psychological investigation into tempo has involved interpreting the expressive timing deviations of musician's performances. As these deviations often mark moments in the music that heighten the expressive nature of a piece, such evidence is clearly important to our understanding of music. It has been argued that detecting deviations from global tempi in real-time performances is highly developed in musicians due to their exposure to listening and playing music (Repp, 1992). Subsequent evidence indicated that the perceptual experience of temporal deviations is influenced by musical structure and not necessarily through specialised musical experience (Repp, 1999). Therefore, individuals with no formal training should be able to perceive deviations in tempo.

Temporal variability between successive note onsets is due to the musician's interpretation of the expressive nuances of the music (Repp, et al., 2002). Large and Palmer (2002) point out that listeners are remarkable in their ability to hear musical events in terms of discrete durational categories (such as whole notes or eighth notes), even in the presence of changes in continuously varying timings that are present in human performance. It is supposed that there is an internal system that allows listeners to keep track of the durational values of notes in a piece of music and to perceive deviations in these during human performance.

Dynamic Attending Theory (Jones, 1976, 1987, 1990; Jones and Boltz, 1989) suggests that attention to an auditory stimulus is directed by an oscillator (Drake, Jones and Baruch, 2000). In the presence of musical stimuli, the oscillator will synchronize itself to the periodicity of physical characteristics of the music, such as accents and the steady beat of the music. This theory also allows for music containing periodicity at more than one level, such as melodic and harmonic changes, by suggesting that the listener makes use of multiple oscillators. These multiple oscillators are thought to produce an expectancy schema, which anticipates the succeeding temporal pattern. Specifically, the expectancy schema is organised at three levels of attending; referent level, future-oriented level and an analytic level. The last two levels are utilised for attending to key and harmonic changes (future-oriented attending) or dynamics and tone onsets (analytic attending). Listeners are thought to switch between levels through a process of focal attending, to the strongest expectancy within the stimulus. The particular oscillator that is used will be determined by the most salient features of the music. Although this theory can explain the way listeners deploy attention over time whilst listening to music, it is not intended as an explanation for the perception of global tempo in music.

Schulze (1978) attempted to assess the way listeners perceive temporal regularity in auditory sequences by testing three models of tempo perception. The first model assumes that a listener perceives temporal regularity by comparing neighbouring intervals to decipher whether the sequence is regular. The second model supposes that listeners produce their own rhythmic pattern and use that to judge the regularity of the sequence. This possibility is based upon the assumption that an internal mechanism is involved called a 'time keeper' that synchronises with the input sequence, but in order to do so has an identical base duration to the input sequence. Lastly, the third model suggests that once the listener has heard the first few intervals of a sequence they produce their own internal representation of the interval. This is used as a reference to judge all other incoming intervals and the stability of the sequence. The first model relates to local tempo from note onset to note onset, whilst the other two relate to global temporal effects.

Each listener heard a regular sequence and one of three comparison sequences where the regularity of the tones had been displaced. Listeners were asked to judge the comparison sequences to determine whether they were either regular or irregular. The results suggested that listeners were capable of judging the regularity of the sequences. Particularly, the results supported the notion that an internal time keeper was involved and that it was not based upon comparisons with further intervals. This is essentially similar to the assumptions made by Dynamic Attending Theory and supports the global internal mechanism put forward by Schulze.

The stimuli in the Schulze study were sequences of tones where the pitch was held constant. Given this, the extent to which those results can be generalised to music listening is unclear. It would be of interest to assess musical sequences that contain variations in pitch as well as the rhythmic structure where there was a musical context for assessing the perception of tempo. Furthermore, the fact that listeners were told where to expect changes in these tonal sequences seems to suggest that listeners had an expectation of what might follow. Had they introduced a control group to assess listener's capabilities where they had no prior information, the results may have suggested that the other models were more appropriate in this context. Whilst this is not clear from these results, future studies found support for these findings by utilising a psychophysical methodology.

Based upon the assumptions made by Schulze's Internal Time Keeper, Vos, Assen and Fraňek (1997) attempted to assess listener's detection of tempo changes. This study investigated the role of the internal time keeper suggested by Schulze (1978) in order to predict whether listeners could detect accelerations and decelerations in the tempo of sequences of music. It was thought that if the base tempo was around 500ms to 700ms, the results would not be biased. If the tempo was faster than this, the listeners were assumed to become biased towards responding that they perceived an acceleration in tempo. Where the base tempo was slower than this they were thought to be biased towards judging the sequence as decelerating. In order to assess this, listeners were required to make a Forced Directional Change Response of 'tempo acceleration' or 'tempo deceleration' for sequences of tones. They were told that the stimulus would change in tempo at some point during the listening phase and they should make a decision once the sequence had finished. Even when they were unsure of their response they were asked to use their best guess. In essence this methodology is more useful because it allows the possibility of assessing the listener's sensitivity to changes in the stimulus as well as their detection of such changes directly.

The psychometric functions of the proportion of acceleration responses were plotted as a function of tempo change. The psychometric functions clearly showed that listeners were capable of detecting temporal changes in both directions and indicated their sensitivity to those changes. The results provided support for Schulze internal system of tempo perception and the assumptions made by the authors. Indeed, they showed that listeners responded as often on both choices and were unbiased when global tempo was around 500 to 700ms. Secondly, they showed that the listeners would become biased towards acceleration responses when the global tempo was faster than the unbiased tempi and deceleration responses when the sequence was slower than the unbiased tempi.

Few studies have evaluated whether there is a global tempo that listeners perceive as optimal for the expressiveness of music. The amount of expressiveness not only relies on temporal deviations from global tempi, but for such deviations to occur global tempi must play a crucial role. If a piece of music is played at a tempo that is too slow for a piece then this would affect any subsequent temporal deviations that performers use to heighten its expressiveness.

One study assessed whether memory for global tempo was absolute (Levitin and Cook, 1996). Their study assessed the ability to match a tempo accurately to pieces of music that were familiar to the participants. Like absolute pitch, where some individuals have an ability to recognise the exact pitch of a note, they suggest that people may hold a similar ability to hear tempo in an absolute manner, where they retain information about the exact tempo for a piece of music. They referred to this as absolute tempo. Data from a previous study evaluating whether people could sing a familiar song at the same pitch as the original composition was used. It was found that the participant's reproductions were at similar tempi to the original compositions. From this, they conclude that long-term memory for tempo in music is accurate at least for music that is very familiar to the listener. As memory for tempo in music seems to be highly accurate it could be argued that presenting a stimulus more than once allows a listener the opportunity to pull upon information they have stored in memory. It follows that studies measuring perception (rather than memory) should not present musical stimuli more than once to participants.

Lapidaki (2000) attempted to assess whether listeners could set the tempi of pieces of music in a consistent manner. It was assumed that if preferred tempi exist, different listeners would produce similar settings. Using the Method of Adjustment, subjects set each piece of music to their own preferred tempo. Judgements were relatively consistent across most listeners. The initial tempo dominated the settings of preferred tempo: a slow initial tempo seemed to provoke slower tempo selections, and so on. The results showed a bias effect and do not necessarily show the optimal tempi for pieces of music and the musical features that drive such responses.

Repp (1994a) attempted to assess whether changes in global tempi could affect the timing of aspects of a piece of music. Schumann's Träumerei for the piano was chosen as it was thought that this piece was not technically difficult, but demanded a great deal of interpretation by a performer. Two pianists played the piece three times, at a slow, medium or fast tempi that were chosen because they reflected ranges of appropriate tempi used by famous pianists in recorded performances of this piece (Repp, 1992). Each tempo was introduced to the pianists by a metronome so that they could get a feel for the specified tempi before they performed the piece. The metronome was switched off prior to their performance and the piece was recorded onto MIDI. In the perceptual condition, each performance was played to highly trained pianists who were asked to identify the original performances from those that had been artificially speeded up or slowed down to the same performed speeds the pianists had played.

The results showed that the tone onsets timings were similar over all, but deviated from proportional tempi. The perceptual tasks showed no real pattern of identification and suggests that very little difference in the musical expression was found when tempo was increased. It may also suggest that musical ability does not seem to aid listeners' discriminations. The timing of the grace notes of the performances at all tempi showed a consistent pattern of regularity. It was thought that this was due to the relatively slow general tempi of the performances and may not have been the case had the piece included faster ornaments. Synchronisation of the chords in each performance varied little, even when they were particularly important to the overall expressive shape of that moment in the music. It was expected that as the speed became faster the legato overlap between note-to-note onsets would become less pronounced. The amount of overlap between legato notes was shown to differ as temporal variability changed at a local level. However, this was only the case for one of the pianists as their use of legato was of a more typical level and such differences between each performer were thought to be the result of differences in technical ability. Another aspect that was not affected by global temporal changes was pedal timing which showed no real perceptible effects. Lastly, the performance intensities were also assessed by looking at changes in velocity and although these were shown to increase, the pattern of changes was held relatively constant. As these results seem to focus on the technical aspects of the musician's performance, it is not clear whether there were musical features such as rhythm or pitch that might play a role. Desain and Honing (1994) found evidence to suggest that expressive timing in musical performance does not rely on global tempo. Their results imply that timing in human performance is related to the actual structure of the music rather than the global tempo chosen by the performer.

Repp (1994a) suggests the need to explore a variety of different types of music to evaluate these effects more fully. He also highlights that rhythmically unpredictable music that includes short notes, rests, and a variety of articulation might be affected more by changes in global tempi. However, the limited number of pieces included in both Desain and Honing (1994) and the Repp (1994a) study cannot substantiate this claim. Repp (1994b) implies that pianists will have an interpretation for familiar pieces of music and when they are asked to play these pieces at a different tempi without much notice, their interpretation will not change. Had the musicians had more time to practice this at a different speed they may have changed their interpretation. In contrast, one might argue that the overall similarities between these performances and those recorded performances where global tempo seems to differ and the time spent on ensuring the interpretation was secure, leads one to believe that this might not be the case, certainly for this piece of music. A more typical argument would be that overall expressive nuances are the same in most performances of experienced musicians, but vary perceptually in listeners levels of satisfaction, interpretation or the appropriateness of the piece (one aspect not assessed in the study).

Understanding the basic tempo is necessary in order to clarify whether there are interpretative tempi that are deemed to be most appropriate rather than only measuring the ability to match performances with other such performances in the musical domain. Similar to the study conducted by Lapidaki, it could be argued that performers are biased towards particular tempi because of their familiarity for music. To avoid these biases the need to explore unfamiliar pieces of music should be conducted. Indeed, listeners were not asked whether the tempi being performed were deemed to be optimal or appropriate. As this might add to a body of knowledge on this topic, a thorough investigation of whether there are optimal global tempi for pieces of music should be made.

Boltz (1998) investigated whether temporal and melodic accent structure also influences tempo discrimination. In the first study, listeners were asked to complete a paired tempo comparison task where the comparator melody differed from the standard in terms of the number of changes in pitch contour direction and the number of pitch skips. It was assumed that the number of contour changes and pitch skips involved in listening to music would affect the perceived tempo of a piece of music. Particularly, it was assumed that the greater the number of contour changes and the greater the number and magnitude of pitch skips, the more likely the piece would be perceived as being slower than a comparison melody. It was suggested that these changes could cause listeners to perceive prolonged accents and retards at phrase ending points and are enhanced or reduced further by actual changes in tempi. Boltz found that manipulations of pitch content did affect the perceived tempo. This offers some insight into the musical features that affect listener's perception of tempi in music.

A second study examined the effect of rhythm in combination with changes in pitch on the perceived tempo of a melody (Boltz, 1998). It was thought that compatible and incompatible rhythms might differentially affect the perception of tempo. Where a melody and rhythm complement one another they contain temporal and melodic combinations that allow the listener to perceive the music in a single coherent way in order to make sense of the music. Where this is not the case, the music is perceived as being ambiguous. The study assessed this by asking the listeners to judge the tempi of melodies that contained compatible or incompatible rhythms. Boltz suggests that a rhythmic structure that conflicts with the melodic aspects within a piece can lead the listener to disengage from the anticipated flow of the tempo. This lack of cohesion leads the listener to perceive the music ambiguously.

It is the purpose of this chapter to explore and identify those factors that are involved in the perception of the appropriateness of a given global tempo. The method of constant stimuli will be used to measure this and avoid the bias encountered in the Lapidaki study. Participants will be asked to judge whether a piece of music is played too slow or too fast. The point at which the probabilities of the two responses are equal will be called the optimal tempo: it is the point at which the tempo is neither too fast nor too slow and is therefore presumably judged appropriate.

Three key issues will be focused on. Firstly, are listeners able to make reliable judgements about the appropriateness of the tempo of pieces of music? If listeners are capable of this, then is there agreement between listeners about the optimal tempo? Finally, does the optimal tempo vary from piece to piece? If it does, then the implication is that factors within the music, (e.g. pitch contour and/or rhythm) are determining the optimal tempo.

METHODS

Participants

Participants were undergraduate students from the University of Stirling. The number of participants varied from 17 to 21 in each study. Most were Psychology students; none were music students. No attempt was made to either select or filter out musically trained subjects: the incidence of musical training in the sample was around 2%, and probably reflects the population incidence generally.

Materials

The stimuli were 23 extracts of Scottish fiddle music (See Appendix 1). They were all monophonic, were in various keys and had indicative score tempi that ranged from 31.5 to 132bpm. Although the genre of the music was familiar to the listeners, the stimuli were selected on the basis that the pieces selected would be unfamiliar to the listeners. Each extract was constructed by the experimenter who performed each extract on a Roland E-28 intelligent keyboard whilst recording them as MIDI data (APPENDIX 2 - CD DISC) using Sibelius 3 or Cubase VST Score. All note durations were set to the full length of the note and velocity was set to a value of 80 on the MIDI data. The MIDI instrument was set to a violin and all notes started at the score time without variations such as rubato. Each piece of music was played at seven different tempi. In studies 1 & 2, the tempi were set to the score tempo ±0, 10, 20, 30 beats per minute. In study 3, the tempi were set to variations about the optimal tempi found in studies 1 & 2, rather than about the score tempi. All expressive features were removed to assess the musical structures rather than human performance. Stimuli were played to participants through high quality headphones (Sennheiser HD 250 linear II headphones) at a volume level that was comfortable via the line out socket on the PC.

Procedure

Before the beginning of the experiment proper, each participant was given a short practice trial in order to familiarise them with the task. The stimuli were then played to the participants and after hearing each stimulus they were asked to make a response by indicating whether they thought it was played too fast or too slow by pressing designated keys on a computer keyboard. They were instructed to make just one of the two responses, even if they were unsure about their response.

Design

The experiment made use of a within subjects design. There were two independent variables of musical extract and tempo. The dependent variable was a forced choice of 'too fast' or 'too slow' response. The design was constructed to avoid, as far as possible, the possibility of the stimulus set providing a normative tempo against which participants could make the tempo judgement. Participants heard the set of stimuli in a completely randomized order with neither extract nor tempo blocked. This avoids any effects of a tendency to judge the tempo of any given stimulus with respect to its immediate predecessors. Each participant heard each stimulus level just once, to avoid any possible memory effect. Each participant was assigned to one of two groups. Both groups heard 4 of the extracts (extracts 1, 2, 9 and 11) and the remaining extracts were split between the two groups so that group 1 heard an additional 10 melodies and group 2 heard 9 additional melodies. Each melody was played at seven different tempi making a total of 98 trials for group 1 and 91 trials for group 2. This was done to make the total length of time required from each participant manageable.

Three separate studies were conducted. In study 1, the participants were required to wait until the stimulus had been played in its entirety before they could make a response. In study 2, the participants were free to respond as soon as they were sure of their response, irrespective of whether the extract had finished or not. In study 3, the stimuli were made of a non-pitched drum beat rather than a pitched instrument, thereby removing the melodic component of the stimuli but leaving the rhythmic element unchanged.

RESULTS

Study 1 and Study 2:

The data were collected into psychometric functions for each extract, giving the proportion of participants responding 'too fast' as a function of the tempo of the music. Three sample psychometric functions and the data are shown in Figure 1. For each extract, the proportion rises as tempo increases; in most cases the psychometric functions produced a full range of responses from 0 'too slow' to 1 'too fast'. Each psychometric function was analysed by finding the least squares fitting cumulative normal function using a simplex minimization routine (this is equivalent to the pre-computer technique of probit analysis). The curve is then represented by 2 parameters, the PSE (the point at which the function crosses 50%) and the SD (inversely proportional to the slope of the curve). The PSE is the point at which the responses 'too fast' and 'too slow' are equally likely, and is thus the point at which the tempo is perceived as optimal by the participants.

Figure 1: Proportion of 'too fast' responses as a function of tempo for 3 extracts used in study 1

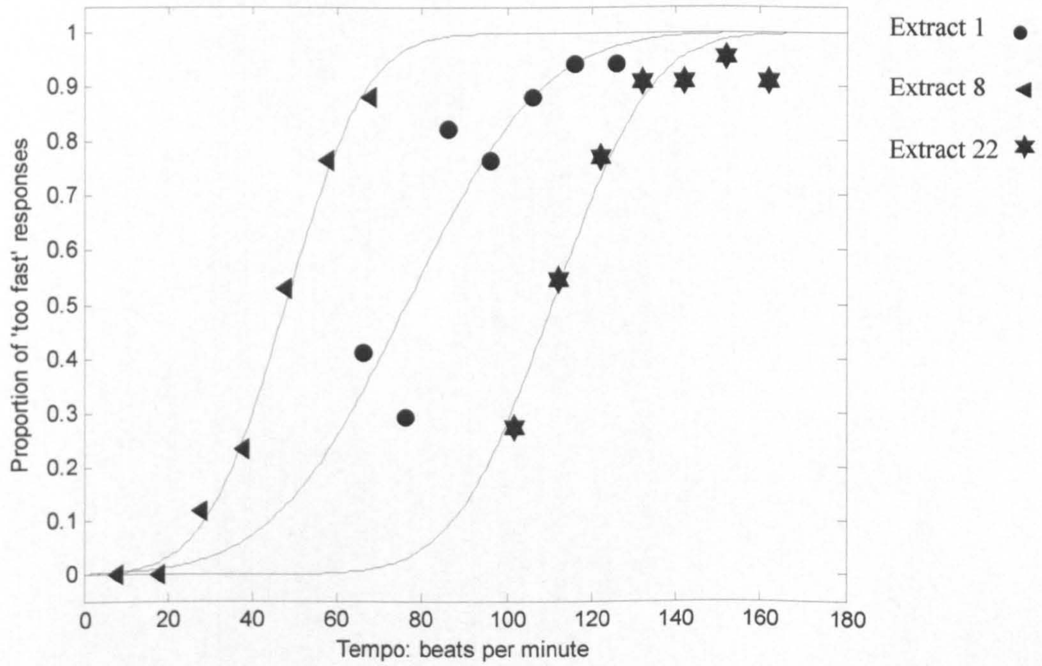

Figure 1: The proportion of 'too fast' responses as a function of tempo as beats per minute for extracts 1, 8 and 24. The symbols are data points; the smooth curves are least squares fits to the data. The psychometric functions rise as tempo increases and participants produced a full range of responses from 0 'too slow' to 1 'too fast'. The optimal tempi can be obtained at the point where the psychometric curves cross 0.5 and produced a range from 45 b.p.m. to 120 b.p.m.

Figure 2 shows the values obtained for the optimal tempi for each extract. Some of the extracts appear more than once to show how replicable the results were in study 2. The main observation is that these are uniformly scattered between 40 and 120 bpm. The 95% confidence limits, derived by a bootstrap technique, also shown on Figure 2, clearly indicate that variation in optimal tempi from extract to extract is systematic and not due to random sampling of some single underlying value. Each extract has its own optimal tempo, which is significantly different from the optimal tempi of other extracts. To assess the statistical significance of this observation, a global optimal tempo was calculated using all the data together and compared with the individual optimal tempi for the different extracts. Bootstrap analysis shows that the null hypothesis that all optimal tempi are the same can be rejected ($p<0.0001$).

Figure 2: Optimal tempi for each extract with 95% confidence limits in study 1

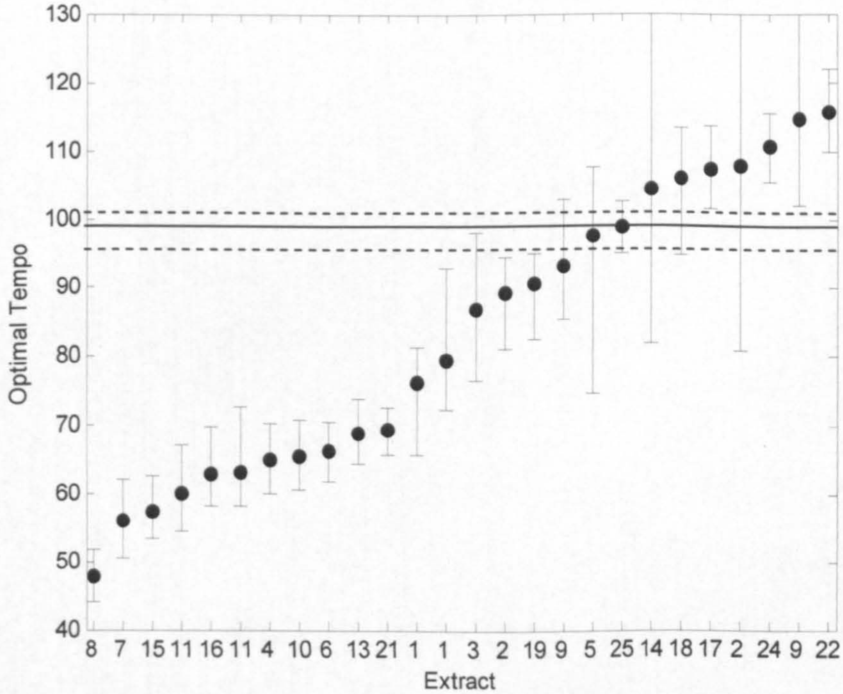

Figure 2: The optimal tempi for each extract with 95% confidence limits on each extract (shown as vertical bars). The extracts are ordered by their optimal tempo. Were there any clustering of values, then these would appear as plateau in the graph, and it can be seen that the optimal tempi are uniformly distributed over a wide range. The conclusion is that each extract has its own characteristic optimal tempo. It also shows the overall global optimal tempo calculated using all the data together shown as a horizontal black

line and the standard deviation shown as dotted lines. Extracts 1, 2, 9 and 11 appear more than once and are taken from study 2. It is clear that extract 1, 2 and 11 turn out to be similar in both studies and are therefore replicable. The optimal tempo for extract 9 is different in the two studies. This is perhaps a result of the task demands in study 2.

It is also clear that extracts 1, 2 and 11 are very similar as the confidence limits overlap. This suggests that these extracts may turn out to have the same optimal tempi. Some of the variability and the different optimal tempi for extract 9 could be explained by the task demands in study 2, as they were required to make a response as soon as they wanted to do so. Hearing the whole extract might allow participants to make more assured responses because they have more information available to them. However, the results seem to suggest that the optimal tempi of these extracts are similar.

Before moving on, we consider the issue of whether each individual participant is making self-consistent judgements about the tempo of each extract. Since each participant made only one response to each extract and tempo combination, the psychometric functions for individual participants are comprised of only 0 and 1 values. If self-consistent, each participant would have a pattern where all the responses for tempi less than some critical value would be 'too slow' and all responses for faster tempi would be 'too fast': ideally there would be no more than one transition in response from 0 to 1 or vice versa as tempo increases. By counting the number of response transitions for each participant and for each extract, it is possible to assess how far each individual is behaving self-consistently. Figure 3 shows the distribution of number of transitions per participant and per extract. It also shows the expected distribution on the null hypothesis that transitions are randomly distributed with 95% confidence limits, as calculated by a bootstrap technique using a repeated random sampling of the responses made by participants. It can be seen that on 71% of occasions there was no more than 1 transition in response, compared with 12% expected by chance ($p<0.001$).

Figure 3: Distribution of the Number of Transitions per Participant and per extract for study 1

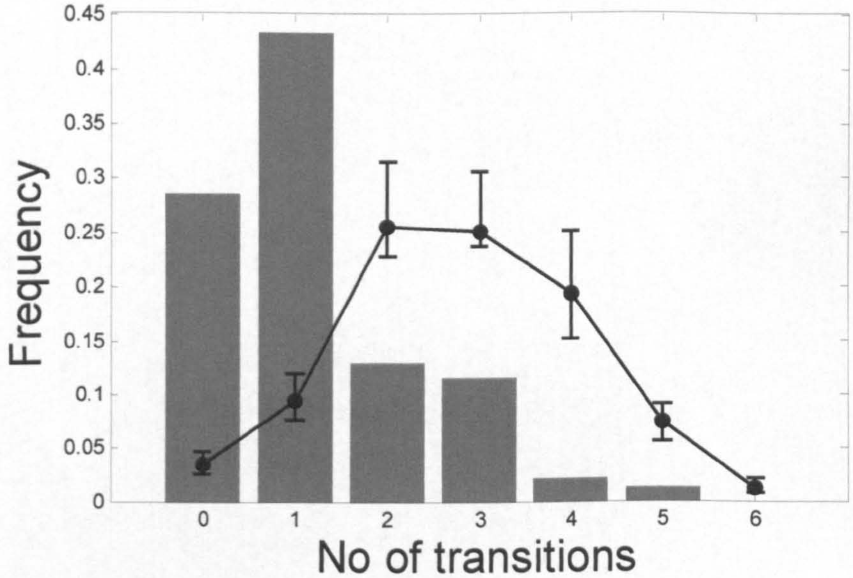

Figure 3: Each participant heard each extract once at each tempo. The psychometric function for an individual participant will switch between 0 and 1 as tempo increases. The number of transitions in either direction is a measure of how consistent the participant is in their response to tempo: one or less transitions would indicate perfect consistency. The figure shows the distribution of the number of transitions in response from "too slow" to "too fast" for each extract and each participant and the expected distribution (with 95% confidence limits) on the null hypothesis that transitions are randomly distributed. On 71% of occasions participants made no more than 1 transition in response (compared with 12% expected by chance ($p<0.001$). The conclusion is that participants are responding in a self-consistent manner.

The second issue is whether different participants show similar optimal tempi for the same extract. For each extract we can estimate the optimal tempo for each participant, by fitting a cumulative normal curve to the individual data. The difference between each individual participant's optimal tempo and the group optimal tempo can be calculated. The distribution of these differences will show how similar participants are: if it is narrow they are similar; if it is wide they are dissimilar. Figure 4 shows the distribution of individual optimal tempi about the mean optimal tempo for each extract for study 1. It also shows the distribution expected on the null hypothesis, with 95% confidence limits, obtained by bootstrap. It can be seen that the spread of individual optimal tempi is narrower than expected by chance ($p<0.01$).

Figure 4: Distribution of individual optimal tempi about the average optimal tempo for Study 1

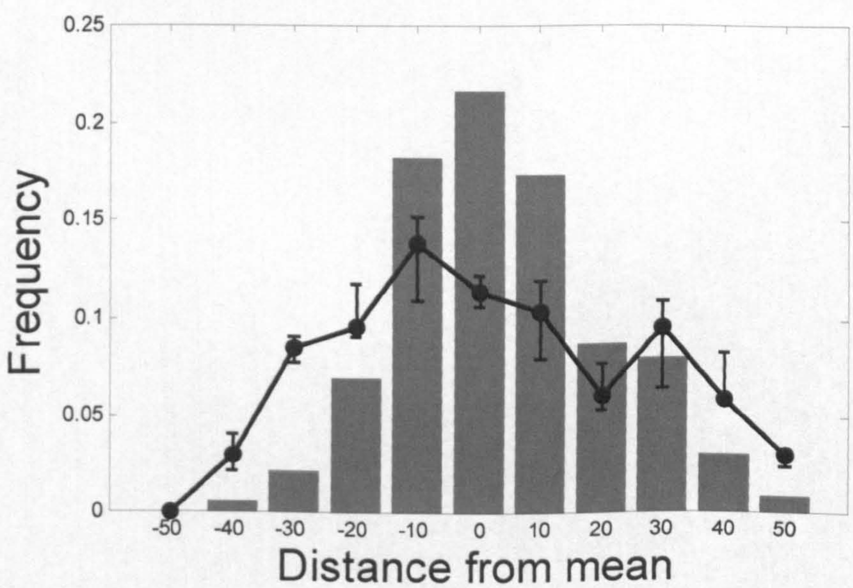

Figure 4: The dispersion of optimal tempo measures for individual participants and individual extracts, about the optimal tempo for each extract (bars). The figure shows the expected distribution on the null hypothesis that individual participants are not sensitive

to differences between extracts (line with confidence limits). The spread of individual tempi is narrower than expected by chance ($p<0.01$).

Study 2:

The second study was a repeat of the first study, subject to the difference that participants could respond at any time after the start of the stimulus. For each extract a different range of test tempi was chosen, varying around the measured optimal tempi of study 1, rather than around the score tempi. Study 2 was carried out principally to establish whether the measured optimal tempi for the various extracts were stable properties that could be replicated with a fresh sample of participants and a slightly different technique. Figure 5 plots the optimal tempo from study 2 against the optimal tempo from study 1. The data are closely correlated, indicating that the measured optimal tempo for an extract can be replicated.

Figure 5: Optimal Tempo for study 2 as a function of Optimal Tempo for Study 1

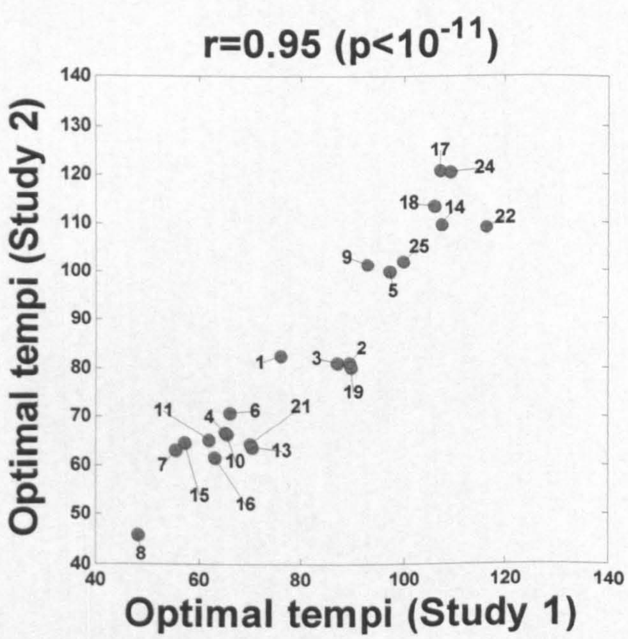

Figure 5: The optimal tempi for study 2 compared to the results of study 1. This shows that the optimal tempi for the same extract from the two studies are similar. The reported p-value relates to the null hypothesis that the two variables are unrelated

Study 3:

A final study repeated the procedure, but with a non-pitched drum beat rather than a pitched instrument, thereby removing the melodic component of the stimuli. The optimal tempi for these stimuli are plotted in Figure 6 against the optimal tempi measured in study 1. As can be seen there is still some relation, but the scatter is larger than in study 2. It is concluded that rhythm plays some part in the judgements of the optimality of tempo, but that the pitch contour is also important.

Figure 6: Optimal tempo in Study 3 as a function of Optimal Tempo in study 1

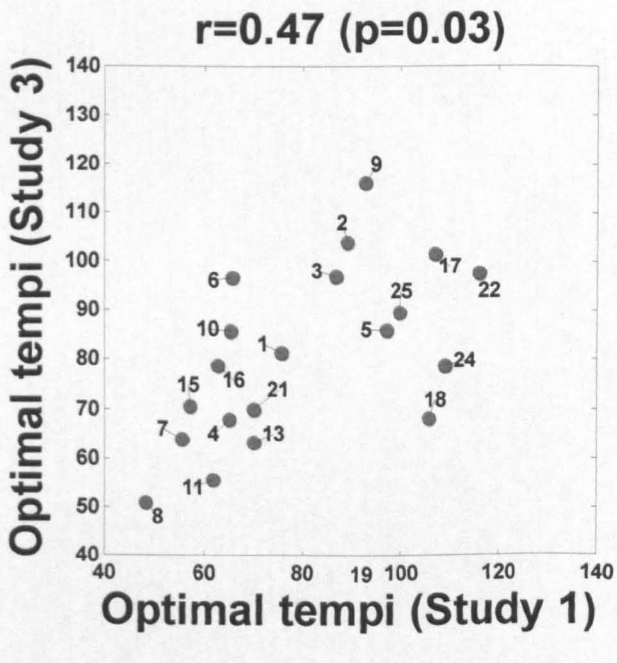

Figure 6: The optimal tempi for study 3 compared to study 1. This shows that the optimal tempi are similar. The scatter is wider than figures 5, showing that rhythm alone can deliver the optimal tempo in some cases but not all. The reported p-value relates to the null hypothesis that the two variables are unrelated

DISCUSSION

It has been shown that untrained (but not un-exposed) listeners can make consistent judgements about the appropriateness of the tempo of pieces of music, and can judge whether a piece of music is played too fast or too slow. The main result is that the judged optimal tempo varies significantly from piece to piece, and this variability is robust across three different studies. This result, although perhaps not entirely surprising in musical terms, has some significant implications.

An important part of the procedure was the randomization of all stimuli so that neither piece nor tempo was blocked. This precaution means that the possibility that listeners were basing their judgements on a comparison between their memory for a previously heard extract with the extract they were responding to at that time is unlikely to explain the pattern of results. The variation of optimal tempi across different extracts showed that listeners were making responses that were independent of the context they heard each extract.

Having repeated the procedure with stimuli from which pitch variations were removed, making rhythm-only stimuli, the basic result is very similar to that obtained with the melody present. The lack of melodic structure does not seem to inhibit listener's abilities to make decisions about the appropriateness of the tempo, in 2/3 of the cases. As previous studies show that ambiguity arises when rhythm and melodic structure are not compatible it is interesting that in absence of a melody, consistent 'too fast' and 'too slow' judgements could still be made.

One possible explanation for these findings might be drawn from a cultural perspective. It could be argued that listeners hold expectations about the specific tempo that Scottish music should be performed at because it is frequently associated with dancing. However, the results showed that the optimal tempo for each extract varied across a wide spectrum. This suggests that the results are not influenced by any cultural preconceptions.

The most likely musicological explanations do not prove useful in explaining these results (see table 1). It could have been expected that the major and minor distinction might have caused the results so that listeners prefer the music to be fast in a major key or slow in a minor key. This is not the case in our studies. Although chords were not played to the listeners explicitly, it is possible that the number of harmonic changes implied by the musical structure may have provided them with a cue to judgements about tempo. A greater number of harmonic changes might cause listeners to judge the music as faster than those with fewer changes. However, the number of chord changes did not correlate with optimal tempo.

Table 1: Optimal Tempo and Musical Features

Musical Features	**Correlations**
Major/Minor	-0.16
Number of different pitches	0.41
Note length Standard Deviation	0.40
Frequency of most common pitches	-0.28
Mean interval direction	-0.37
Mean interval size	0.31
No. ascending intervals	0.37
No. descending intervals	0.49*
No. of chord changes	0.17
* $p < 0.05$	

Table 1: The correlation between optimal tempi in study 1 and musical features. The table shows that almost all the musical features were not correlated with optimal tempo,

except the number of descending intervals. Most musicological features do not account for the optimal tempo for extracts of music.

Boltz (1998) found that the greater the number of changes in pitch contour direction (ascending or descending) and the greater the number and magnitude of pitch changes contained in the music influence the perceived tempo of the music. In contrast to this, our results showed that the number of pitches, the mean interval size and the mean interval direction were not correlated with perceived optimal tempi. However, the number of descending intervals did partially correlate with optimal tempo, although the number of ascending intervals did not.

Lastly, no correlation was found between the variability of note durations (standard deviation of the length of the notes) or the frequency of the most common notes with optimal tempo. Apart from pitch contour, the most likely musicological features do not seem to be involved in the perception of tempo. In absence of any clear explanation we need to turn to a perceptual account in order to explain these results more fully.

Having ruled out most musicological accounts for the main result, we turn to consider possible mechanisms for the formation of a reliable judgement of tempo optimality. In order to assess the optimal tempo of music the listeners might first decide how fast the music is being played: the tempo of the stimulus. A further stage would be necessary as a tempo measurement would only state how fast the music is being played, not whether it is 'too fast' or 'too slow'. In order to make a response, listeners would need to make a comparison between the supposed tempo measurement and some criterion optimal tempo that they consider to be appropriate. The present findings show that the criterion by which they make a response is driven by the stimulus and not internally generated.

The notion that internal devices either in terms of an internal measuring device or clock (Clarke, 1987) or an internal oscillator (Drake, et al., 2000; Large and Jones, 1999; Schulze, 1987) are involved in the experience of temporal aspects of music has proven useful. However, the present study begins to shape an entirely different picture of the usefulness of internal mechanisms for perceiving tempo in music. If internal mechanisms are involved either in terms of a measuring device or an oscillator it would be expected that such a measurement would be highly accurate and precise. Indeed, oscillatory attending is thought to be highly accurate where the musical stimulus has a regular rhythm (Large and Jones, 1999). As our stimuli contained no variations in IOIs or changes in velocity that are produced in expressive performances then it could be argued that oscillatory attending is heightened further. In these circumstances the way in which the listener attends to the music is extremely focused and can accurately predict the occurrence of the underlying temporal regularity.

Such exact predictions or measurements of the stimulus tempo may or may not be a part of the process that delivers a 'too fast' response, but it is not the whole process. If a tempo measurement is made, then some other measurement, indicating the preferred tempo (rather than the actual tempo) also needs to be made, in parallel, from the same stimulus. This is possible, but we would like to explore a different approach: that the tempo of the stimulus is measured, not in absolute terms as events per unit time from which 'too slow' or 'too fast' is derived, but instead is measured directly as being 'too slow' or 'too fast'. We wish to consider the possibility that tempo relative to optimum is perceived directly rather like a higher order invariant (Gibson, 1966).

This direct perception of tempo relative to optimal could arise because the events were occurring faster or slower than predicted by some internal device, but there are no grounds for supposing that such devices would make predictions that differed from the actual timing of the stimulus events. An alternative is that the nature of events (such as individual pitches) coupled with their temporal sequence, are used to make predictions about the timing of the next event. Where the event nature and timing are incompatible the tempo is perceived to be suboptimal.

Our results indicate that the perception of the appropriateness of the tempo for an extract of music is determined by the contents of the extract itself. We have not been able to identify any simple musical correlate of this and instead turn to consider perceptual features of the music. It might be thought that variations in the speed with which the different musical extracts can be processed sets different optimal tempi, but there is no evidence of a failure to process the stimuli. Furthermore, the stimuli, being simple tonal and monophonic, are very much simpler than much music that people encounter.

Suppose that music has events that vary in their character, possibly along several simultaneous dimensions. The perceptual effect of the music will be due to the time-varying nature of this character but also to the speed with which the character varies. To say this is to say little, except that it then opens up the possibility that there is an important perceptual interaction between event character and event duration. The overall effect of a piece of music depends on getting the tempo right for the character of the particular events in the music. This can be illustrated by supposing that events in musical stimuli vary in strength, and that strong events lose effectiveness by being rushed.

It is then reasonable to suggest that a piece of music that is played too quickly and is filled with strong events would be perceived by the listener as 'too fast'. Likewise, a piece of music that is played slowly and is filled with weak events would be perceived as 'too slow'. The strength of events within a piece of music lead the listener to determine what tempi is appropriate. The relative strengths of these events may require them to unfold more or less quickly in order to construct a perceptual representation of them that makes sense to the listener. These assumptions are speculative and necessitate further research in order to establish what event strength is and its relationship with perceived tempo in music.

CHAPTER 4: IS OPTIMAL TEMPO DEPENDENT UPON CONTEXT

INTRODUCTION

The previous chapter showed that optimal tempi varied from extract to extract. This variation of optimal tempi across different extracts suggested that listeners were making responses that were independent of the context within which they heard each extract. Psychophysical methods used in the previous chapters (like most methods in psychology) present participants with numerous trials and require them to make responses on each. We argued that the randomization of all stimuli meant that listeners were not basing their 'too fast/slow' judgements on a comparison between responses for a previously heard extract with the extract they were responding to at that time. Although this is most likely we wanted to assess this more carefully to eliminate any likelihood that previous responses were contaminating future responses.

One possible way of testing whether this is the case is to present just one trial to each participant. In making only one response any possibility of the previous extract or extracts biasing participants responses would be eliminated. If the pooled responses of many participants produced results similar to those in the previous chapter, then this would show that optimal tempo is not dependent upon context. This would show that 'too fast/slow' judgements are made on the basis of features of the individual stimuli rather than context.

The purpose of the present paper is to test this by presenting the extracts from the previous study. Each participant listened to one of the extracts at one level of the cue and was asked to judge whether it was 'too fast' or 'too slow'.

METHODS

Participants

Participants were visitors at the Glasgow Science Centre. 644 participants took part in the study. No attempt was made to either select or filter out musically trained subjects.

Materials

The materials were exactly the same as those described in the previous chapter except that the tempi were set to variations around the optimal tempi found in studies 1 and 2 (±0, 10, 20, 30 beats per minute).

Procedure

Participants were told that they would be taking part in a short experiment where they would be asked to listen to a piece of music and make one of two responses on a keyboard. One stimulus was played to each participant and after hearing it they were asked to make a response by indicating whether they thought it was played 'too fast' or 'too slow' by pressing designated keys on a computer keyboard. They were instructed to make just one of the two responses, even if they were unsure about their response. Four responses were obtained for each extract at each level of the cue, therefore requiring 644 participants.

RESULTS

The data were collected into psychometric functions for each extract, giving the proportion of participants responding 'too fast' as a function of the tempo of the music. For each extract, the proportion rises as tempo increases; in most cases the psychometric functions produced a full range of responses from 0 'too slow' to 1 'too fast'. Each psychometric function was analysed by finding the least squares fitting cumulative normal function using a simplex minimization routine (this is equivalent to the pre-computer technique of probit analysis). The curve is then represented by 2 parameters, the PSE (the point at which the function crosses 50%) and the SD (inversely proportional to the slope of the curve). The PSE is the point at which the responses 'too fast' and 'too slow' are equally likely, and is thus the point at which the tempo is perceived as optimal by the participants.

Three sample psychometric functions are shown in Figure 7, 8 and 9 (the dotted line in each figure shows the psychometric functions for the present study). The psychometric function for extract 22 (see figure 9) is very similar in both studies. Although the curves for extract 1 and 8 are different there are two important points to take into consideration. The first is that extract 1 has a similar PSE value in both studies. Secondly, the best fit of the curve for extract 8 in this study may actually be closer to study 1. A fuller range of responses at each level of the cue may have contributed to the slope of the curve. Having a larger range of responses at each level of the cue may have produced a much better fit and therefore similar slope of the function for the present study and study 1. It is clear that listeners are capable of making 'too fast' and 'too slow' responses that collectively produce s-shaped psychometric functions.

Figure 7: The proportion of 'too fast' responses as a function of tempo for extract 1 in study 1 and study 4

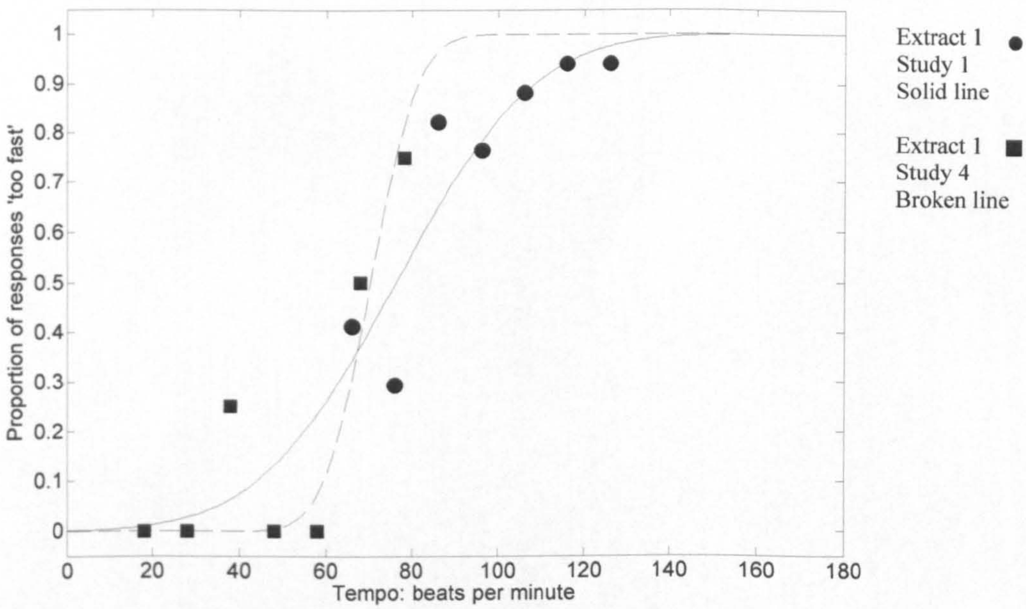

Figure 7: The proportion of 'too fast' responses as a function of tempo as beats per minute for extract 1 in study 1 and study 4. The symbols are data points; the smooth curves are least squares fits to the data. The circles are the data points from study 1 for extract 1 with a solid psychometric function and the squares are the data points for extract 1 from study 4 with a psychometric function represented by a dotted line. The psychometric functions rise as tempo increases and produce a full range of responses from 0 'too slow' to 1 'too fast'. The optimal tempi can be obtained at the point where the psychometric curves cross 0.5 and produced a tempi of approximately 70 b.p.m.

Figure 8: The proportion of 'too fast' responses as a function of tempo for extract 8 in study 1 and study 4

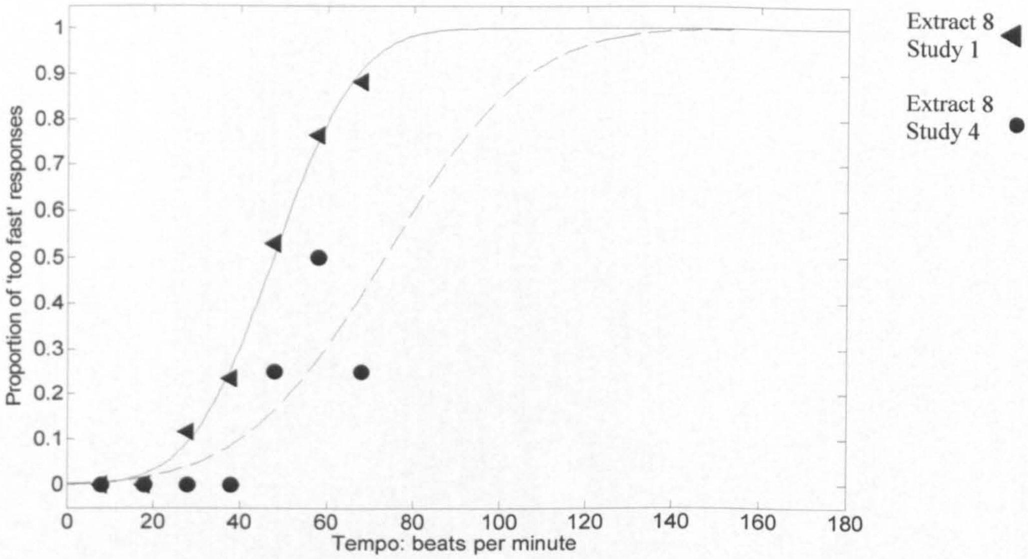

Figure 8: The proportion of 'too fast' responses as a function of tempo as beats per minute for extract 8 in study 1 and study 4. The symbols are data points; the smooth curves are least squares fits to the data. The triangles represent the data points from study 1 for extract 8 with a solid psychometric function and the circles are the data points for extract 8 from study 4 with a dotted psychometric function. The psychometric functions rise as tempo increases and produce a full range of responses from 0 'too slow' to 1 'too fast'. The optimal tempi can be obtained at the point where the psychometric curves cross 0.5 and produced a tempi from approximately 50 b.p.m. to 70 b.p.m.

Figure 9: Proportion of 'too fast' responses as a function of tempo for extract 22 in study 1 and study 4

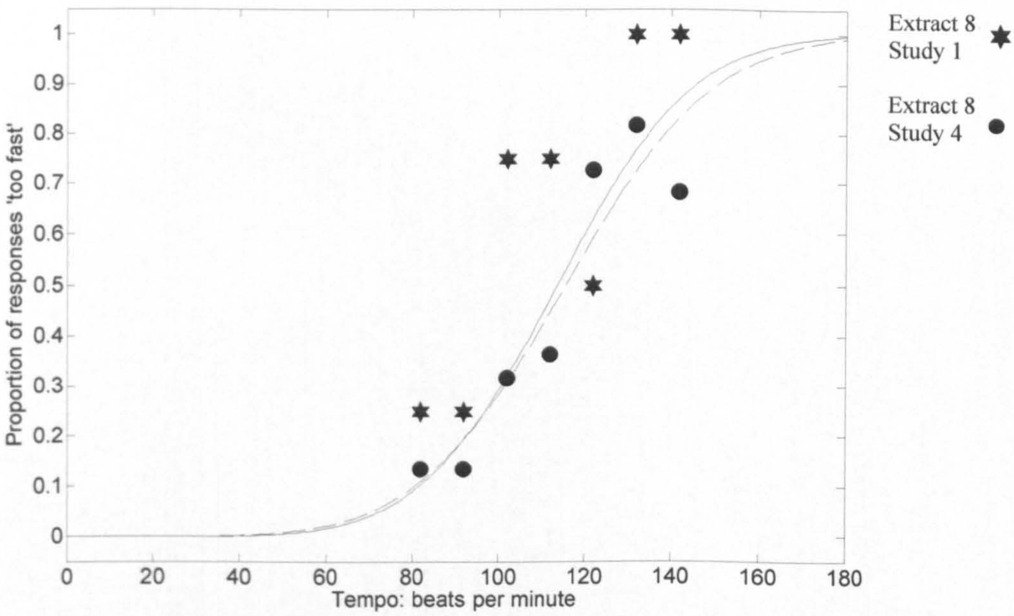

Figure 9: The proportion of 'too fast' responses as a function of tempo as beats per minute for extract 22 in study 1 and study 4. The symbols are data points; the smooth curves are least squares fits to the data. The stars are the data points from study 1 for extract 22 with a solid psychometric function and the circles are the data points for extract 22 from study 4 with a psychometric function represented by a dotted line. The psychometric functions rise as tempo increases and produce a full range of responses from 0 'too slow' to 1 'too fast'. The optimal tempi can be obtained at the point where the psychometric curves cross 0.5 and produced a tempi of approximately 110 b.p.m.

The final issue addresses the relationship between the optimal tempi in study 1 with the present study. Establishing the similarity between them would demonstrate that listeners are capable of making responses that are independent of the context. Figure 10 plots the optimal tempo for study 1 as a function of study 4. The data are closely correlated showing that the measure of optimal tempo can be replicated without any effect of the context.

Figure 10: Optimal Tempo per extract for study 1 as a function of Optimal Tempo for study 4

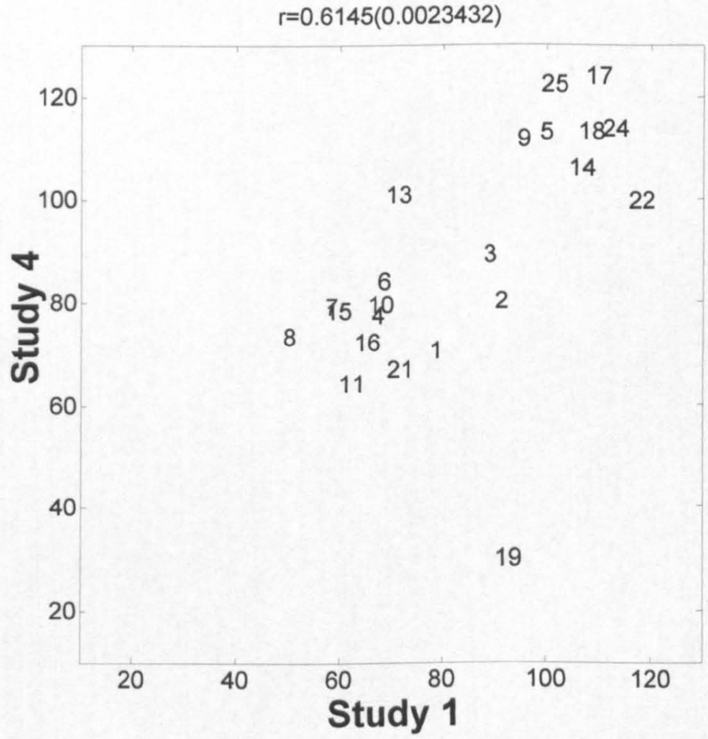

Figure 10: The optimal tempi for study 4 compared to study 1 for each extract. This shows that the optimal tempi for the same extract from the two studies are similar.

DISCUSSION

The main issue was to assess whether the listeners were capable of making responses on single trials. When the responses from all of the participants are pooled it was of interest to assess whether they produced s-shape curves. This would show that the listeners were making responses that are independent of the context and driven by the stimulus. The results produced s-shaped curves that suggest that the listeners were capable of making responses based upon the characteristics of the stimulus rather than information produced from previously heard extracts.

Another important observation is that the optimal tempi for both studies are similar. This also supports the notion that the listeners are capable of perceiving the optimal tempo without the context. Certainly it suggests that when the listeners are making their responses they do so by comparing the tempi the music is being played at with an optimal tempo in mind. In some way it suggests that having heard a piece of music they have some sense of the music dragging or rushing beyond its appropriate speed. The ability to make a 'too fast/slow' response is therefore based on features of it that are not delivered by a comparison with other such pieces of music.

A larger number of responses might produce a more reliable data set. In order to make an exact comparison between study 1 and this study another 2,576 participants would be required. This would be the equivalent of getting the same number of responses as the previous chapter. The present results are promising and support the usefulness of the method of constant stimuli.

Therefore, it seems that listeners can make responses that are dependent just upon the stimulus being presented. Listeners are not basing their responses on previously heard extracts and are using the features of the stimulus to make judgements about its tempo.

CHAPTER 5: THE PERCEPTION OF HAPPY AND SAD EMOTIONS IN MUSIC

ABSTRACT

It is believed that tempo is one of the most important musical features involved in the perception of happy and sad emotions. The present study was conducted to assess the effect of changing the tempo of music on the perception of happy and sadness. Listeners heard short extracts of music that varied in tempo and were asked to make a 2AFC of happy or sad for each extract. Separate psychometric functions were obtained for each extract of music, and the points where these crossed 83% and 17% happy were calculated, and treated as happy tempo and sad tempo respectively. The results show that most extracts can be perceived as both happy and sad just by varying the tempo. However, the tempo at which extracts become happy or sad varies widely from extract to extract. We show that the sad and happy tempi are related to the size of the intervals (pitch changes) in the extract.

INTRODUCTION

Research into emotion and music has attempted to understand how we identify emotions when we perform or listen to music. Some of the earliest work in the area attempted to use adjectives to describe emotional experience (Hevner, 1935; Farnsworth, 1954). Whilst recognising that emotional experience can be described in this way it is important to acknowledge the distinction between direct affective experience and indirect perceptual experience when we use these techniques. The first of these processes involves some change in an individual's internal emotional state whilst the latter refers to the listener simply perceiving the presence of emotion. It is the latter of these two experiences that this paper will focus on. Of interest is the change in the emotion perceived by the listener when we alter the tempo of the music. As tempo is thought to be the most important feature that delivers the emotional expression of music (Hevner, 1937; Juslin, 1997; Rigg, 1939) it is crucial to understand whether perceived emotion might change when the tempo is changed.

Chapter 3 showed that even short extracts of music have an optimal tempo such that performances on either side are reliably heard as 'too fast' or 'too slow'. This optimal tempo varies from piece to piece over a wide range of tempi. This finding is significant in the present context since it makes the issue of tempo relative, rather than absolute. So tempi at which music sounds sad could be absolutely slow, or could be slow compared to optimal tempi. It was argued that the optimal tempo for a piece of music must be derived from the details of the piece itself. If a piece of music can sound 'too fast' or 'too slow', then this may interact with it sounding happy or sad.

Studies that have assessed the role of tempo and emotion have explored: the influences of culture (Balkwill, Thompson, Matsunaga, 2004; Balkwill and Thompson, 1999); identification of emotions expressed during performance (Juslin, 1997; Gabrielsson and Juslin, 1996), emotional ratings to musical stimuli after brain damage (Peretz, Gagnon, and Bouchard, 1998), child development and emotional responses to music (Dalla Bella, Peretz, Rousseau and Gosselin, 2001) and the effect of music on mood and arousal in spatial tasks (Husain, Thompson and Schellenberg, 2002).

Various attempts have been made to assess the structural features of the music that might play a role in the experience of emotion. A full study of these features of music was done by Hevner (1935a, 1935b, 1936, 1937) that assessed mode, melodic contour, harmony, rhythm, tempo and pitch. After hearing pieces of music listeners were required to select as many adjectives as they wished from a list assembled into eight clusters to describe the emotional meaning being expressed. Each cluster contained adjectives that had a similar meaning. The clusters were arranged in a circle so that each one had a set of descriptors on the opposite side that contained the opposite meaning to that set. The results suggest that the strongest influences overall on listeners judgements were tempo and mode. Music that was judged to be sad was heavily influenced by the minor mode, low pitch and slow tempo with happy music being influenced more by the major mode and fast tempo. Hevner suggests that we should be cautious over the results since any manipulations made by the performer when they played the music to participants may have sounded unmusical: the performer may have found it difficult to play the modified versions of the music because it sounded unnatural by comparison to the original. Furthermore, it is suggested that the use of music that held a constant emotional expression throughout the piece would be necessary because it would ensure that listeners would choose descriptors that contained a similar meaning. Presumably this would allow a direct assessment of the structural features that are influential in happy music as opposed to sad music.

Riggs (1940) reports further research that focused on short extracts composed by the experimenter that were thought to express happiness and sadness. Like the Hevner study each composition (five in total) was played at 6 different tempi from 60 bpm to 160 bpm and was performed on the piano. Listeners were required to decide whether the music was pleasant/happy or sad/serious and then choose descriptors from the lists contained within subcategories. It was found that faster tempi were likely to provide more pleasant/happy than serious/sad judgements. Likewise, slower tempi were likely to produce more sad/serious judgements. This is not entirely clear cut as listeners did judge two of the pieces of music to be sad at the fastest tempo it was presented at. Arguably there are other musical features that might have played a role. However, since an analysis of the musical features involved was not conducted it makes it difficult to assess what might be responsible for the results obtained. Making use of a larger range of tempi may also have shown a greater degree of variation in perceived happy and sad judgements. Had this been the case it would have shown that perceived happy and sad judgements can be made by altering the tempo of the music.

Kamentsky, Hill and Trehub (1997) conducted a study to assess the effects of various parameters on the perception of emotion in music. Participants listened to 4 pieces of classical music performed by a musician that were manipulated by sequencing software. The study contained 4 conditions: changes in dynamics, changes in tempo, no variations in either tempo or dynamics, and changes to both tempo and dynamics. To ensure that the music did not sound artificial, all of the changes were rated by 3 trained musicians who agreed that they sounded appropriate. Listeners were then asked to rate each piece of music on a 7-point scale for emotional expressiveness, how much they liked it, the degree of variation in tempo, and familiarity. The results suggested only a partial effect of tempo and dynamics on the expressiveness of the music. They suppose that this might be due to the listeners lack of formal training or that they may not be well versed in the cultural affects of the musical genre used. Both of these arguments seem unlikely since these types of manipulations are a feature of all musical genres and even non-musicians would have been well used to hearing them. It could be argued that local variations in tempo have little effect upon the expressive character. Additionally, the emotion the music was expressing to the listener was not assessed which makes it difficult to really interpret what emotional experience they had. In some cases music may not express anything in particular to the listener which could explain why the effect of tempo was only small. In order to change the character of music the global tempo may be the only way to alter the expressive quality of a piece. Minor changes in tempo that are heard during musical performance may only aid the listening experience by making the music sound more natural, but not cause changes in the perception of emotion in music.

A number of different approaches have been used to assess our experience of emotion in music. Whilst these techniques offer insight into our experience they can be problematic in their approach. Providing the participant with a variety of adjectives to choose from can lead to confusion over the listeners experience. If the listener chooses multiple adjectives in response to the music that have different meanings but underlying similarities (i.e. they may describe the music as both exciting and fearful) it could leave the experimenter with a dilemma about what emotion they were actually experiencing or what they believed was being expressed. Indeed, asking participants to choose from lists of this sort can create a situation where the listener has to continuously shift their attention through the numerous possibilities (Schubert, 2001). In doing so, the task demands could impact on their responses because of increasing levels of cognitive workload placed upon the listener.

One study that attempts to deal with this issue assessed the interactive effect of rhythm and pitch on ratings of happy, sad and scary emotions (Schellenberg, Krysciak and Campbell, 2000). The initial phase of the study required participants to make a single judgement about a collection of short melodies and decide which of the three emotions they would use to describe its character. These descriptors were chosen specifically because they are often used most often to describe emotion in music. Those pieces with the most consistent responses were chosen and a total of 6 pieces were selected that include 2 pieces from each emotional character (2 happy, 2 sad and 2 scary pieces). There were four conditions that were used to assess the effects of rhythm and pitch on ratings of emotion (original; rhythm removed; pitch removed; and pitch and rhythm removed). They found that for the happy pieces the original extracts were happiest; the rhythm removed were next happiest; the pitch removed was less happy; and the pitch and rhythm removed were least happy. For sad pieces the original extracts and those with the rhythm removed were saddest. The results also suggested that pitch and rhythm were important and had differential effects on ratings of happiness, but only manipulations in pitch had an effect over ratings of sadness, at least for these pieces.

Although this begins to provide some indication of what features might be important in making these judgements it is not clear whether these observations would transmit across other pieces of music. Pre-selecting those melodies with the most extreme emotion may have also pre-selected uncharacteristic melodies and a broader spread might have given different results. The perceived emotion in those extreme pieces might be over-constrained – i.e. concurrently indicated by many different musical cues. By removing any one or two structural cues the remaining musical features would still be intact. It would have been more useful to consider pieces where the emotion was less clear. Furthermore, an analysis of the detailed pitch structures and rhythmic structures might have shed light on the differences they found. Further studies should assess musical features in the first instance in order to provide a clearer interpretation of the structural characteristics that are responsible for sad and happy judgements.

Scherer and Oshinsky (1977) conducted a study that used a number of tone sequences that manipulated the pitch, contour, volume, tempo, tonality and rhythm through a synthesizer and therefore removed any expressive variations provided by the performer. Participants were asked to rate each tone sequence for emotionality. Interestingly, the results suggest that the tempo manipulations were the most powerful predictor for the emotion ratings. This was particularly the case for emotions such as sadness and happiness, although other features also played an important role in emotional ratings. For happy pieces of music, features such as a fast tempo and large pitch variations were important. Sequences that were rated as sad were strongly influenced by a slow tempo, low pitches, few harmonics and descending pitches. In terms of the effects of tonality, the major mode was most influential for music that was rated as happy and the minor mode was most influential for pieces that were rated as sad. Rhythm did not seem to be important for ratings of happiness and sadness. However, these features were not mutually exclusive and did play a role in emotional ratings other than happiness and sadness.

These studies highlight the role of particular musical features in making ratings on emotionality. However, there are a number of issues that limit their potential impact. In most studies the manipulation of tempo is limited to 2 values (slow or fast) an observation that has been recognised by other researchers (Gabrielsson and Lindström, 2001). Since these studies do not indicate what tempo the music was played at and only include 2 variations it is difficult to assess the real impact of tempo on happy and sad emotions. The tempi that are assumed to be slow or fast may not be fast enough or slow enough for the individual pieces of music used. Chapter 3 indicates that there are tempi for pieces of music that are perceived to be optimal. They also indicate what tempi the music would have to be in order to be 'too fast' or 'too slow' for a piece of music and show that these vary widely from piece to piece. They point out that one melody would be 'too slow' at one tempi and another at that same tempi would be 'too fast'. To assess the influence of fast and slow tempi on emotional ratings it is important to start out by using a large range of tempo variations to ensure that the fast and slow tempi really are fast and slow enough for that particular piece of music.

There is ample evidence that tempo correlates with happiness and sadness, but the nature of the relationship is not yet clear. We have broken this down into 3 questions:

1) Can the same melody be both happy and sad, depending solely on tempo?

2) Does the tempo at which a melody becomes happy or sad vary from extract to extract?

3) Do any identifiable musical features interact with the effect of tempo on perceived emotion?

METHODS

Participants

Participants were undergraduate students from the University of Stirling. 30 participants took part in the study. Most were Psychology students; none were music students. No attempt was made to either select or filter out musically trained subjects: the incidence of musical training in the sample was around 2%, and probably reflects the population incidence generally.

Materials

The materials were the same as those in chapter 3 (study 1 and 2) but the stimuli were played at 16 different tempi from 10 bpm to 160 bpm in increments of 10 bpm.

Procedure

The procedure was the same as the one used in study 1 and 2, but in this study participants were asked to indicate whether they thought the extracts were happy or sad by pressing designated keys on a computer keyboard.

Design

The experimental design was the same as the one described in Chapter 3 except that each melody was played at sixteen different tempi making a total of 368 trials.

RESULTS

The data were collected into psychometric functions for each extract, giving the proportion of participants responding 'happy' as a function of the tempo of the music. Three sample psychometric functions and the corresponding data are shown in Figure 11. For each extract, the proportion rises as tempo increases; in all but 2 cases the psychometric functions produced a full range of responses from 0 'sad' to 1 'happy'. The responses for extract 1 and 10 did not provide a full range of responses, (shown in Figure 12). For these 2 extracts the responses reach around 0.5 at 140bpm and then level off for both curves. Therefore, these pieces seem to be predominantly sad up to neutral. Each psychometric function was analysed by finding the maximum likelihood fitting cumulative normal function using a simplex minimization routine (this is equivalent to the pre-computer technique of probit analysis). The curve is then represented by 2 parameters, the PSE (the point at which the function crosses 50%) and the SD (inversely proportional to the slope of the curve). The PSE is the point at which the responses 'happy' and 'sad' are equally likely, and is thus the point at which the tempo is perceived as neutral by the participants.

Figure 11: Proportion of 'happy' responses as a function of tempo

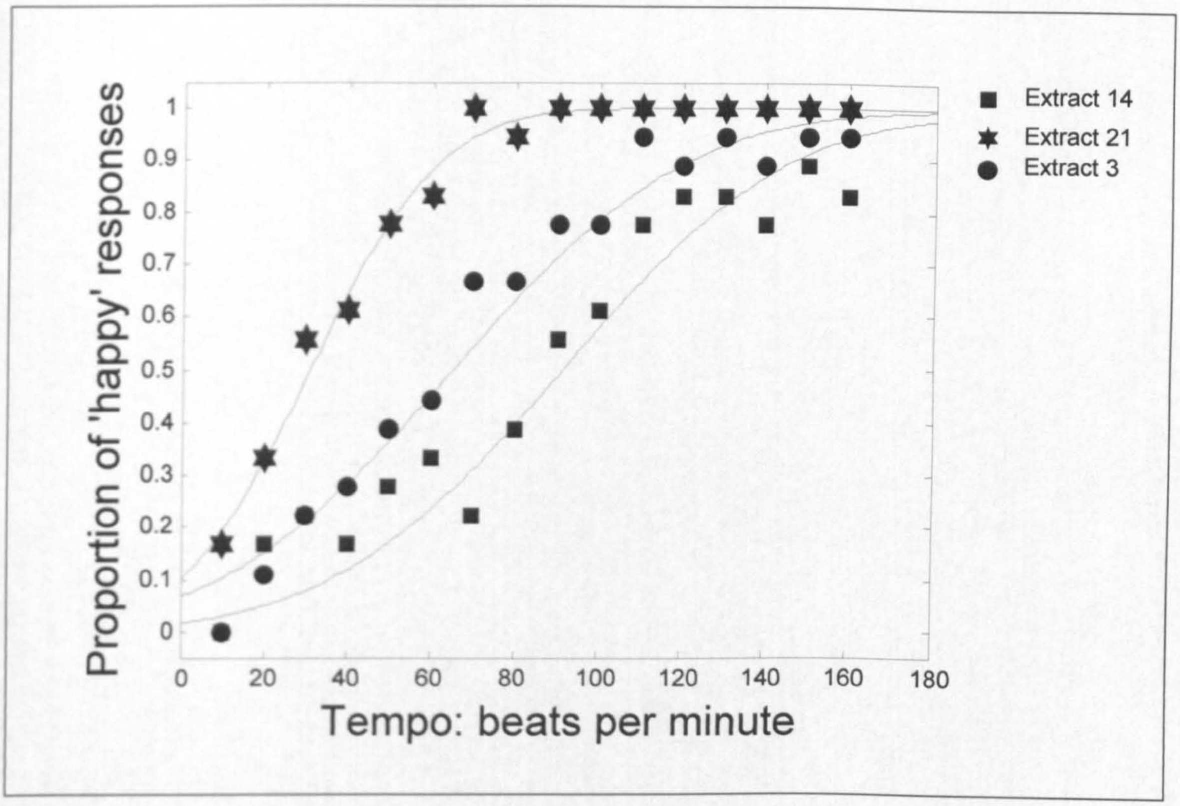

Figure 11: The proportion of 'happy' responses as a function of tempo as beats per minute for extracts 14, 21 and 3. The symbols are data points; the smooth curves are least squares fits to the data. The psychometric functions rise as tempo increases and participants produced a full range of responses from 0 'sad' to 1 'happy'. The happy tempo is the point where the curve crosses 83% and the sad tempo is the point where the curve crosses 17%. The happy tempo for extract 21 is 50 bpm where the curve crosses 83% and 5bpm where the curve crosses 17% for the sad tempo.

Figure 12: Proportion of 'happy' responses as a function of tempo for extract 1 and 10

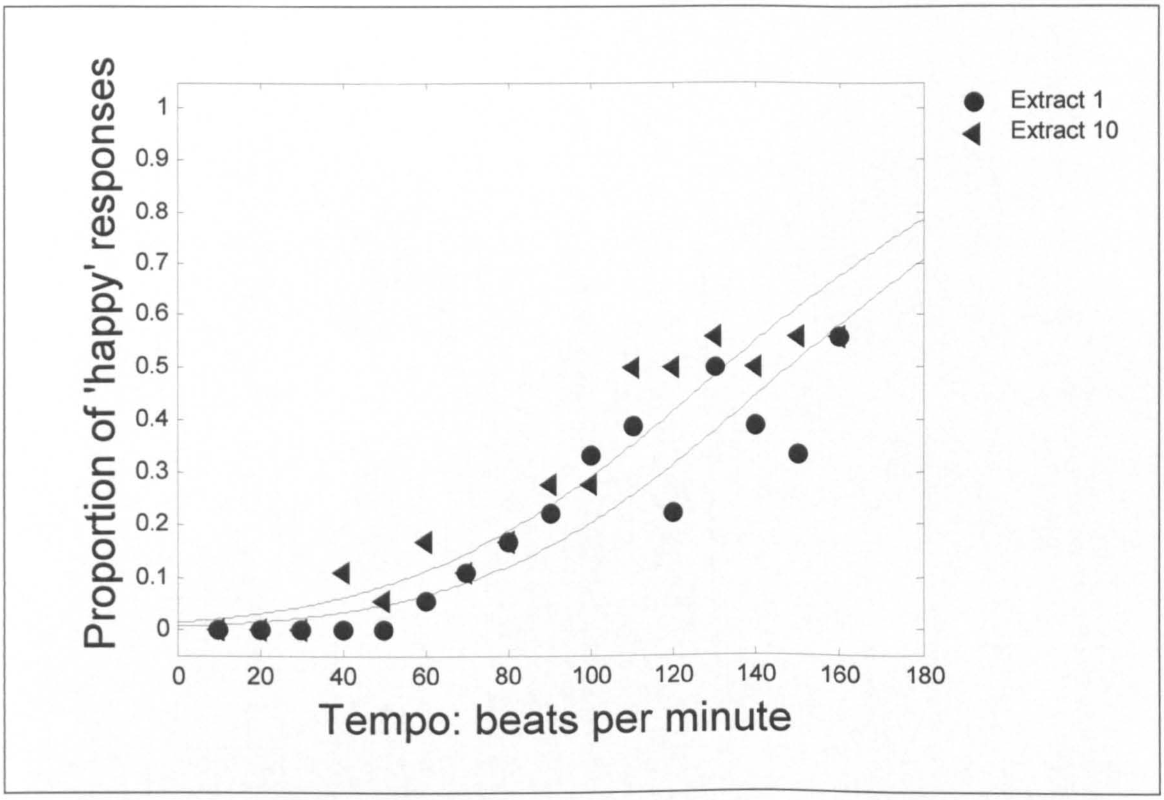

Figure 12: The proportion of 'happy' responses as a function of tempo as beats per minute for extracts 1 and 10. The symbols are data points; the smooth curves are least squares fits to the data. The psychometric functions rise as tempo increases, but the range of responses from 0 'sad' to 1 'happy' plateau at 0.5. These extracts are only sad at a tempo of 130 bpm for extract 10 and 140 bpm for extract 1.

We are interested in the tempo at which an extract is reliably happy and reliably sad. We define the tempo at which the curves cross 83% as the happy tempo and the 17% as the sad tempo. By choosing the 83% point and the 17% point (one standard deviation from the point of subjective equality) we are assuring that the responses will be significantly different from neutral. The happy tempo is the slowest tempo at which an extract sounds happy. The sad tempo is the fastest tempo at which an extract sounds sad. Figure 12 shows that the happy tempo for extract 21 is 50 bpm where the curve crosses 83% and that the sad tempo is 5 bpm where the curve crosses 17%.

From this point on the happy and sad tempo relates to the number of notes played per minute rather than tempo as bpm. There are a number of reasons why tempo expressed in this study as crotchets (quarter notes) per minute, is not entirely satisfactory as a measure of how quickly music is played. For example, the exact same performance can be notated as quarter notes, at a tempo of 50 bpm, or notated as eighth notes at a tempo of 25 bpm. Moreover, several pieces in our selection had a basic pattern of eighth notes grouped in triplets (musically 6/8 time signatures). For these, the basic perceived pulse is 1.5 quarter notes long. In this case we have adopted a less ambiguous measure of tempo: notes per minute (npm) to resolve this issue.

Figure 13 shows the values obtained for the happy tempi for each extract. The main observation is that these are scattered between 90 and 200 npm. Extracts 1 and 10 produce large error bars and are not represented on the graph because the participants did not produce a full range of responses from 0 to 1. The happy tempo for these extracts, are higher than is represented on the graph. The 95% confidence limits, derived by a bootstrap technique, also shown on Figure 13, clearly indicate that variation in happy tempi from extract to extract is systematic and not due to random sampling of some single underlying value. Each extract has its own happy tempo, which is significantly different from the happy tempi of other extracts. To assess the statistical significance of this observation, a single happy tempo was calculated using all the data from all extracts together and compared with the individual happy tempi for the different extracts.

Bootstrap analysis shows that the null hypothesis that all happy tempi are the same can be rejected (p<0.000001).

Figure 13: Happy tempo for each extract with 95% confidence limits

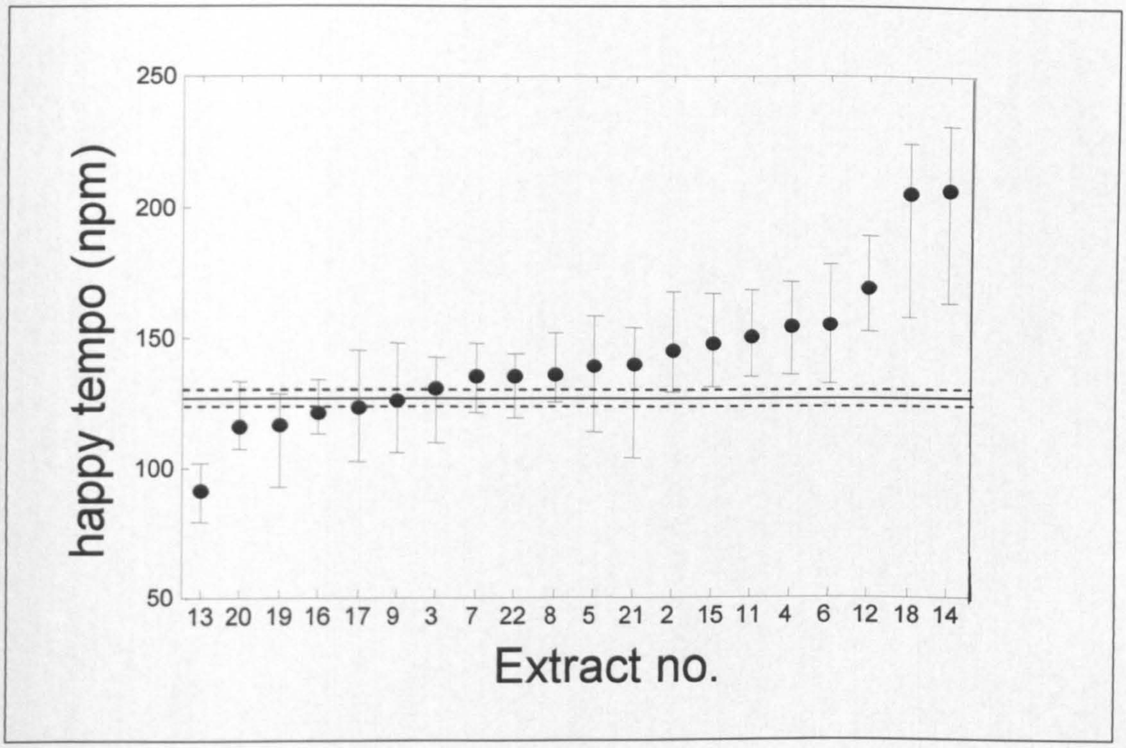

Figure 13: The happy tempi (npm) with 95% confidence limits for each extract (shown as vertical bars). The extracts are ordered by their happy tempo. It also shows the overall happy tempo calculated from the 83% point on the psychometric function created by pulling all of the data together. This is shown as a horizontal black line and with the standard deviation shown as dotted lines. Extracts 1 and 10 produce large error bars and are not represented on the graph because the participants did not produce a full range of responses from 0 to 1. Were there any clustering of values, then these would appear as plateau in the graph, and it can be seen that the happy tempi are uniformly distributed over a wide range. The conclusion is that each extract has its own characteristic happy tempo.

Figure 14 shows the sad tempi for each extract with 95% confidence limits. The main observation is that these are scattered between 10 npm and 140 npm. The 95% confidence limits, derived by a bootstrap technique, also shown on Figure 14, clearly indicate that variation in sad tempi from extract to extract is systematic and not due to random sampling of some single underlying value. Each extract has its own sad tempo, which is significantly different from the sad tempi of other extracts. To assess the statistical significance of this observation, a single sad tempo was calculated using all the data from all extracts together and compared with the individual sad tempi for the different extracts. Bootstrap analysis shows that the null hypothesis that all sad tempi are the same can be rejected ($p<0.0001$).

Figure 14: Sad tempo for each extract with 95% confidence limits

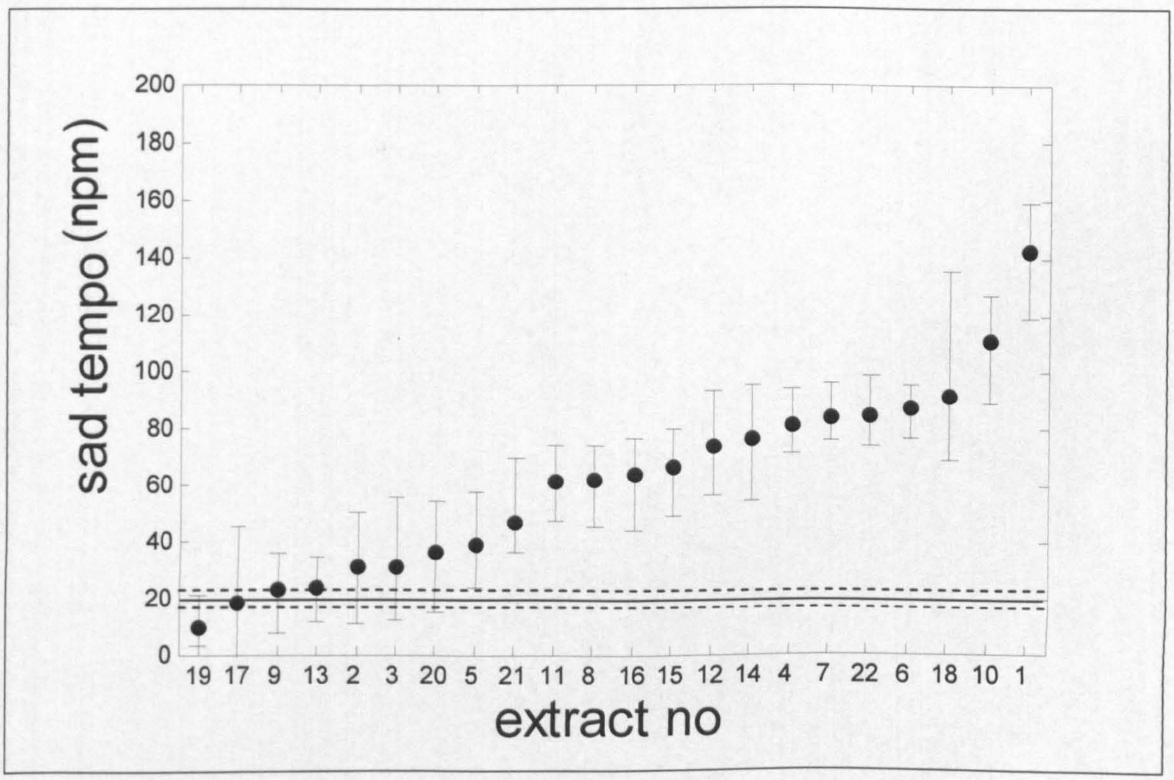

Figure 14: The sad tempi (npm) with 95% confidence limits for each extract (shown as vertical bars). The extracts are ordered by their sad tempo. It also shows the overall sad

tempo calculated from the 17% point on the psychometric function created by pulling all of the data together. This is shown as a horizontal black line and with the standard deviation shown as dotted lines. Were there any clustering of values, then these would appear as plateau in the graph, and it can be seen that the sad tempi are uniformly distributed over a wide range. The conclusion is that each extract has its own characteristic sad tempo.

Relationship to Optimal Tempi

Chapter 3 measured optimal tempi for the same extracts used in the present study. There is no correlation between the present happy tempi and those optimal tempi ($r = 0.07$) or sad tempi and the optimal ($r = -0.33$).

The probability of getting a happy response at the optimal tempo for each extract can be obtained from the psychometric functions for happy responses. Figure 15 shows the probability of a happy response for extracts played at their optimal tempi, with 95% confidence limits. The spread of values suggests strongly three clusters:

1) extracts 1, 10, 6 and 15 have very low probability of happy responses at the optimal tempo (i.e. high probabilities of sad responses). These sound sad when played at their optimal tempo.

2) extracts 17, 25, 18, 21, 22 and 24 have very high probability of happy responses at the optimal tempo. These sound happy when played at the optimal tempo.

3) extracts 8, 9, 7, 16, 2 and 4 that are neither happy nor sad at the optimal tempo and therefore neutral. These have no emotional character when played at the optimal tempo.

Figure 15: The probability of a happy response at the optimal tempo for each extract

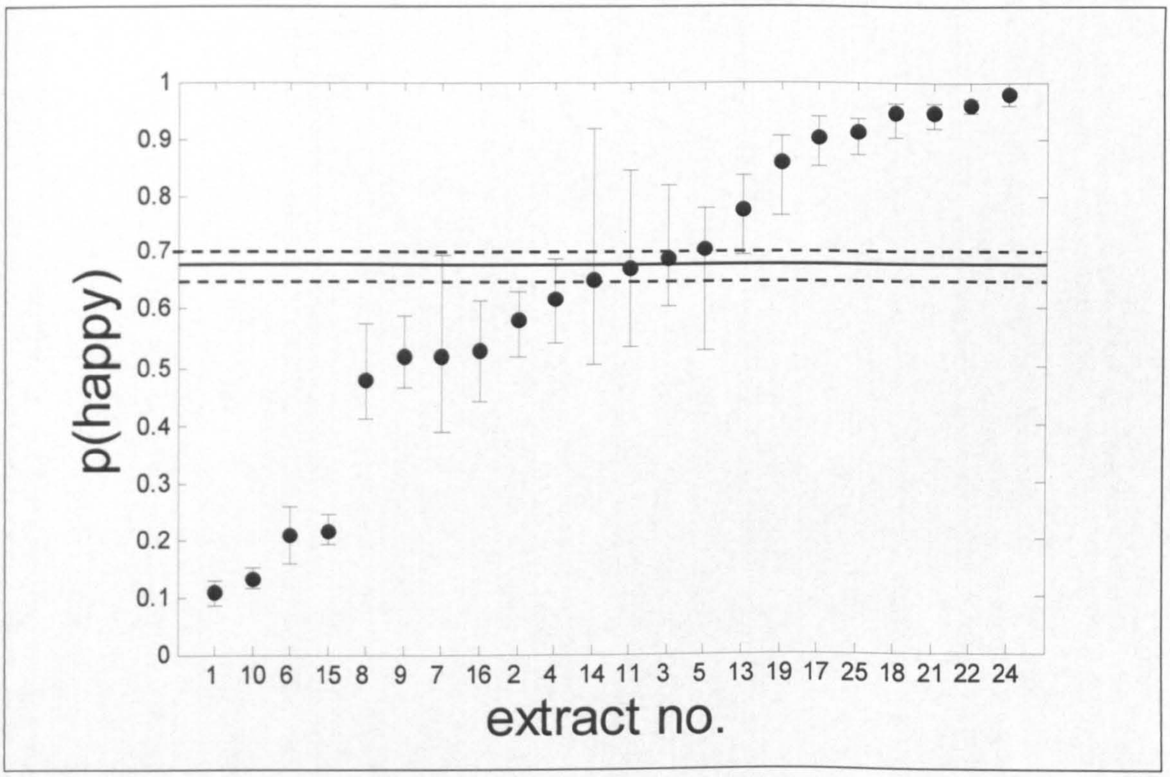

Figure 15: The probability of a happy response at the optimal tempo for each extract was obtained by interpolation on the psychometric functions. It also shows the overall global optimal tempi calculated using all the data together shown as a horizontal black line and the standard deviation shown as dotted lines. Each extract is ordered by its probability of a happy response in order to make any clustering apparent. There are three clusters suggested by the data: a low probability cluster (sad cluster), a high probability cluster (happy cluster) and some intermediate data (neutral).

The remaining extracts (11, 3, 5, 13, 19) are somewhat intermediate between clusters 2 and 3. Although there is no overall relationship between either happy or sad tempo and optimal tempo, figure 15 shows that, for the most part the optimal tempo tends to lie either close to the sad tempo (cluster 1) or close to the happy tempo (cluster 2) or at the point of neutrality (cluster 3). Interestingly, the extracts in clusters 1 (sad) and 2 (happy) have no sad tempi that are not at the common 30 bpm level. Sad tempi of 30 bpm seem to apply only to the neutral extracts.

DISCUSSION

This study was conducted to assess the relationship between tempo and the perception of happy and sad responses for pieces of music. It has been shown that listeners can make consistent happy and sad judgements for pieces of music that vary in tempo. The results also show that by varying the tempo of a piece of music, the same piece can be perceived as happy or sad. The happy and sad tempi varied from piece to piece, where the sad tempi varied from 5 bpm to 90 bpm and the happy tempi varied from 55 bpm to 180 bpm. However, there was no evidence of a common happy tempo or common sad tempo.

The next point was whether there was a relationship between optimal tempo and the happy tempo and sad tempo. Overall the happy and sad tempi were not correlated with the optimal tempo. Whilst this is the case, calculating the probability of a happy response at the optimal tempo proved more useful and it seemed that 3 clusters emerged. Where this probability is high then the music would be happy at the optimal tempo and is essentially happy music. A low probability would mean that the music would be sad at the optimal. Those extracts in between are the least likely to be happy or sad at the optimal tempo and are neutral. Those at the extremes are of interest as they have a tendency towards being more happy and sad than the other extracts. These extracts would be perceived as happy or sad when they were played by a musician at the tempo that sounds right.

It was of interest to assess the relationship between structural features of the music and the happy and sad tempi. The first analysis took all the extracts together. We assessed a range of different features that include the mode of the music (major or minor), the frequency of the most common pitch, the mean interval direction, mean absolute interval size, proportion of ascending and descending pitches, no. of chord changes per second, no. of different pitches used per second and note length standard deviation. The only feature that was significantly correlated with tempi was mode: mode with sad tempo ($r = 0.64$; $p<0.01$) and mode with happy tempo ($r = 0.57$, $p<0.01$). This supports findings in the literature (Hevner, 1935a, 1935b, 1936, 1937; Scherer and Oshinsky, 1977). Since the major and minor mode has this influence it might be expected that the number of chord changes would also have a similar effect as these reinforce the mode of the music. Although the chords were not played to them explicitly it might be possible that the listeners pick up on the relationship between melodies and their underlying harmonies through musical listening. Surprisingly, this feature does not have a relationship with the tempi the music is perceived to be most happy or sad: no. of chord changes versus sad tempo ($r = -0.006$, $p = 0.98$) and no. of chords versus the happy tempo ($r = 0.038$, $p = 0.87$).

It is suggested that descending and ascending intervals are associated with the perception of emotion where descending intervals are likely to make the music sad whilst ascending intervals are likely to make the music happy (Cooke, 1959). Whilst this is the case it seems that neither of these features played a role in the tempi the music was perceived to be happy or sad: descending intervals versus sad tempo ($r = -0.095$, $p = 0.67$), descending intervals versus happy tempo ($r = 0.347$, $p = 0.11$), ascending intervals versus sad tempo ($r = 0.129$, $p = 0.56$) and ascending intervals versus happy tempo ($r = -0.244$, $p = 0.27$).

We refined the analysis of structural features in the music by removing the neutral extracts and analysing just those extracts in the happy and sad clusters. All of the extracts in the happy cluster are major and all of the extracts in the sad cluster are minor. As before, no other musical features were correlated with the happy or sad tempi for these extracts, except the mean interval size which shows a clear pattern (see figure 16a and 16b). As the size of the intervals increase for the music in the happy and sad clusters the sad tempo decreases (see figure 16a). Similarly, as the size of intervals increase for music in the happy and sad clusters the happy tempo decreases (see figure 16b). What this means is that music that contains lots of large intervals can be played quite slowly and still be perceived as happy, but would need to be played very slowly in order to be perceived as sad. Likewise music that contains small intervals must be played very fast to be perceived as happy, but can be played quite fast and still be perceived as sad. However, this is not the whole story. Figure 16 shows one outlier (actually extract 1) and some scatter which suggests that other factors must also be involved. Moreover, the relationship between interval size and tempo does not hold for neutral extracts. It is difficult to identify why this might be the case because no other musical feature seemed to be involved.

Figure 16: The sad tempo and happy tempo as a function of the interval size per note

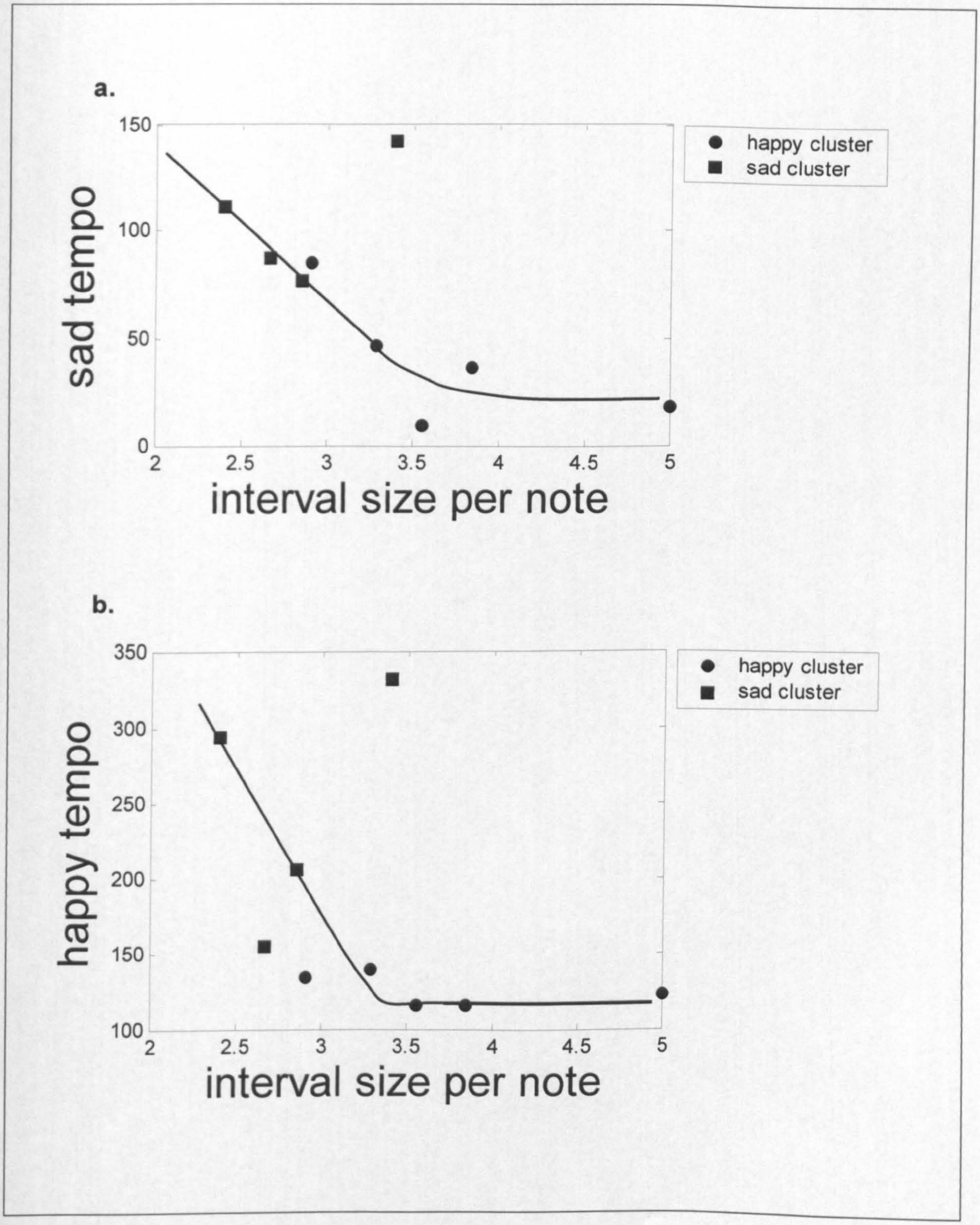

Figure 16: The mean interval size in the happy and sad cluster as a function of the happy tempo and the sad tempo. As the size of the intervals increase for the music in the happy

and sad clusters the sad tempo decreases (figure 16a). Similarly, as the size of intervals increase for music in the happy and sad clusters the happy tempo decreases (figure 16b).

Regardless of this, there is an important implication of this finding. Consider two pieces of music both with the same rhythm, but one of them contains small intervals and the other contains large intervals. Figure 17 presents an example of the psychometric functions for these two extracts: the left curve has large intervals; the right curve has small intervals. Both curves show that reducing tempo increases sad responses. The music that contains large intervals (left curve) could be played quite slowly and sound happy, but if the large intervals were replaced by smaller intervals at the same tempo it would have the effect of making the music sound sad (see figure 17 arrow B). If we replace large intervals by small intervals in this way but wish to keep its affective character unchanged (proportion of happy responses), we need to make a compensating increase in tempo (see figure 17 arrow A). Likewise, if we have a piece of music that contains small intervals it could be played quite fast and sound sad. If we replace small intervals by large intervals and wish to maintain its character, it would have to be played even slower in order to sound sad (see figure 17 arrow C). In this sense interval size and tempo work together in their effect on the happy and sad character of a piece. Given this, it is possible, at least in this specific regard, to treat interval size as if it has a tempo value. So, reducing tempo makes music sad. Similarly, reducing interval size makes music sad. In some way, reducing interval size and reducing tempo are equivalent to each other in terms of their effect. It is as if strong events (large intervals) cause the effect of an increase in tempo for some perceptual purposes; weak events (small intervals) have the opposite effect.

Figure 17: The effect of large and small intervals on perceived happy and sad emotions

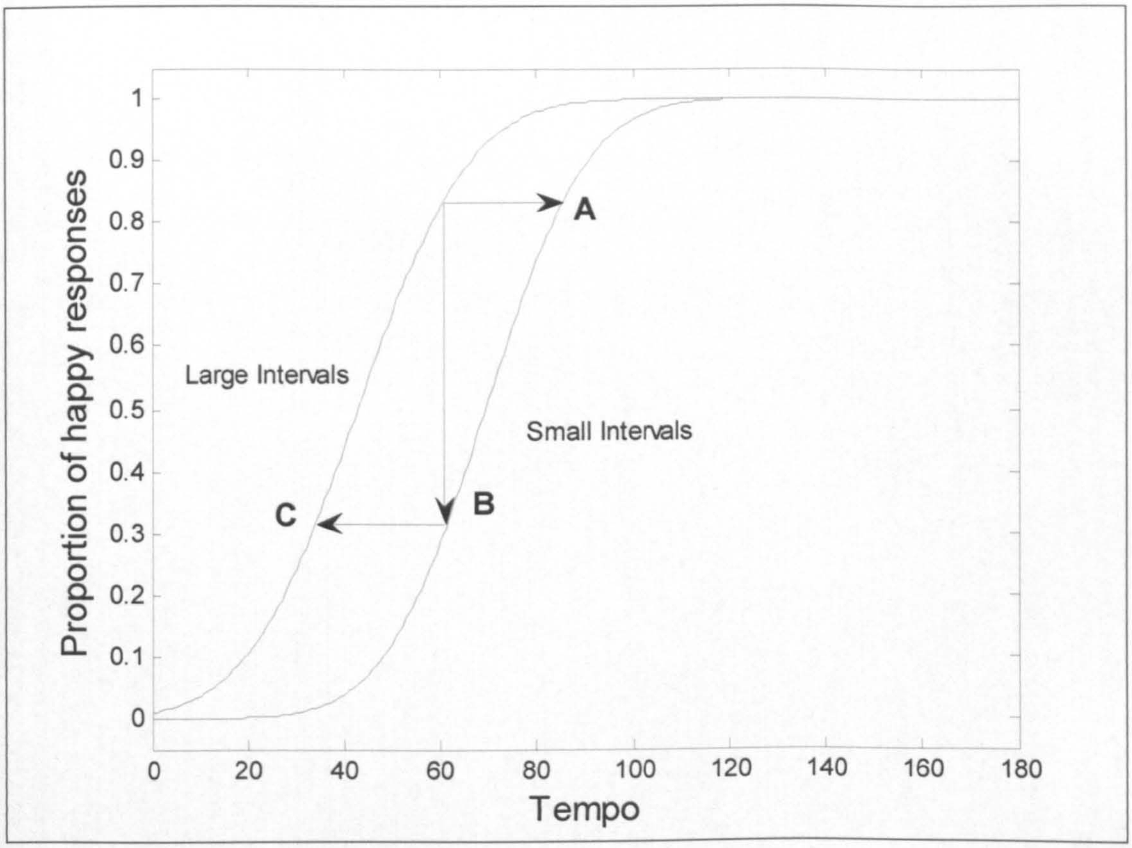

Figure 17: The effect of large and small intervals on perceived happy and sad emotion. If the music is happy and contains large intervals, but the tempo is increased the intervals would have to get smaller in order to remain happy (arrow A). If the music is sad and contains small intervals, but the tempo is decreased the intervals would have to get larger to remain sad (arrow C). When the tempo is held constant for large intervals the effect of making it sad would require it to contain smaller intervals (arrow B).

Why is there a relationship between interval size and timing? In considering what might control optimal tempi, Chapter 3 argued that music contains events that vary in their strength and that it might be these variations that give rise to variations in optimal tempo from piece to piece. They argued that there is an interaction between the strength of an event and its preferred duration. It seems that the present results begin to support their suggestion, where different interval sizes have varying levels of strength. Return to the example above: we have two pieces of music with the same rhythmic content but one contains large intervals and the other contains small intervals. In order to maintain the same emotional character the one with small intervals would need to be performed more quickly. The number of notes per second heard in any piece with small intervals would need to increase by comparison to a piece with large intervals to maintain its position on the happy or sad dimension. Large intervals are therefore stronger because we do not need to hear as many of them per second in order to maintain a particular level of happy or sad response. Small intervals are weaker because we need to hear more of them in order to maintain the same level of emotional response.

There are two possible explanations for these findings. The first assumes that large intervals compress time (as if increasing the tempo) and small intervals dilate time (as if decreasing the tempo). In these circumstances it is timing that determines the emotional character where a slow tempo is sad and a fast tempo is happy. The alternative assumes that a slow tempo reduces the interval size and fast tempo increases the interval size. By doing so the interval size determines the emotional character where large intervals make music happy and small intervals make music sad. We prefer the first of these since it allows other factors (e.g. unexpected pitches) to have similar effects.

CHAPTER 6: CHANGES IN AUDITORY PITCH CAN CAUSE CHANGES IN PERCEIVED DURATION

ABSTRACT

We report a study that shows that the perceived duration of a sequence of 3 auditory tones with a physical duration of 0.5 seconds can be increased by up to 35% depending on the changes of pitch between the tones. Current explanations for the experience of duration rely on internal measuring devices based on specifically generated time signals, such as a pacemaker-accumulator (Treisman, 1963) or an internal clock (Church and Broadbent, 1990). There has been debate about the usefulness of these systems for explaining differences between physical and perceived time, as it is not clear why the internal signal would have a variable rate for different stimuli. Our results show that mere differences in pitch can cause our perception of time to dilate by a large amount. We conclude that it is stimulus properties per se that are determining perceived duration rather than any internal time signal gated by the stimulus. We explain this finding in terms of stimulus strength where strong stimuli that are less frequent cause perceived time to dilate compared with weak stimuli. Perceived time is therefore only a sequence of weak and strong events with no independent internal time base.

INTRODUCTION

Time perception is central to our daily experience. A number of different factors are believed to affect our experience of time and explain why it fluctuates such as: age related affects (Rakitin, Stern, and Malapan, 2005); body temperature (Hancock, 1993; Wearden and Penton-Voak, 1995); arousal (Penton-Voak, et al., 1996; Angrilli, Cherubini, Pavese, and Manfredini, 1997); and dopaminergic dysfunction in Parkinsons disease (Perbal, Deweer, Pillon, Vidailhet, Dubois, and Pouthas, 2005; Malapani, Rakitini, Meck, Levy, Deweer, Dubois, Agid, and Gibbon, 1998). Early attempts to explain the experience of time suggested that internal mechanisms or clocks might be involved (François, 1927). Since then the area has gone on to propose a number of revisions to these internal mechanisms and some alternatives. However, we present data that questions the fundamental basis of such models to explain time. The purpose of this chapter is to show that the perception of time is driven by the features of the stimulus and not internal time keepers.

There are 2 internal systems that have been suggested to be responsible for the way we process time: an internal clock known as the pacemaker-accumulator model; an internal clock involving oscillators. To begin with we shall deal with the pacemaker-accumulator version. The pacemaker-accumulator contains a pacemaker, accumulator, switch, storage devices and a response mechanism (Wearden, 2001; Wearden, 1999). The pacemaker is the central part of the clock mechanism that creates a series of pulses. It was initially suggested that the rate at which pulses are produced is regular (Treisman, 1963), but subsequent revisions proposed that it was a Poisson pacemaker (Gibbon, 1977). Poisson pacemakers generate pulses at random such that the duration between pulses are variable, but over longer time spans the duration between them is averaged out (Wearden, 1999). When asked to make duration judgements the switch between the pacemaker and accumulator closes allowing a series of pulses to flow between the pacemaker and the accumulator. When the stimulus is no longer present the switch opens and the route from the pacemaker to the accumulator is cut off. The switch opens and closes with a variable latency that can be affected by attention or other factors related to the time taken to react to a stimulus. The accumulator acts like a container for the pulses that are generated by the pacemaker. The amount of time that has passed is given by the number of pulses in the accumulator.

Since most studies use paired stimuli in order to assess our perception of time it makes sense that some sort of storage system would be required at this stage. Wearden (2001) suggests that the type of storage device used depends upon task demands. If the task requires an immediate response then information may only need to be stored in short-term memory. However, tasks involving judgements where a response is based upon comparisons of previously learned durations with new stimuli entering the system would require information to be stored in long term memory. Responses are therefore made by comparing the differences between stored durations and a comparison with new stimuli where disparities greater than some threshold would be used to create a response. If the difference between the two is greater than this threshold then one of the stimuli would be recognised as different from the other. On the other hand if the difference is less than this threshold then a 'same' response would be made.

Errors in making duration judgements stem from two main sources. The first is variable error and the second is constant error. In order to explain what these mean we shall turn to an example of a study we might use to assess perceived duration. In this study we present two stimuli (a standard stimulus lasting 0.5 seconds and a test stimulus that varies in its physical duration at set values above and below the duration of the standard). Each level of the cue (the set values of the duration of the test stimulus) is presented to the participant 3 times and for each presentation they have to decide whether the standard or test stimulus is longer. On each of the 3 occasions the participant perceives the duration of the standard to be 0.8, 0.9, and 0.7 seconds. In this case the variable error is 0.1 seconds whilst the constant error is 0.3 seconds which is an average of the bias from each occasion. The variable error refers to changes in perceived duration from trial to trial (in this case it is 0.1 seconds) whilst the constant error refers to the consistently 'longer' responses the participant makes (in this case 0.3 seconds).

All parts of the pacemaker-accumulator model could contribute to variable errors. Variability would occur because the pacemaker produces pulses that are irregular over short periods and because the switch varies in its onset and offset latencies. The memory store would also cause some variable error because storage and retrieval from short-term and especially long-term memory might not be accurate. Finally, the decision process would also cause some variable error because the threshold used to make a response would be an average of previous responses. Since previous responses would contain some variable error in themselves this would also affect the threshold created.

Constant errors would be explained by systematic variations in the rate of the pacemaker and the onset and offset of the switch. The pacemaker could produce pulses at faster or slower rates dependent upon the stimulus being presented and the switch could have a faster or slower latency for the same reason. This type of explanation has been used to account for studies showing perceived differences in the duration of sounds compared to lights (Goldstone and Lhamon, 1974; Wearden, et al., 1998), and filled durations compared to unfilled durations (Buffardi, 1971). Clearly these stimuli are very different. It would be expected that stimuli that are very similar would cause the pacemaker and switch to run at a similar rate overall and any constant errors involved would be the same. However, the present study shows that even small changes in a stimulus can have significant consequences for perceived duration.

The oscillator-based system is thought to contain a number of oscillators (11 in total) that have cycles containing different periodicities (Church and Broadbent, 1990). Originally, it was suggested that the oscillators might be driven by a range of biological functions including: a circadian rhythm; the central nervous system; respiratory systems; circulatory systems; hormonal systems; and behavioural systems (Gallistel, 1990). They suggest that the regularity of these multiple oscillatory systems would permit precise measurements ranging from years to milliseconds. The period of each oscillator is a multiple of 2 of the next most fast such that the 1st oscillator would have a cycle of 0.2 seconds, the 2nd oscillator would have a cycle at 0.4 seconds and so on.

As the set of oscillators switch from positive to negative phases at different rates, the pattern of phases across oscillators at any given moment produces a unique pattern that can be used to determine stimulus duration. Each pattern has been learned at some point and is stored in long-term memory. Comparisons between those patterns created by the oscillators and previously learned patterns provide the information needed to make a judgement about the duration of a stimulus.

The difference between the pacemaker-accumulator model and this one is that it is not possible for one oscillator to vary independently from the others because the pattern created would not be comparable with those stored in reference memory. Essentially it would create a nonsense pattern that the individual would not have stored and therefore understand. The oscillators have to be coupled so that their rate varied together. Then the oscillators could give rise to variable errors by random variations in rate and to constant errors by systematic variations in rate. Differences between constant errors for different stimuli would only arise if stimuli were very different from one another and, small changes in a stimulus should therefore have no effect. Coupled oscillators and pacemaker-accumulators are essentially similar because they both involve regular timing mechanisms. The arguments presented earlier regarding the pacemaker-accumulators are therefore the same for the oscillator based system.

We report a study that shows that small changes in pitch can have a dramatic effect on perceived duration. The results show that by altering the pitch of the middle tone of a 3 tone stimulus its perceived duration can increase by up to 30% by comparison to a stimulus of equal duration with no change in pitch. The results presented here question the usefulness of the pacemaker-accumulator model and the oscillatory based system because the size of the constant errors observed vary across the pitch changes. Perceived duration is therefore driven by some feature of the stimulus and not some internal timing system.

METHODS

Participants

3 Participants took part in the study (the author, colleague and an undergraduate psychology student from the University of Stirling). The participants were selected on the basis that they were highly experienced in psychophysical studies. The undergraduate was naïve to the purpose of the study.

Materials

Listeners were asked to compare the durations of a test and standard stimulus each containing 3 tones. The stimuli were created by using a MATLAB function called sound play that delivers the waveform directly to the line out socket on the PC. Each tone was made up of the fundamental and the first 3 harmonics, all of equal amplitude. This gave the tone a rich sound and a readily identifiable pitch. Each tone had a linear attack growth in amplitude over the first 50ms, and a similar decay over the last 50ms. This waveform envelope avoided any clicks or other audible transients. The test stimulus contained 3 tones without any silence between them; the first and last were the same pitch (middle C) whilst the middle tone varied in pitch. The pitch of the middle tone varied from trial to trial from middle C to an octave above middle C using all of the semitones in between. The standard stimulus contained 3 tones that were all the same pitch (middle C). On each trial, the duration of the standard stimulus was randomly set to a value between 385ms and 650ms. The test stimulus had a duration that was the duration of the standard multiplied by the cue amount which varied between 0.25 and 4. The boundaries from one tone to the next inside the stimulus were placed randomly, subject only to the constraint that each tone should last a minimum of 18% of the total stimulus duration. Each stimulus level was presented 10 times in a fully randomised sequence. Stimuli were played to participants through high quality headphones (Sennheiser HD 250 linear II headphones) at a volume level that was comfortable.

Procedure

The method of constant stimuli was used to measure the point of subjective equality (PSE) which is a way of establishing the perceived duration of the test stimulus. Before the beginning of the experiment proper, each participant was given a short practice trial in order to familiarise them with the task. The stimuli were then played to the participants and after hearing each stimulus they were asked to make a response by indicating whether the first or second stimulus was longer in duration by pressing designated keys on a computer keyboard. They were instructed to make just one response for each pair, even if they were unsure about their response.

Design

The experiment made use of a within subjects design. There were two independent variables of duration and pitch change. The dependent variable was the forced choice response. The design was constructed to avoid, as far as possible, the possibility of the stimulus set providing a normative duration and beat against which participants could make judgements. Participants heard the set of stimuli in a completely randomized order with neither pitch nor duration blocked. This avoids any effects of a tendency to judge the duration of any given stimulus with respect to its immediate predecessors.

RESULTS

Figure 18 shows the PSE as a function of the size of the pitch change in the test stimulus. If we expect participants to be accurate at this task we would expect them to perceive the test stimulus to be equal in duration to the standard stimulus at 0.5 seconds because that was the only occasion where they were both physically equal. The data shows that this is not the case. For small pitch changes between 0 and 3 they perceive the standard and test equally. As we increased the size of the pitch change to medium sized pitches from 4 to 6 the perceived duration of the test increases. Large pitch changes between 7 and 11 cause perceived duration to increase further still with the largest perceived duration of 0.682 seconds. When we reach a pitch change of 12 we observe that participants perceive the stimuli equally.

Figure 18: Point of subjective equality as a function of size of pitch change

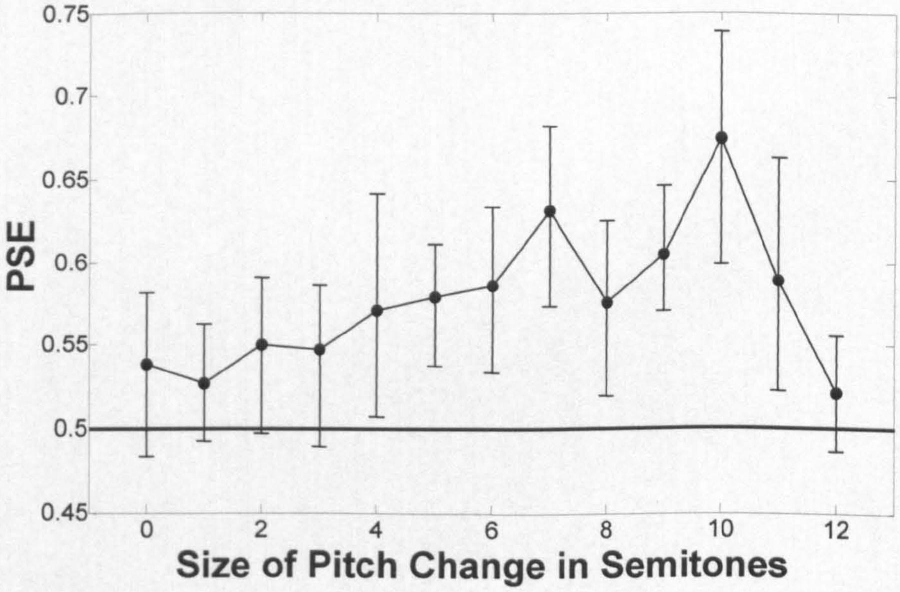

Figure 18: Point of subjective equality (ie perceived duration of the test stimulus) as a function of the size of pitch change (measured in semitones where 0 is middle C and 12 is

an octave above middle C). The error bars are 95% confidence limits. The perceived duration increases as the size of the pitch change increases, but falls back down at a pitch change of 11 and 12 (12 is an octave). As can be seen, the perceived duration is reliably greater than the physical duration for pitch changes between 0 and 3.

DISCUSSION

The results show that the size of pitch changes within a stimulus can cause differences in perceived duration. It is important to note that the two features that delimit the stimulus – the onset of the first tone and offset of the third tone – were always identical as the first and third tones were the same in all stimuli. The only systematic variable across test stimuli was the size of the pitch change to and from the second tone rather than the durational and rhythmic characteristics of the stimulus. In light of this, constant errors in perceived duration depend solely on the size of the pitch changes in the stimulus rather than any internal timing within the stimulus.

Most explanations for the experience of time suggest the use of internal clocks of one type or another. The pacemaker-accumulator model and the oscillator based system can account for constant errors in perceived duration by supposing that the clock rate varies depending on the stimulus. This is plausible for markedly different stimuli, such as sounds that are perceived longer than lights or filled durations that are perceived longer than unfilled durations.

However, the constant errors observed in the present study would not be accounted for by the pacemaker-accumulator model or the oscillator model, unless the underlying timer mechanism was substantially influenced from moment to moment by irrelevant features of the stimulus. A device that follows the stimulus so closely can hardly serve as an independent timer. By changing the middle tone by some small amount the constant errors should either be the same across each stimulus level or be zero.

To explain what is responsible for these results we shall turn to 4 main findings from our study that showed that the listeners perceived:

i) the standard stimulus containing 3 Cs (lasting 0.4 seconds) as shorter than the test stimulus containing 3 Cs (lasting 0.5 seconds);

ii) the test stimulus containing the same pitch (CCC lasting 0.5 seconds) as shorter than the standard stimulus (CCC lasting 0.6 seconds);

iii) the standard stimulus containing 3 Cs (lasting 0.8 seconds) as longer than the test containing a change in pitch (CBC lasting 0.5 seconds);

iv) the standard stimulus containing 3 Cs (lasting 0.6 seconds) as shorter than the test containing a change in pitch (CBC lasting 0.5 seconds).

We suggest that these judgements are based upon the strength of the stimulus and not some internal time base. One dimension that influences the strength of a stimulus is how rare the events contained within it are. Where the stimulus has a large change in pitch, this is a rare event for the listener. Figure 19 shows the frequency that pitch changes occur in 750 simple single line melodies. The distribution shows that increasing the size of changes in pitch decreases the frequency with which they occur. Small pitch changes are very common whilst large pitch changes are least common. Listeners would have heard music that uses these types of changes through normal musical listening and would recognise how usual or unusual any pitch change was. We suggest that common pitch changes (small pitch changes) are perceived briefest as they are heard most often, and uncommon pitch changes (large pitch changes) are perceived to be much longer because they are heard less often.

Figure 19: Frequency of pitch changes as a function of the size of the pitch change

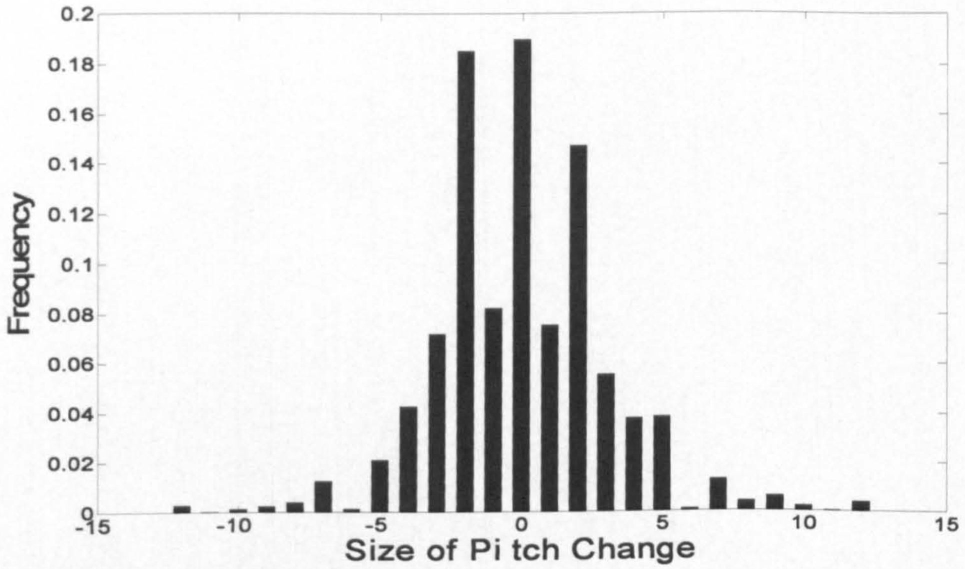

Figure 19: The frequency of occurrence of pitch changes as a function of the size of pitch change for 750 simple single line tonal melodies. Small pitch changes occur most frequently whilst large pitch changes occur least frequently.

The second component that causes stimulus strength to grow is stimulus duration. As a weak stimulus continues for a much longer period it grows in strength. The longer the stimulus is sustained the stronger it becomes so that weaker stimuli can become stronger simply by playing them for longer. Long events are therefore rare in a similar way to large pitch changes. Figure 20 shows the frequency that durations occur in the 750 melodies used in the previous example. The distribution shows that increasing the duration of a note decreases the frequency with which it is used. It is important to notice though that these two features are not mutually exclusive and there may be other features that are equally influential in the strength of a stimulus.

Figure 20: Frequency of note durations as a function of the size of a note duration

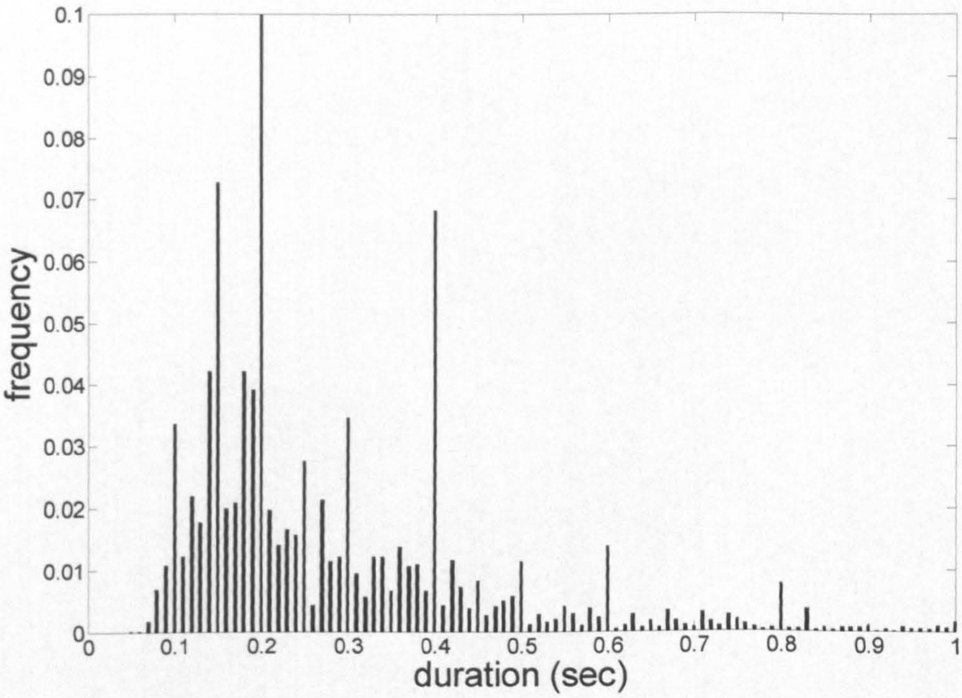

Figure 20: The frequency of occurrence of note durations for 750 simple single line tonal melodies. Small note durations occur most frequently whilst large note durations occur least frequently.

Now we show how the present results can be explained by suggesting that participants based their responses on stimulus strength and not duration. Figure 21 presents a schematic representation of the results for the CBC stimulus (test) and the CCC stimulus (standard).

(i) Stimulus B (CCC lasting 0.5 seconds) is stronger than stimulus A (CCC lasting 0.4 seconds) because it has grown in strength as its duration has increased.

(ii) Stimulus B (lasting 0.5 seconds) is perceived shorter than stimulus C (lasting 0.6 seconds) because C has grown in strength as its duration has increased.

(iii) Stimulus D (CCC lasting 0.8 seconds) has a greater strength than stimulus E (CBC lasting 0.5 seconds) because its strength has grown as its duration has increased. The change in pitch increases the strength of E (the CBC stimulus) by comparison to D (the CCC stimulus), but it is not as strong in this case as stimulus D.

(iv) Lastly, stimulus E (CBC stimulus lasting 0.5 seconds) is stronger than C (CCC lasting 0.6 seconds) because of the change in pitch.

The results are therefore explained by the notion that the strength of the stimulus is used to judge its perceived duration.

Figure 21: Event Strength as a function of duration for the Standard Stimulus (CCC) and the test stimulus (CBC)

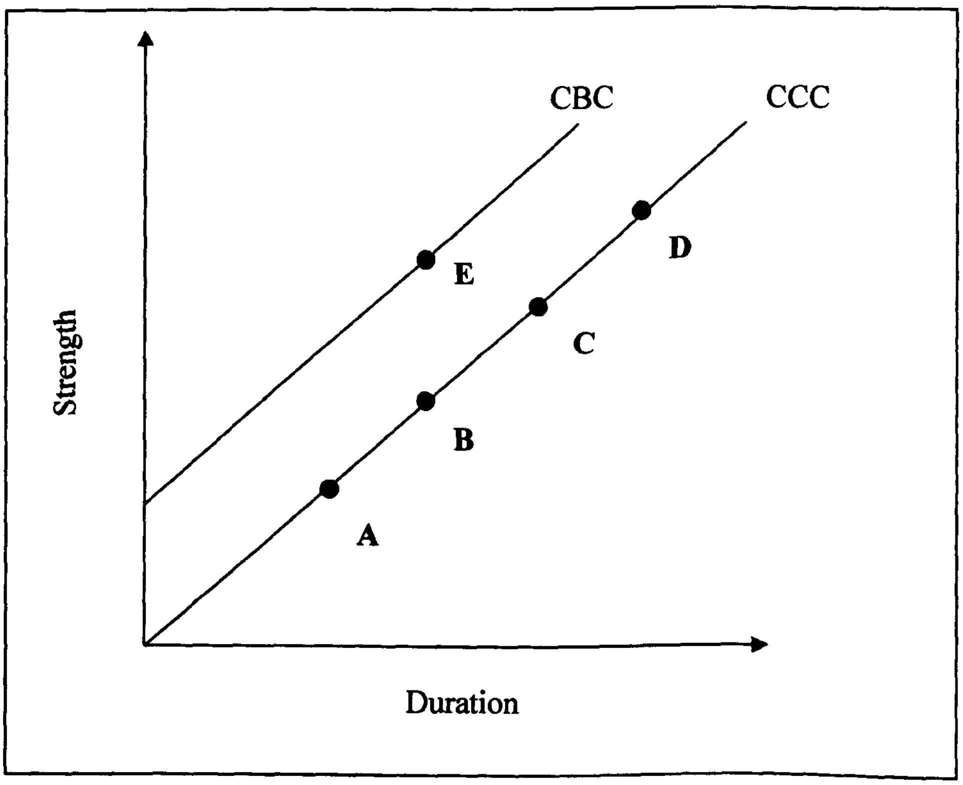

Figure 21: A schematic diagram of the effects of event strength as a function of duration for the standard stimulus (CCC) and the test stimulus (CBC). Stimulus D is stronger than A, B, and C because it lasts longer. Stimulus C lasts longer than stimulus E, but E is stronger than C because of the change in pitch. Stimulus D lasts longer than E because its strength has grown as its duration increased. The change in pitch increases the strength of E by comparison to D, but it is not as strong as stimulus D because its duration is much longer.

The conclusions made here are not entirely without foundation as it has been suggested that the intensity of a stimulus can contribute to its perceived duration where brighter lights are judged longer than dimmer ones (Goldstone and Goldfarb, 1964) and reducing auditory intensity reduces perceived duration (Walker and Scott, 1981). These findings have also been supported by Tse, Intriligator, Rivest and Cavanagh (2004) who recognised not only the importance of the stimulus for perceived duration, but the effects of changing their characteristics.

In a series of studies the researchers assessed the effects of changes within a stimulus and its effects on perceived duration in both the visual and auditory domains (Tse, Intriligator, Rivest and Cavanagh, 2004). In two of the studies reported they found that an auditory stimulus that used a glissando effect that rises smoothly and a visual equivalent that increased in size caused an increase in perceived duration. The authors suggest that this effect was due to the demands of the task where high levels of attention processing cause an increase in the perceived duration of the stimuli. They also suggest that a low-probability stimulus would orient more attentional processing that a high-probability stimulus. However, it could be argued that an increase in pitch causes listeners to perceive the stimulus as longer because it creates a sense of motion. The same is true of the visual stimulus they used where an increase in size produces the effect of the stimulus moving towards the participant. If they are perceiving motion when they are presented with these stimuli, then their results could be explained in these simple terms. However, this explanation could not account for the present results as the stimuli used changed in location (individual tones changed in pitch) rather than increased in size (glissando effect). Interestingly they do mention that variations in perceived duration might be due to the stimuli having a low or high-probability. The lack of any definitive description of what causes a stimulus to be either of these merely explains their effect rather than the actual features involved. We go somewhat further than this by defining what strength might refer to and suggest that both duration and the contents of a stimulus (pitch in this case), contribute to stimulus strength.

To recap, the results show that the perceived duration of our stimuli increased by up to 35% depending upon the pitch change involved. We suggest that this is due to the strength of the pitch changes and increases in duration. Event strength is achieved by increasing the duration of a stimulus that causes its strength to grow. Likewise, pitch changes that are rare cause variations in perceived duration. It is therefore the stimulus that is responsible for variations in perceived time rather than an internal time base. The perception of time is therefore a consequence of the strength of the events we experience where weak events and strong events have differential effects on perceived time.

CHAPTER 7: DISCUSSION

This thesis reports on a series of studies that were developed to examine the perception of time in music. Recall that the areas of interest included global optimal tempo, tempo and its relationship with emotion, and the duration of 3 tone stimuli with changes in pitch. These were briefly outlined in the first chapter and individually focused on during the course of each subsequent experimental chapter. Initially we wanted to establish whether there were optimal tempi for pieces of Scottish fiddle music and extract from this any musical features that were responsible. Having assessed this we wanted to ensure that the method of constant stimuli was an appropriate and useful tool for investigating tempo in music. Following on from this the relationship between tempo and perceived emotion of happy and sad responses was investigated. Finally, we evaluated the effects that small changes to 3 tone stimuli have on perceived duration. The main findings from each experimental chapter will be dealt with in the order they appeared in the thesis with a summary of what the results collectively tell us about the perception of time in music.

SUMMARY

The Primal Sketch in Music

The primal sketch was introduced at the beginning of this thesis and was intended to provide a framework for explaining the processes involved in the perception of music. It was suggested that the raw primal sketch in music provides a description of the basic features of the notes in a piece of music and the full primal sketch pulls these notes together into larger musical groupings. A final musical representation follows these 2 perceptual processes and can include a large range of possible musical conceptualisations where listeners produce optimal tempo representations, emotional representations and durational representations that are the result of the assimilated features contained within the full primal sketch and raw primal sketch. Each of the tasks described in this thesis would require both the raw primal sketch, full primal sketch and the representational stage.

It was also argued that particular notes have greater influence than others and are determined by the properties they contain and the context in which they are placed. We suggested that notes have varying degrees of strength that are determined by their physical properties and by their relationship with the notes that surround them. The amount of strength a piece of music contains will determine the degree to which it is happy rather than sad or how fast or slow it ought to be played. The strength of changes in the pitch of groups of notes also determine whether they are perceived to be longer of shorter than a group of notes without a change in pitch. It is therefore the strength of musical events that shape the perceptual processing that listeners experience.

Tempo Perception

We wanted to find out whether there were optimal tempi that listeners perceived to be appropriate for pieces of Scottish fiddle music using the method of constant stimuli. The results revealed that there were optimal tempi for extracts of music and that these varied from piece to piece, but were consistent across a number of different studies. A second study explored the possibility that the rhythm was responsible by using just the rhythmic patterns from the extracts in the previous study. Listeners were capable of making tempo judgments for the rhythmic patterns alone and at least for some of these the optimal tempi turned out to be very similar to those in the first study. We concluded that rhythm must play some role in this process, but not the only one. As some of the extracts in the rhythm study did not produce the same optimal tempi as they had in the 1^{st} study, it seemed to suggest that other structural features of the music were also important.

At the beginning of this thesis we suggested that some musical features will have a greater influence over others in shaping the response a listener makes (as outlined in the primal sketch for music). Tempo judgements are therefore made by pulling upon the strongest features contained in the music and these have to be played at particular tempi in order for them to make sense to the listener. It is reasonable to assume that the rhythm was a much stronger feature for some extracts and less so than pitch for others, or in some cases require an equilibrium between pitch and duration to be achieved.

Method of Constant Stimuli

As we discussed in Chapter 3 we considered the possibility that listeners might be using an average tempo from previously heard extracts in order to do the task of making a 'too fast' or 'too slow' judgment. We therefore wanted to assure ourselves that the listeners were not creating their own bias to make the task easier. The method of constant stimuli was modified to cater for this by presenting a single extract at one level of the cue per participant and therefore eliminating this issue. To conduct such a study required a large sample of the population, but the experimenter and author of this thesis was mad enough to give it a bash. The results were surprising and reassuring in that they provided s-shaped psychometric functions that were similar to those presented in Chapter 3. On this basis we argued that listeners can make 'too fast' or 'too slow' responses without the need to create an average of previously heard extracts. In addition, it provides support for using this technique in future studies where this might be a confounding variable.

Optimal tempo and the perception of happy and sad emotions in music

The emotion study examined the relationship between global tempo and the perception of happy and sad responses for the pieces of music used in the previous 2 Chapters. A large range of tempi was employed to assess this. The results revealed that the same piece of music could be perceived as both happy and sad by manipulating the global tempo and that the happy and sad tempi varied from piece to piece. We then assessed whether the music was happy or sad at the optimal and found that some pieces of music were more happy than sad at the optimal tempo, or more sad than happy at the optimal tempo. However, a number of the pieces were neither happy or sad at the optimal tempo which suggested that they were neutral at the optimal. Having assessed a range of musical features the size of pitch changes in the music was most relevant. We showed that the tempi at which the music was happy and sad were related to the size of the pitch changes contained in the music.

In explaining this finding we suggested that the only way a piece of music can become happy or sad would require 2 types of changes to be made: an increase or decrease in interval size or an increase or decrease in tempo. For example, if the music was played at a medium tempo and contained small intervals where the listeners perceived it to be sad, one way to make it happy would be to increase the size of the intervals. Similarly, if we took a piece of music that was played at the same speed, but it contained large intervals and was perceived as happy, one way of making it sad would be to decrease the size of intervals contained in the piece. Alternatively, if we wanted to maintain the same musical structure as the first example, the other way of making this same piece of music happy would involve increasing its tempo. When we take the second example mentioned above we would need to decrease its tempo to make it sad in order to maintain its musical features. On the basis of these observations it became clear that changes to the size of intervals and tempo could achieve the same effects and that they were somehow linked.

Time Perception and Pitch Changes

In considering what might be involved in the perception of time in music we wanted to assess what effect small changes to a stimulus would have on perceived duration. As interval size and tempo seemed to produce the same effects, it seemed necessary to illustrate this more clearly by looking at it in without an emotion judgment being made. The results revealed that perceived duration increased as the size of the pitch change increased. We suggested that uncommon pitch changes are perceived as longer than common pitch changes that are perceived much shorter by comparison. We went on to suppose that common pitch changes are weak events and uncommon pitch changes are strong events. Additionally, long durations were suggested to be strong and short durations were suggested to be weak. Strong events in this case cause perceived duration to increase whilst weak events are perceived to be shorter than strong events.

Time Perception in Music

The key assumption that has been made throughout this thesis is the notion that music contains strong and weak events and that particular musical features that the listeners perceive to be salient are responsible. There is an interesting implication for the findings in chapter 3 and 4 that lead us to understand what the relationship between time and interval size is. Lets remind ourselves of the main finding from chapter 3. We showed that increasing the size of an interval has the same effect as an increase in tempo. Large intervals should therefore feel as though they are going faster than they actually are and therefore be perceived as shorter in perceived duration. However, in the 4th chapter we showed that the perceived duration of a large interval is perceived to be longer than a small interval.

How might we explain this conundrum? In order to explain this we shall turn to an example. A piece of music would not be rushed if it contained small intervals and was being played at a moderate tempo. In order to make this piece of music rushed the tempo would either have to increase or the interval size would need to increase to gain the same effect. Changing the music to include larger intervals, at a moderate speed, would sound rushed because a large interval requires a longer duration. Try and think of this as squeezing a duration that is longer into a shorter duration. In doing so the auditory sequence would sound rushed.

If we were capable of experiencing time we would see no need for watches, clocks or stopwatches to tell us what time it is or how long something lasts. If we could experience time we would be able to do these tasks properly without the need for an external device to do it for us. As a result I suggest that time does not exist and that what we actually experience are a series of events. Our results showed that differences in perceived duration were due to measurements of stimulus strength and not measurements of time. The data from study 4 explicitly show that this is the case and that we are not capable of perceiving time. The reason why duration judgements vary is due instead to the strength of the stimulus we perceive.

Event strength is delivered by the contents of the events we experience. As we outlined at the beginning, the primal sketch creates a description of the events we experience that we then use to make a response. A 'too fast' or 'too slow' judgment is made by measuring the strength of each note within a passage by comparison to those that surround it. Musical phrases are unique in the sense that they include note durations and pitch structures that are specific to that phrase. As a result 2 notes with the same pitch and duration might contain different levels of strength because of what surrounds them. Each phrase will contain units of strength that determine its optimal tempo. If these notes are not played at the right tempo they lose their effectiveness. Event strength is the measurement used to determine the optimal tempo of a piece of music.

REFERENCES

Adachi, M., Trehub, S.E., and Abe, J-I., 2004. Perceiving emotion in children's songs across age and culture, *Japanese Psychological Research*, **46**(4), 322-336.

Angrilli, A., Cherubini, P., Pavese, A., and Manfredini, S., 1997. The influence of affective factors on time perception, *Perception and Psychophysics*, **59**(6), 972-982.

Balkwill, L-L., Thompson, W.F. and Matsunaga, R. 2004. Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners. *Japenese Psychological Research*, **46** (4), 337-349.

Balkwill, L-L. and Thompson, W.F. 1999. A cross-cultural investigation of the perception of emotion in music: psychophysical and cultural cues. *Music Perception,* **17**, 1, 43-64.

Behrens, G.A., and Green, S., 1993. The ability to identify emotional content of solo improvisations performed vocally and on three different instruments, *Psychology of Music*, **21**, 20-33.

Block, R.A., and Zakay, D., Models of psychological time. In R.A. Block (Ed.), *Cognitive models of psychological time*, (1-35). Hillside, NJ: Erlbaum.

Boltz, M.G., 1998. Tempo Discrimination of musical patterns: Effects due to pitch and rhythmic structure, *Perception and Psychophysics*, **60**(8), 1357-1373.

Brown, S.W., 1995. Time, change, and motion: the effects of stimulus movement on temporal perception, *Perception and Psychophysics*, **57**(1), 105-116.

Buffardi, L., 1971. Factors Affecting the Filled-Duration Illusion in the Auditory, Tactual, and Visual Modalities, *Perception and Psychophysics*, **10**, 292-294.

Casini, L., and Macar, F., 1999. Multiple approaches to investigate the existence of an internal clock using attentional resources, *Behavioural Processes*, **45**, 73-85.

Church, R.M., and Broadbent, H., 1990. Alternative representations of time, number and rate, *Cognition*, **37**, 55-81.

Clarke, E.F., 1987. Levels of structure in the organization of musical time, *Contemporary Music Review*, **2**, 211-238.

Cooke, D. 1959. *The language of music*. Oxford: Clarendon Paperbacks.

Dalla Bella, S., Peretz, I., Rousseau, L. and Gosselin, N. 2001. A developmental study of the affective values of tempo and mode in music. *Cognition*, **80**, B1-B10.

Desain, P. and Honing, H., 1994. Does expressive timing in music performance scale proportionally with tempo, *Psychological Research*, **56**(4), 285-292.

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

Farnsworth, P.R. 1954. A study of the Hevner adjective list. *Journal of Aesthetics and Art Criticism*, **13**, 97-103.

Fraisse, P., 1984. Perception and estimation of time, *Annual Review of Psychology*, **35**, 1-36.

François, M., 1927. Contribution à l' etude du sens du temps: La temperature interne comme facteur de variation de l'appreciation subjective des durées, *L'Année Psychologique*, **27**, 186-204.

Gabrielsson, A. and Juslin, P.N. 1996. Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music*, **24**, 68-81.

Gabrielsson, A., and Lindström, E., 2001. The influence of musical structure on emotional expression. In J.N. Juslin, and J.A. Sloboda (Eds.) Music and Emotion: Oxford, Oxford University Press.

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

Gallistel, C.R., 1990. *The organisation of leaning.* Cambridge, MA: MIT Press.

Gibbon, J., 1977. Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, **84**, 279-325.

Gibson, J.J., 1975. Events are perceivable but time is not. In J.T. Fraser and N. Lawrence (eds.), *The Study of Time II. Proceedings of the Second Conference of the International Society for the Study of Time*, Berlin: Springer-Verlag, 295-301.

Gibson, J.J., 1966. *The senses considered as perceptual systems* (Boston, MA: Houghton and Mifflin)

Goldstone, S., and Goldfarb, J.L., 1964. Auditory and visual time judgement. *Journal of General Psychology*, **57**, 243-248.

Goldstone, S., and Lhamon, W.T., 1974. Studies of auditory-visual differences in human tie judgement: 1. Sounds are judged longer than lights, *Perceptual and Motor Skills*, **39**, 63-82.

Goldstone, S., and Lhamon, W.T., 1972. Auditory-visual difference in human temporal judgment, *Perceptual and Motor Skill*, **34**, 623-633.

Gundlach, R.H., 1935. Factors determining the characterisation of musical phrases, *American Journal of Psychology*, **47**, 624-644.

Hampton, P.J., 1945. The emotional element in music, *The Journal of General Psychology*, **33**, 237-250.

Hancock, P.A., 1993. Body temperature influence of time perception. *Journal of General Psychology*, **120**(3), 197-215.

Hevner, K. 1937. The affective value of pitch and tempo in music. *American Journal of Psychology,* **49**, 621-30.

Hevner, K. 1936. Experimental studies of the elements of expression in music. *American Journal of Psychology,* **49**, 621-30.

Hevner, K. 1935a. The affective character of the major and minor modes in music. *American Journal of Psychology,* **8**, 101-42.

Hevner, K. 1935b. Expression in music: A discussion of experimental studies and theories. *Psychological Review,* **47**, 186-204.

Hunter, J., 1979. *The Fiddle Music of Scotland* (Edinburgh: T and A Constable Ltd)

Husain, G., Thompson, W.F. and Schellenberg, E.G. 2002. Effects of musical tempo and mode on arousal, mood and spatial abilities. *Music Perception*, **20** (2), 151-171.

Jones, M.R., 1976. Time, our lost dimension: toward a new theory of perception, attention, and memory, *Psychological Review*, **83**, 323-355.

Jones, M.R., 1987. Dynamic pattern structure in music: recent theory and research, *Perception and Psychophysics*, **41**, 631-634.

Jones, M.R., 1990. Learning and the development of expectancies: an interactionist approach, *Psychological Review*, **96**(3), 459-491.

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

Juslin, P.N. 1997. Perceived emotional expression in synthesised performances of a short melody: Capturing the listener's judgement policy. *Musicae Scientiae,* **1**, 225-56.

Kamenetsky, S.B., Hill, D.S. and Trehub, S.E. 1997. Effect of tempo and dynamics on the perception of emotion in music. *Psychology of Music*, **25** (2), 149-160.

Lapidaki, E., 2000. Stability of Tempo Perception in Music Listening, *Music Education Research*, **2**(1), 25-44.

Large, E.W., and Jones M.R., 1999. The dynamics of attending: How people track time-varying events, *Psychological Review*, **106**(1), 119-159.

Large, E.W., and Palmer, C., 2001. Perceiving Temporal Regularity in Music, *Cognitive Science*, **26**, 1-37.

Levitin, D.J., and Cook, P.R., 1996. Memory for musical tempo: Additional evidence that auditory memory is absolute, *Perception and Psychophysics*, **58**, 927-935.

London, J., 2002. Cognitive constraints on metric systems: some observation and hypotheses, *Music Perception*, **19**(4), 529-551.

Malapani, C., Rakitini, B.C., Meck, W.H., Levy, R., Deweer, B., Dubois, B., Agid, Y., and Gibbon, J., 1998. Coupled temporal memories in Parkinson's disease: a dopamine-related dysfunction, *Journal of Cognitive Neuroscience*, **10**, 316-331.

Martin C, 2002a *The Fiddle Music of the Scottish Highlands; Ceol Na Fidhle* **1 and 2** (Isle of Skye: Taigh Na Teud)

Martin C, 2002b *The Fiddle Music of the Scottish Highlands; Ceol Na Fidhle* **3 and 4** (Isle of Skye Taigh Na Teud)

Ornstein, R.E., 1969. *On the experience of time*. Penguin, Middlesex, England.

Perbal, S., Deweer, B., Pillon, B., Vidailhet, M., Dubois, B., and Pouthas, V., 2005. Effects of internal clock and memory disorders on duration reproductions and duration productions in patient with Parkingson's disease, *Brain and Cognition*, **58**(1), 35-48.

Penton-Voak, I.S., Edwards, H., Percival, A., and Wearden, J.H., 1996. Speeding up the internal clock in humans? Effects of click trains on subjective duration, *Journal of Experimental Psychology: Animal Behaviour Processes,* **22**(3), 307-320.

Peretz, I., Gagnon, L., and Bouchard, B., 1998. Music and emotion: perceptual determinants, immediacy, and isolation after brain damage, *Cognition*, **68**, 111-141.

Poynter, W.D., 1989. Judging the duration of time intervals: A process of remembering segments of experience. In I. Levin and D. Zakay (Eds.), *Time and human cogntion: A life-span perspective*, 305-321, Amsterdam: Elsevier.

Proffitt, D.R., and Kaiser, M.K. 1995. Perceiving events. In W. Epstein and S.J. Rogers (Eds.), *Perception of space and motion: Handbook of perception and cognition*, 2nd edition: San Diego: Academic Press.

Quinn, S. and Watt, R. 2006. The perception of tempo in music, *Perception*, **35**(2), 267-280.

Rakitin, B.C., Stern, Y., and Malapani, C., 2005. The effects of ageing on time reproduction in delayed free-recall, *Brain and Cognition*, **58**, 17-34.

Rickard, N.S., 2004. Intense emotional responses to music: a test of the physiological arousal hypothesis, *Psychology of Music*, **32**(4), 371-388.

Repp, B.H., 1999a. Relationships between performance timing, perception of timing perturbations, and perceptual-motor synchronization in two Chopin preludes, *Australian Journal of Psychology*, **51**(3), 188-203.

Repp, B.H., 1999b. Detecting deviations from metronomic timing in music: effects of perceptual structure on the mental timekeeper, *Perception and Psychophysics*, **61**(3), 529-548.

Repp, B.H., 1999c. Control of expressive and metronomic timing in pianists, *Journal of Motor Behaviour*, **31**(2), 145-164.

Repp, B.H., 1998. A microcosm of musical expression. I. Quantitative analysis of pianists' timing in the initial measures of Chopin's Etude in E major, *Journal of the Acoustical Society of America*, **104**(2), 1085-1100.

Repp, B.H., 1997a. The effect of tempo on pedal timing in piano performances, *Psychological Research*, **60**, 164-172.

Repp, B.H., 1997b. Variability of timing in expressive piano performance increases with interval duration, *Psychonomic Bulletin and Review*, **4**(4), 530-534.

Repp, B.H., 1997c. Expressive Timing in a Debussy prelude: A comparison of student and expert pianists, *Musicae Scientiae*, **1**(2), 257-268.

Repp, B.H., 1995. Expressive timing in Schumann's "Träumerei:" An analysis of performances by graduate student pianists, *Journal of the Acoustical Society of America*, **98**(1), 2413-2427.

Repp, B.H., 1992. Probing the cognitive representations of musical time: Structural constraints on the perception of timing perturbations, *Cognition*, **44**, 241-281.

Repp, B.H., 1994a. Relational invariance of expressive microstructure across global tempo changes in music performance: An exploratory study, *Psychological Research*, **56**, 269-284.

Repp, B.H., 1994b, On determining the basic tempo of an expressive music performance, *Psychology of Music*, **22**, 157-167.

Repp, B.H., 1994c. On determining the basic tempo of an expressive music performance, *Psychology of Music*, **22**, 157-167.

Repp, B.H., 1992a. A constraint on the expressive timing of a melodic gesture: Evidence from performance and aesthetic judgment, *Music Perception*, **10**(2), 221-242.

Repp, B.H., 1992b. Diversity and commonality in music performance: An analysis of timing microstructure in Shumann's "Träumerei:", *Journal of the Acoustical Society of America*, **92**(5), 2546-2568.

Repp, B.H., 1990a. Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists, *Journal of the Acoustical Society of America*, **88**(2), 622-641.

Repp, B.H., Windsor, L., and Desain, P., 2002. Effects of tempo on the timing of simple musical rhythms, *Music Perception*, **19**(4), 565-593.

Rigg, M.G. 1940. Speed as a determiner of musical mood. *Journal of Experimental Psychology,* **27,** 566-571.

Rigg, M.G. 1939. What features of a musical phrase have emotional suggestiveness? *Publications of the Social Science Research Council of the Oklahoma Agricultural and Mechanical College,* No.1.

Rigg, M., 1937. Musical expression: an investigation of the theories of Erich Sorantin, *Journal of Experimental Psychology*, **21**, 442-455.

Schellenberg, E.G., Krysciak, A.M., and Campbell, J. 2000. Perceiving emotion in melody: Interactive effects of pitch and rhythm. *Music Perception*, **18** (2), 155-171.

Scherer, K.R. and Oshinsky, J.S. 1977. Cue utilisation in emotion and attribution from auditory stimuli. *Motivation and Emotion*, **1** (4), 331-346.

Schubert, E. 2001. Continuous Measurement of Self-Report Emotional Response to Music. In P. Juslin and J. Sloboda (Eds.), *Music and Emotion: Theory and Research.* (pp. 393-414). Oxford: Oxford University Press.

Schulze, H-H., 1978. The detectability of local and global displacements in regular rhythmic patterns, *Psychological Research*, **40**, 173-181.

Smith, J.D., 1987. Conflicting aesthetic ideas in a musical culture, *Music Perception*, **4**, 373-392.

Thomas, E.A. and Brown, I. Jr., 1974. Time perception and the filled-duration illusion, *Perception and Psychophysics*, **16**(3), 449-458.

Thomas, E.A., and Weaver, W.B., 1975. Cognitive processing and time perception, *Perception and Psychophysics*, **17**, 363-367.

Timmers, R., Ashley, R., Desain, P., and Heijink, H., 2000. The influence of musical context on tempo rubato, *Journal of New Music Research*, **29**(2), 131-158.

Trehub, S., and Thorpe, L.A., 1989. Infants' Perception of Rhythm: Categorization of Auditory Sequences by Temporal Structure, *Canadian Journal of Psychology*, **43**(2), 217-229.

Treisman, M., 1963. Temporal discrimination and the indifference interval: Implications for a model of the internal clock, *Psychological Monographs*, **77**(13), Whole No. 576.

Tse, P., Intriligator, J., Rivest, J., and Cavanagh, P., 2004. Attention and the subjective expansion of time, *Perception and Psychophysics*, **66**(7), 1171-1189.

Vos, P.G., Assen, M.V., and Fraňek, M., 1997. Perceived tempo change is dependent on base tempo and direction of change: Evidence for a generalized version of Schulze's (1978) internal beat model, *Psychological Research*, **59**, 240-247.

Walker, J.T., and Scott, K.J., 1981. Auditory-visual conflicts in the perceived duration of lights, tones, and gaps, *Journal of Experimental Psychology: Human Perception and Performance*, **7**, 1327-1339.

Wearden, J.H., 2001. Internal Clocks and the Representation of time. In Hoerl, C. and McCormack, T. (Eds.) *Time memory: Issues of philosophy and psychology*. Oxford University Press: Oxford.

Wearden, J.H., 1999. "Beyond the fields we know…": exploring and developing scalar timing theory, *Behavioural Processes*, **45**, 3-21.

Wearden, J.H., Edwards, H., Fakhri, M., and Percival, A., 1998. Why "sounds are judged longer than lights": Application of a model of the internal clock in humans, *The Quarterly Journal of Experimental Psychology*, **51B**(2), 97-120.

Wearden, J.H., and Penton-Voak, I.S., 1995. Feeling the heat: body temperature and the rate of subjective time, revisited, *Quarterly journal of experimental psychology*, **48B**, 129-141.

Zakay, D., and Block, R.A., 1997. Temporal Cognition, *Current Directions in Psychological Science*, **6**, 12-16.

APPENDIX 1

Extracts 1 to 16 are from Hunter (1979); extracts 17, 18, 19, 21, 24 and 25 from Martin (2002a) and extracts 20, 22 and 23 from Martin (2002b).

1. Angus Camerons complements to Alex Webster
2. Bonnie Glenfarg
3. Lament for the death of Rev. Archie Beaton
4. Lament of Flora MacDonald
5. MacPherson's Rant
6. Mairi Bhan Og Mary, young and fair
7. Mr Garden Troup's Farewell to France
8. Mrs Hamilton of Pentcaitland
9. Mrs Helen N. Robertson
10. Roslin Castle
11. Sitting on the stern of a boat
13. The Duchess of Manchester's farewell to the highlands of Scotland
14. The fallen chief
15. The Marchoness of Huntly's favourite
16. The Marquee of Huntly's snuff mill
17. The Kilworth Hills
18. Miss Drummond of Perth
19. Loch Torridon
21. Queen Victoria's Diamond Jubilee
22. Stornoway Castle

24. Bob Steele

25. Lochanside