Some robust higher level percepts for music

Roger Watt and Sandra Quinn

Department of Psychology, University of Stirling, Stirling FK9 4LA, UK

Abstract

Studies are reported to explore a range of higher-level percepts in music. Participants were asked to make two-alternative forced choice judgements of extracts of instrumental music on various dipole categories, such as happy/sad or male/female. The consistency with which each stimulus was judged on a response category across listeners provides an indication of the extent to which the musical percept can be mapped reliably onto that dimension. High consistency would suggest that the response category is related to one of the natural perceptual dimensions for music.

We found very high consistency (90%+) for various response categories normally used as descriptions of people (such as male/female and happy/sad). Other types of response category gave much lower consistency. Perhaps our participants are experts in making fine distinctions in person-related categories for almost any stimulus. We tested this with a control experiment where food-stuffs replaced the musical stimuli. We did not find high agreement for person-related categories. The differences between responses to music and food were highly statistically significant.

Introduction

Listeners tend to agree when judging one piece of music as happy and another as sad, at least within a single culture. The minimum that this means is that some salient aspect of the music is being related in the same way by listeners to some aspect of happiness or sadness.

Every piece of music has multiple perceptual qualities and different pieces differ in these qualities. Similarly, the adjectives happy and sad refer to concepts that differ in a number of different ways. When a listener judges music as happy or sad, they form a relationship between the two sets of qualities. In principle, the task faced by the listener is to choose a combination of perceptual qualities from the stimulus and to map this onto a combination of conceptual qualities from the response labels. So, the musical percepts might vary on brassiness and the concept sad and happy might vary on forcefulness. The participant could then simply decide that high brassiness resembles high forcefulness (and low brassiness resembles low forcefulness), and use happy and sad as surrogate responses for brassy or not.

It would be open to different listeners to choose different perceptual or different conceptual dimensions and to choose different response mappings. Responses with high inter-listener agreement only arise when listeners use the same dimensions and the same mapping. The term intersubjective percept will be used to describe this state of affairs where listeners make judgements with high degrees of agreement. The most widely studied intersubjective percept in music is emotion.

Since music is an open category, always growing and changing, it is logically impossible to limit what emotion percepts might be possible, either by empirical means or by theoretical means. However, an important principle to emerge is that there are various sources of emotion within music. Sloboda and Juslin (2001) divide these sources into two fundamental types. Intrinsic emotion is where the source is entirely within the music itself. For example, familiarity with music leads to a set of expectations about the behaviour of music; where these expectations are violated or fulfilled, the listener will experience different effects (Meyer, 1954). Extrinsic emotions are where the music is perceived with respect to something extra-musical. Extrinsic emotions may be iconic (see Dowling and Harwood, 1986) – so loud music may sound threatening. Extrinsic emotions may also be entirely associative, such as in national anthems.

Several studies have explored the range of emotions that listeners perceive within music, using a range of different types of stimuli and response formats. The results are summarized in Gabrielsson and Lindström (2001). There are three types of result.

The first type of result lists the types of adjective chosen by listeners as appropriate responses to musical stimuli. The general result is captured in the adjective circle of Hevner (1936, 1937), which comprises a self-returning sequence of dignified, sad, dreamy, serene, graceful, happy, exciting, vigorous. There is ample evidence for each of these, using various forms of stimulus (real music, Behne, 1972, Kastner and Crowder, 1990; artificial tone sequences, Motte-Haber, 1968, Gabriel, 1978; modified music, Juslin, 1997, Lindstrom, 1997; specifically composed music, Thompson and Robitaille, 1992) and various types of stimulus manipulation (rhythm, Motte-Haber, 1968, Gabrielsson, 1973; tempo, Juslin, 1997; mode, Crowder, 1985; timbre, Juslin, 1997).

The second type of result relates the responses made to an underlying continuous space. Using a method called the semantic differential Osgood et al(1957) had subjects rate a stimulus on a range of different dimensions stretched between polar opposite adjectives, such as happy – sad. The mean rating on each dimension is then a measure of the position of the stimulus on the dimension in question. The vector of positions on the various dimensions is then taken to be the perceptual measure of the stimulus. Factor analysis tends to find three underlying dimensions behind the pattern of responses: valence (pleasant-unpleasant), activity (active-passive) and potency (strong-weak). For musical stimuli, Scherer and Oshinsky (1977) had listeners rate synthesised tone sequences on these three dimensions and explored the relationship between acoustic properties of the stimuli and the ratings. The semantic differential has been used to study the semantic structure of actual music (Hacohen and Wagner, 1996) with broadly similar results. They used extracts from Wagner's Der Ring des Nibelungen and listeners rating responses on eight bipolar adjective scales. Three scales produced high ratings: 'sadness/joy', 'weakness/strength', 'restraint/impetuosity'. These are clearly closely related to valence, potency and activity respectively.

The third type of result has explored behavioural signs of emotional response. Sloboda (1991, 1992) obtained accounts from listeners of musical situations where music caused bodily responses, such as piloerection. To a degree it was possible to relate specific musical features to specific bodily responses. An important feature of this research is that it does not involve any form of cognitive appraisal and is therefore not contaminated by what listeners know about music.

A different question, less frequently addressed concerns the range of other higher level features that are intersubjective percepts of music. This is an important question because an explanation of emotion in music might be too narrow to account for the full phenomenon – it might be a special case of a wider explanation for music.

There are studies that have considered the possibility that music might give rise to perceptions of motion (see Clarke, 2006, chapter 3 for a review). An important link between this and emotion percepts in music is the observation that a common cue to the emotional state of another is the dynamic structure of movements of their face, hands and body.

Gabrielsson (1973) found that certain rhythmic patterns could be readily described by listeners with words that relate to motion: "running", "limping", "flowing" and "crawling". Clynes (1977) and Scherer (1995) relate perceived musical motion to gestural motion of the human body. Todd (1999) suggests that music generates the sensation of ego-motion. Shove and Repp (1995) have emphasised that one obvious source of perceived motion in music is the movements of performers (seen or not). Clarke (2006) observes that music is frequently intended for listeners to make movements to, such as dancing. In all these case, the movements are movements of people.

However, Clarke (2006) also draws attention to non-animate motions that can be perceived in music, so for example, a crescendo – increasing amplitude would indicate motion towards the listener that does not need to be animate at all. Such sounds do, of course have implications for the behaviour of the listener, suggesting avoidance, which is animate.

Intersubjective percepts of music include various different human emotions and motions. Because these are normally properties of people, it is not a large jump to speculate that something in music generates the impression of there being an illusory presence of person. It is then a short further move to speculate that music might have other qualities which resemble those normally perceived in people, traits as well as states. The studies reported in this paper consider the possibility that the intersubjective percepts of music might widely overlap with the percepts formed of a person. Of course, there are trivial cases to be avoided where the music is perceived as a direct product of a person (composer, performer) and responses refer to the person not to the music.

This crude notion of a perceived imaginary person is has been touched on in a more sophisticated way in the literature on the philosophy of musical aesthetics. Kivy (1980) considers the idea that a listener to music may hear a resemblance to a state of mind. Levinson (1996) explores the possibility that music can be heard as if it is the personal expression of emotion by an imaginary individual. The purpose of the present paper is to add empirical support to the general notion, rather than consider the important ramifications of the idea. The interested reader is referred to Scruton (1997).

The basic design is to measure the information content of music by assessing the degree of agreement between participants in their choice from opposing pairs of verbal descriptors (such as happy/sad) as appropriate responses to a number of different pieces of music. Where high agreement exists, then some fairly direct relationship will exist between the perceptual qualities of the music and the conceptual qualities of the verbal label. We ask the simple question: what types of verbal label generate the highest agreement.

A procedural difficulty is that music also has conceptual qualities. People know that music can be happy or sad; this will facilitate listeners in giving music a high rating for happy or sad. It is not a commonplace that music can be male or female, and this knowledge will tend to inhibit listeners from giving music a high rating in either of these regards. For this reason, the present method employs a two-alternative forced choice response (2AFC). If listeners are committed to making one of the two responses "male" or "female" then they are no more inhibited from using one of them than they would be if the response labels were "happy" or "sad".

There are three practical merits to the 2AFC method. The first merit is its extreme simplicity: the task for the listener requires just one partition on the stimulus space, not multiple partitions. The second merit is that 2AFC is a very sensitive method for measuring very small differences in the perceptual consequences of stimuli. The third benefit lies in the statistical power of the binomial distribution of responses which is produced by the 2AFC method. The data can be fully characterized by the binomial distribution and analytical solutions exist for confidence limits, including situations where probability values are very low.

In common with all laboratory studies of music perception, we cannot be sure that the type of listening engaged by our participants is naturalistic. The 2AFC procedure will also influence this. However, the consequences of listening to make 2AFC judgements are unlikely to add information to the stimulus and a re more likely to reduce the information content.

Study 1

Methods

The first study comprised two parts: an experimental part and a control part. The two parts were identical, except that whereas the stimuli in the first part were music, the stimuli in the second part were food.

Experimental stimuli:

Four pieces of music were used as materials for the first study (full details are given in appendix 1). None had any vocal lines, all were real performances (ie not synthesised). Each was played at a clearly audible level. For each, a short 10 second excerpt was played without an abrupt start or end. Two were chosen from Wagner's *Siegfried* and were deliberately selected to present very different effects. The first, *Siegfried's Horn-Call* is a strident energetic brass solo; the second, *Forest Murmurs*, is very gentle and warm scored for strings and soli woodwind. The third and fourth pieces of music were chosen to be much more extreme: an excerpt from a medieval song (performed instrumentally) and an excerpt from a piece of avant-garde 20th century music.

Control stimuli:

Four pieces of food were used as the control stimulus: they have many of the same broad psychological properties as music, without being music. Food is a rich and complex stimulus; it is highly familiar; people have strong likes and dislikes; it varies enormously in its perceptual qualities. In these respects it is very similar to music. However, it is different from music in several important ways: it is not created; it is not expressive; it is not structured hierarchically or temporally.

Participants:

For the experimental part, the participants were first year Psychology students (mean age 17.6 years). People with more than 1 year of formal musical training were excluded. Any who were familiar with any of the extracts of music were excluded from the study. After excluding 8 people, 200 participants remained. A very high number of participants was used to establish any effect magnitude.

For the control part, the participants were 60 young people visiting Stirling University on an open day. With a mean age of 17.1 years, this sample is very similar to the sample for the experimental part.

Responses:

Responses for each stimulus were collected for 12 pairs of adjectives on a paper form. The pairs were chosen from 4 groups: 3 person traits, 3 person states and 6 not person categories. The sequence of pairs down the page, and the polarity left-right of the pairs were randomized from participant to participant and from stimulus to stimulus, to balance out any order effects.

Procedure:

After listening to a music stimulus or eating a food stimulus, each participant was asked to make 12 two-alternative forced choice responses (2AFC), choosing one each from 12 pairs of adjectives.

Results

Of interest is the extent to which participants were in agreement about the appropriate response for any given stimulus. Responses to the left hand adjective (as listed in appendix 1) were scored as -1; responses to the right hand adjective were scored as +1. Agreement level is then the mean score across participants and varies from -1 (all participants responding to the left-hand adjective) through 0 (responses split 50:50) to +1 (all participants responding to the right-hand adjective). A value of ±0.8 corresponds to 90% of participants responding in the same direction.

With a 2AFC design, it is important to distinguish between a bias and sensitivity. If all participants, in response to every piece of music responded “female”, then there would be apparent agreement of 100% because of this bias. To allow for this, information (I) is calculated as the absolute deviation of responses for each response category to one stimulus from the mean response for that category across all stimuli. Information varies from 0 to 1. Chance level responding would yield a value of 0 for information. Invariably responding "female" would lead to an information value of 0. Completely consistent responding "female" to two stimuli and "male" to the other two would yield a value of 1.

The basic agreement levels for the music responses and the food are shown in Figure 1. Each graph corresponds to one stimulus and has the 12 pairs of adjectives as 12 rows in the histogram. The adjectives are arranged so that those with more positive connotations are on the right-hand edge. As can be seen, agreement levels in the music stimuli reach very high values for some adjectives. The equivalent agreement levels for food are not as high.

Figure 1 about here

Figure 2 about here

Figure 2 shows the two values of information for each adjective pair. Treating the food responses as a control, it is possible to compare information in music and food. A simple binomial calculation can be used to assign 99% confidence limits to the data, and ranges of these are shown on the figure as error bars. These confidence limits can be used to determine statistical significance at the 0.01 level.

i) The figure shows that information in all response categories for both types of stimulus is significantly different from zero (chance).

ii) The figure shows that the information level for music is significantly higher than for food in 5 out of the 6 person-related response categories and 2 out of 6 non-person-related categories. The overall difference between music and food is highly significant ($p<10^{-12}$). This is very small, but only because a large number of participants were used.

iii) The figure also shows that the person-related categories of response generate significantly higher levels of information than do the non-person related categories, but only for music stimuli.

With the binomial distribution, it is possible to calculate directly the probability of obtaining a given measured information level, given any expected information level. Table 1 shows the calculated probability of obtaining the information values for person-related categories, on the basis that the expected information level is that given overall by the non-person categories. As can be seen, the main finding is that information levels are significantly higher for music and person-related responses than any other combination of stimulus and response.

Table 1 about here

Discussion

The very high levels of agreement coupled with the large number of listeners suggests a robust phenomenon. However, there are two issues that require exploration before any conclusion may be reached. First, the results of this first experiment only apply to four pieces of music, and so a wider and larger set of stimuli is required. Second, there may be an underlying structure of relations between the adjectives used, either conceptually or musically, which may change the manner in which the results might be understood. In particular, the semantic differential typically results in just 3 independent variables – named as activity (quick/slow), potency (strong/weak) and valence (good/bad). Do these dimensions also determine the present results?

Study 2

The second study uses a much larger and varied range of musical excerpts (24), with a more limited number of participants (50). The first intention is to explore the generality of the finding across a wider range of genres of music. The second intention is to explore relations between the adjectives, by performing a factor analysis of response data to the larger number of stimuli.

Methods

The basic method was similar to the first study.

Participants: 50 Psychology students were the participants, receiving course credit for participation. Any participants with more than 1 year of formal music training or who were familiar with any of the extracts of music were excluded from the study.

Materials: 24 pieces of music were employed as stimuli. None had any vocal lines, all were real performances (rather than synthesised). Each was played at a clearly audible level through closed ear headphones. For each, a short 10 second excerpt was played. The range was all western music, and extended from medieval music, through baroque, classical and romantic to modern popular and art music. It included music of varying modality and atonal music.

Procedure: Each participant attended a separate session. After listening to each piece of music, they were asked to make their responses by completing a form. This form had the same 12 pairs of adjectives as in study 1, randomized in order and polarity as before.

Results

Analysis 1: basic effect

Although the first study found large effects of person-related responses despite averaging data across pieces of music, this does not necessarily mean that every piece of music will yield high levels of information across all such categories. It is possible that 10% of all music will yield an information level of 1 for female/male responses, but the remainder will yield information levels of zero. The averaged information level would be low (0.1), but the conclusion that music cannot provide high information in this response would not be warranted.

The basic results are summarized in Figure 3. The first panel of the figure shows the mean information level for each pair of adjectives taken across participants and stimuli. As can be seen, the person-related categories again show higher levels of information, although the difference is less marked this time. The second panel plots the mean information for person related categories against the non-person categories for each piece of music (represented by a single point). As can be seen, a very high proportion of the pieces have more information for person categories than non-person ones (lie above the positive diagonal). The filled symbols are for the 8 popular music extracts – these behave just the same as the classical pieces. This confirms that the basic finding of study 1 is consistent across various genres of western music. It is interesting to note that the data have a slope that is greater than 1: the advantage of person-related responses over non-person related ones is highest for stimuli that yield the highest overall information level (ie the most expressive).

Figure 3 about here

Analysis 2: synonymy

The person-related response categories are undoubtedly not independent of each other. For example, the 24 agreement scores for gentle/violent and for female/male have a correlation of 0.71, which is high. Given the conventional stereotypes for these words, this is not perhaps surprising. Figure 4 shows a scatter plot for two of the response pairs; each point corresponds to one of the 24 pieces of music. Given that the values for each data point being plotted are binomial variables, it is possible to directly calculate the probability that the two values are derived from the same expected agreement level. The figure shows as filled symbols those cases where this probability is less than 0.005 (a very conservative significance level). Despite the high correlation between female and gentle, there are 5 pieces of music where the two are not remotely similar. These 5 stimuli demonstrate that "female" and "gentle" are not synonyms within music.

Figure 4 about here

The lower panel in Figure 4 shows an analysis of this type for all pairs of response labels. Each cell shows the number of stimuli where the column response category and the row response category are significantly different from each other at the (very low) criterion of $p<0.005$. As can be seen, most response labels are distinctive. There are two responses labels which cannot be distinguished at this criterion: “young/old” and "joyful/sad”.

Analysis 3: no of factors

In the sense of synonymy or interchangeability, our response categories are not identical. However, that is not to say that they are not related. A principal components factor analysis (PCA) on the 24 observations (stimuli) of 6 empirical variables (agreement levels) was conducted to explore further the relations between response categories. PCA assumes that the various observed data are the result of the linear combination (ie summation) of a set of underlying independent factors. Given this assumption, PCA then computes the underlying factors, called principal components. PCA finds factors such that the first factor accounts for the greatest possible amount of variance in the observations, the second accounts for the greatest amount of variance remaining and so on. The total number of principal components equals the number of empirical variables, in this case 6 because we are only analyzing the person-related categories.

The first issue with PCA is to determine the number of principal components that might be required to account adequately for the observed data. The usual method for doing this with normally distributed data is to count only those factors that account for more of the total variance than an equal share, which is essentially arbitrary. The use of binomial data offers a more rigorous method. If we take just the first component, then we can use it alone to reconstruct the empirical data. There will be differences between the original data and this reconstruction and these can be assessed by computing the probability of obtaining the original data given the reconstruction. If this probability is lower than our criterion ($p=0.005$), then we can use the first and second components to reconstruct the data. If the probability of getting our actual data from this reconstruction is still too low, we can add the next component and so on. That method can be used to establish how many components are required to reasonably account for that piece alone. Figure 5 shows the results of this analysis.

As can be seen, 5 components are required to account for all the stimuli even at a conservative criterion of p=0.005.

Figure 5 about here

Analysis 4: individual factors

Finally, we can turn to inspect the full set of individual principal components from the PCA. Each principal component is a set of 6 values, each being the amount of the corresponding empirical variable required to make that component. The 6 components are shown in Figure 6.

Figure 6 about here

The error bars show 80% of the range of component values as obtained by a method based on bootstrap (described in Appendix 3). They are drawn as a guide to how robust the different parts of each factor are. Where the error bar straddles the zero line, then that part is not a reliable feature of this factor and can be disregarded.

Some of the factors are interpretable:

1. The first factor, which accounts for a high proportion of the variance, combines “gentle” “good” “pleased” “joyful” "young" and “female”. This component would appear to be close to the classic variable valence. However, it is not obvious why female is associated with positive valence and male with negative valence.
2. The second factor combines "sad" "gentle" and "old", and appears to be related to the classic variable activity.
3. The third factor provides for independent variation in the female/male dimension. Although "male" might be taken to relate to the classic variable

potency, the other parts of this factor which go with "male" are "old" and "joyful" and neither seems appropriate for potency.

4. The fourth factor provides for independent variation in young and old.
5. The fifth factor provides for independent variation in good and evil.
6. The sixth factor varies mainly in pleased and angry.

General Discussion

Participants were asked to make forced choices from various adjective pairs in response to a range of different pieces of music. The results of each study show high levels of agreement between participants for person-related categories of response to music. What does this mean?

Consider the case of female/male choices. When 95% of participants all choose the same member of the pair, then it is very clear that the music is controlling their choice with high reliability. This does not necessarily mean that the music “is” female or male, just that some salient aspect of the music resembles some aspect of femaleness or maleness for the listener.

Each piece of music has a range of different perceptual qualities. The different pieces vary in these qualities. Similarly, the different members of each adjective pair refer to concepts that vary in a number of different ways. In principle, the task faced by a participant in the 2AFC design is to choose a combination of perceptual qualities from the stimulus and to map this onto a combination of conceptual qualities from the response labels. The measure of information is a measure of the extent to which listeners use the same mapping.

The data show that when the response category is applicable to a person, a high proportion of participants choose the same mapping from stimulus to response, resulting in high information levels. There are three conditions that must be met for this to work.

i) Listeners are highly expert in concepts relating to the qualities of other people, resulting in high reliability and sensitivity to differences in people.
ii) The listeners have a similar level of expertise in judging appropriate qualities of music, resulting in high reliability and sensitivity to differences in music.
iii) Music is perceived to have qualities that map naturally in the same direction for the majority of listeners on to person categories.

The first two conditions are neither surprising nor particularly interesting. Mere exposure over time to variations in people and variations in music is probably enough. The third condition is the only interesting or surprising one: it alone distinguishes music from food in study 1, for example. Salient perceptual qualities of food do not map naturally on to salient qualities of people (although carefully prepared cuisine might).

The difference between information levels for person-related response categories and non-person related categories is important. The failure of music to generate high levels of agreement for some non-person related categories is also due to a failure of the third condition. We are expert in narrow/wide, but there are no obvious and natural mappings between these and the salient qualities of music.

So the simplest conclusion we can reach is that music has, perhaps *inter alia*, higher-level perceptual qualities which can be naturally and reliably mapped onto person qualities. The two analyses of the underlying structure of the responses in the second study both point to this perceptual effect of music having at least 5 or 6 separate variables. This is substantially higher than the typical 3 found by the more normal rating-scale based semantic differential. The power of binomial statistics arising from our use of the 2AFC method underlies this. It is clear that valence, activity and potency do play a large part in the data, but the analysis has identified that these 3 variables are not enough.

It is clumsy to talk of listeners perceiving a person in music. Most of the cues that would indicate the *presence* of that person are not present. Our data only suggest that listeners are able to use the musical cues, presumably temporal and dynamical change, to infer the *qualities* of a person in the absence of that presence. We see a simple analogy with a Gibsonian account: the point becomes easier to express in the language of higher-order invariant and affordance.

Gibson (1966) showed that the dynamic behaviour of the optic array during egomotion provides information about that egomotion irrespective of what the static contents of the optic array might be. In other words, the relative motions of things in the environment and self are specified (without knowing what those things are) by just observing specific types of pattern of change: higher-order invariants. At a simple level, this is similar in logic to our finding. Suppose that the qualities of a person are specified in higher-order invariants: complex temporal and dynamic patterns of change in stimulation arising from that person (visual, acoustic and so on). It then follows that other stimuli with similar temporal and dynamic patterns of change, once the change itself is abstracted from the incidentals of what is being changed, might yield those same higher-order person invariants. Music clearly has complex patterns of temporal and dynamical change. If these patterns of change share appropriate features with the person-quality higher-order invariants, then this would provide an easy account for the present findings.

There is a second Gibsonian style of account of the present finding in terms of affordance. Gibson (1979) emphasised that the most important properties of the environment are those that indicate useful potential interactions between the perceiver and the environment. These potential interactions are called affordances. In the case of a prosaic sense such as vision, the affordances are easy to formulate – each object in our field of view has physical qualities that make that object suitable for certain behaviours. In the case of music, there is no such obvious physical constraint. It is easy enough to see that a pathway affords walking because of the physics of pathways and legs. It is significant harder to see what the physical nature of a piece of music constrains it to afford. There is a useful discussion of this issue in Windsor (2004).

For present purposes, all that can be said is that whatever music affords in a listener, those affordances are closely related to the affordances that arise from interactions with another person. For example, perhaps “female” music affords in the listener a similar mental state as that afforded by interacting with a “female” person, albeit by different routes. Listening to Wagner’s *Forest Murmurs* from Act 2 of *Siegfried* (piece 3 from study 1), affords a set of mental and behavioural dispositions that significantly overlap with those afforded by interacting with a female person, which is not entirely surprising, since the character Siegfried is musing about his mother at this point in the opera.

References

Behne KE, 1972 *Der Einfluss des Tempos auf die Beurteilung von Musik*. (Koln, Germany : Arno Volk Verlag)

Clarke EF, 2006 *Ways of listening: an ecological approach to the perception of musical meaning.* (Oxford: Oxford University Press)

Clynes M, 1977 *Sentics: the touch of emotions*. (New York: Doubleday)

Crowder RG, 1985 “Auditory memory” in *Thinking in sound. The cognitive psychology of human audition*. Eds S McAdams and E Bigand, (Oxford: Clarendon Press)

Dowling WJ and Harwood DL, 1986 *Music Cognition*. (New York: Academic Press)

Gabriel C,1978 “An experimental study of Deryck Cooke’s theory of music and meaning” *Psychology of Music*, **6**, 13-20.

Gabrielsson A, 1973 “Adjective ratings and dimension analyses of auditory rhythmic patterns” *Scand J Psychol* **14**, 244-260.

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

Gibson, J. J. (1979). The Ecological Approach to Visual Perception. Boston: Houghton Mifflin

Hacohen R. and Wagner N, 1997 “The communicative force of Wagner's leitmotifs: complementary relations between their connotations and denotations” *Music Perception* **14**, 445-476.

Hevner K, 1936 “Experimental studies of the elements of expression in music” *American Journal of Psychology*, **48**, 246-268.

Hevner K, 1937 “The affective value of pitch and tempo in music” *American Journal of Psychology*, **49**, 621-630.

Juslin P, 1997 “Emotional communication in music performance: a functionalist perspective and some data” *Music Perception*, **14**, 383-418.

Kastner MP and Crowder RG, 1990 “Perception of major/minor distinction: IV emotional connotations in young children” *Music Perception*, **8**, 189-202’

Kivy P. (1980) *The Corded Shell: Reflections of Musical Expression* (Princeton)

Levinson J (1990). Hope in the Hebrides. In Music, Art and Metaphysics, 336-375 (Ithaca NY).

Lindstrom E, 1997 “Impact of melodic structures on emotional expression” In *Proceedings of the Third Triennial ESCOM Conference, Uppsala, June 1997* Ed A Gabrielsson, pp292-297.

Meyer LB, 1956 *Emotion and Meaning in Music*. (Chicago: University of Chicago Press)

Motte-Haber H. de la, 1968 *Ein betrag zur Klassifikatino meusikalischer Rhythmen.* (Koln, Germany: Arno Volk Verlag)

Osgood CE, Suci GJ and Tannenbaum PH, 1957 *The measurement of meaning*. (Urbana, IL: University of Illinois Press)

Scherer KR, 1995 "Expression of emotion in voice and music" *Journal of Voice*, **9**, 235-248.

Scherer KR and Oshinsky JS, 1977 "Cue utilization in emotion attribution from auditory stimuli" *Motivation and Emotion*, **1**, 331-346.

Scruton, R (1997). *The Aesthetics of Music.* (Oxford: Oxford University Press)

Shove P and Repp B, 1995 "Musical motion and performance: theoretical and empirical perspectives" In *The Practice Of Performance, Studies in Musical Interpretation*. Ed J. Rink (Cambridge: Cambridge University Press)

Sloboda JA, 1991 "Music structure and emotional response: some empirical findings" *Psychology of Music*, **19**, 110-120.

Sloboda JA, 1992 "Empirical studies of emotional response to music" In *Cognitive Bases Of Musical Communication* Eds M. Riess-Jones and S. Holleran, pp33-46. (Washington DC: American Psychological Association)

Sloboda JA and Juslin PN, 2001 “Psychological perspectives on music and emotion” In *Music and Emotion. Theory and Research*. Eds P Juslin and J Sloboda (Oxford: Oxford University Press)

Thompson WF and Robitaille B, 1992 “Can composers express emotions through music?” *Empirical Studies Of The Arts*, **10**, 79-89.

Todd NP, 1999 “Motion in music: a neurobiological perspective” *Music Perception* **17**, 115-126.

Windsor WL, 2004. An ecological approach to semiotics. J Theory of Social Behaviour, 34, 179-198.

Appendices

Appendix 1:

Four pieces of music were employed as experimental stimuli:

1). *Siegfried's horn call* from *Act II Siegfried* (Richard Wagner ca 1865)

2). *Forest murmurs* from *Act II Siegfried* (Richard Wagner ca 1865)

3). *Quant voi en la fin d'este* (Perrin d'Agincourt, ca 1250).

4). *Kontakte fur elektronische: Klange* (Karlheinz Stockhausen 1959)

Four pieces of food were used as control stimuli:

1). Wholemeal bread

2). Cheese

3). Dates

4). Carrot (raw)

Appendix 2:

12 adjectives pairs were used:

1). Person traits

female/male

good/evil

young/old

2). Person states

gentle/violent

joyful/sad

pleased/angry

3). Not person categories

bright/dull

smooth/prickly

sweet/sour

narrow/wide

day/night

dry/moist

Appendix 3:

Like all methods of factor analysis, PCA returns a single set of independent factors. If each of our 24 observations is thought of as being a point in 6 dimensional space, then PCA identifies 6 axes in that space all mutually at right angles to each other. Now any n-dimensional space will have n axes, but how they are jointly aligned with the space is arbitrary. So, for example, the x-axis of 2D space can point in any direction in that space and the y-axis will then point at right angles to it. There is an infinite number of different ways that these 2 axes can be aligned with the space. PCA chooses one specific alignment which places each axis in the direction of greatest remaining variance.

It follows that PCA will find "real" structures only if those structures have nicely graded variances. Where several structures give rise to nearly equal variance in the observations, the components returned by PCA will tend to be arbitrary combinations of those structures. Since one does not know in advance whether this condition applies, there is a difficulty in interpreting PCA factors. Our approach to this is simple. Imagine we could run our experiment a great many times over. We would then accept as "real" any factors which nearly always appeared in the analysis, and discount any others. In practice, this is not feasible, but the statistical process of bootstrap provides a way of simulating this.

We can make 100,000 repeats of the PCA using random re-sampling with replacement of the stimuli. This yields 100000 sets of principal components. Each set has 6 principal components, and each principal component has 6 variables. If we take the first principal component (PC_1) from each we have 100000 estimates of this single component with 6 variables. If the value for the first variable is consistent across all 100000 estimates, then we would conclude that this variable in the first component is a reliable feature of the first component. By inspecting each variable in each component, and then each component in turn, we can assess what are the reliable features of the data.

Tables

Table 1: statistical significance of results from experiment 1

	Music	**Food**
female/male	10^{-7}	0.9
good/evil	10^{-12}	0.05
young/old	0.9	0.01
gentle/rough	10^{-9}	0.9
joyful/sad	0.01	0.7
pleased/angry	10^{-12}	0.9

Table 1 caption: This table shows the probability of obtaining the measured information value for each of the person-related categories on the assumption that the expected information value is the overall information level for all non-person categories. With a significance level of 0.01, information for 5 of the 6 person categories was significantly higher than non-person in music and only 1 in food stimuli.

Table 2: statistical significance for each factor in the PCA

Factors	**female**	**good**	**young**	**gentle**	**joyful**	**pleased**	**p**
Actual	**-0.60**	**-0.80**	**-0.80**	**-0.52**	**-0.80**	**-0.76**	
1	-0.38	-1.04	-0.59	-0.78	-0.56	-0.87	**0**
2	-0.29	-0.93	-0.82	-0.56	-0.78	-0.82	**0.0008**
3	-0.50	-0.86	-0.93	-0.55	-0.68	-0.75	**0.0020**
4	-0.60	-0.85	-0.81	-0.48	-0.78	-0.76	**0.0235**
5	-0.60	-0.80	-0.80	-0.50	-0.79	-0.78	**0.0310**
6	-0.60	-0.80	-0.80	-0.52	-0.80	-0.76	**0.0320**

Table 2 caption: Table 2 shows an analysis of responses to one of the stimuli. The first row (actual) shows the empirical data – the values of agreement measured in the experiment; the next row (labelled 1) shows the reconstruction of these values, using only the first principal component. These calculated values are broadly similar to the actual ones. The final column shows the probability of obtaining the empirical data by chance given the reconstruction as expected. As can be seen, the actual data has a negligible probability of arising from just the first component. The third row (labelled 2) shows the effect of using the first 2 PCs for the reconstruction. The result is closer to the actual data, but still not probable. The table shows that a reconstruction with at least 4 components is required to provide a reconstruction that is not significantly different from then actual data ($p>0.005$).

Figure Legends

Figure 1

This figure shows the responses to the 8 stimuli of experiment 1. Each panel corresponds to one stimulus. The data show the mean response for each response category to that stimulus. Values that are close to +1 or -1, indicate high levels of intersubject agreement. As can be seen, the music tends towards higher agreement than the food.

Figure 2

This figure shows the information content of responses to music (dark bars) and food (light bars) from experiment 1. The error bars show 99% confidence limits, based on the binomial distribution. Music tends to higher information values, especially for the person related response categories (top 6 bars).

Figure 3

The top panel of this figure shows the information in responses to 24 different pieces of music as measured in experiment 2. The information content for person-categories is typically higher than for non-person categories.

The bottom panel shows the relationship between information for person categories of response and non-person categories of response, for each stimulus separately. The darker circles are the popular pieces of music, and the lighter circles are the classical music. As can be seen, there is no appreciable difference for the two genres. Most of our listeners will have been more familiar with popular than classical styles of music.

Figure 4

The top panel of this figure shows the relationship between female/male responses and gentle/rough responses, showing each stimulus separately. The scatter indicates a close correlation, for most stimuli. However, there are 5 stimuli, where the probability that the two responses are drawn from the same underlying binomial distribution is less than 0.005. Consequently, it can be concluded that, although for music, the two descriptions “female” and “gentle” are similar, they are not synonyms.

The bottom panel shows the same result for all response categories. The number in each cell is the number of stimuli, where the two response categories (row and column) are significantly different ($p<0.005$). The only two responses that are interchangeable, at this (low) level of significance are “young” and “gentle”.

Figure 5

This figure shows the number of stimuli (out of 24) whose pattern of response can be successfully reconstructed (at $p>0.005$) as a function of the number of factors used. For example, 3 factors are adequate to account for 17 stimuli, but leave 7 out. At this conservative level of significance, 5 factors are required to reconstruct all stimuli.

Figure 6

This figure shows the 6 factors from the Principal Components Analysis.

The first two are closely related to the classic variables of valence and activity. The other factors are less simple.

Figures

Figure 1: basic results for experiment 1

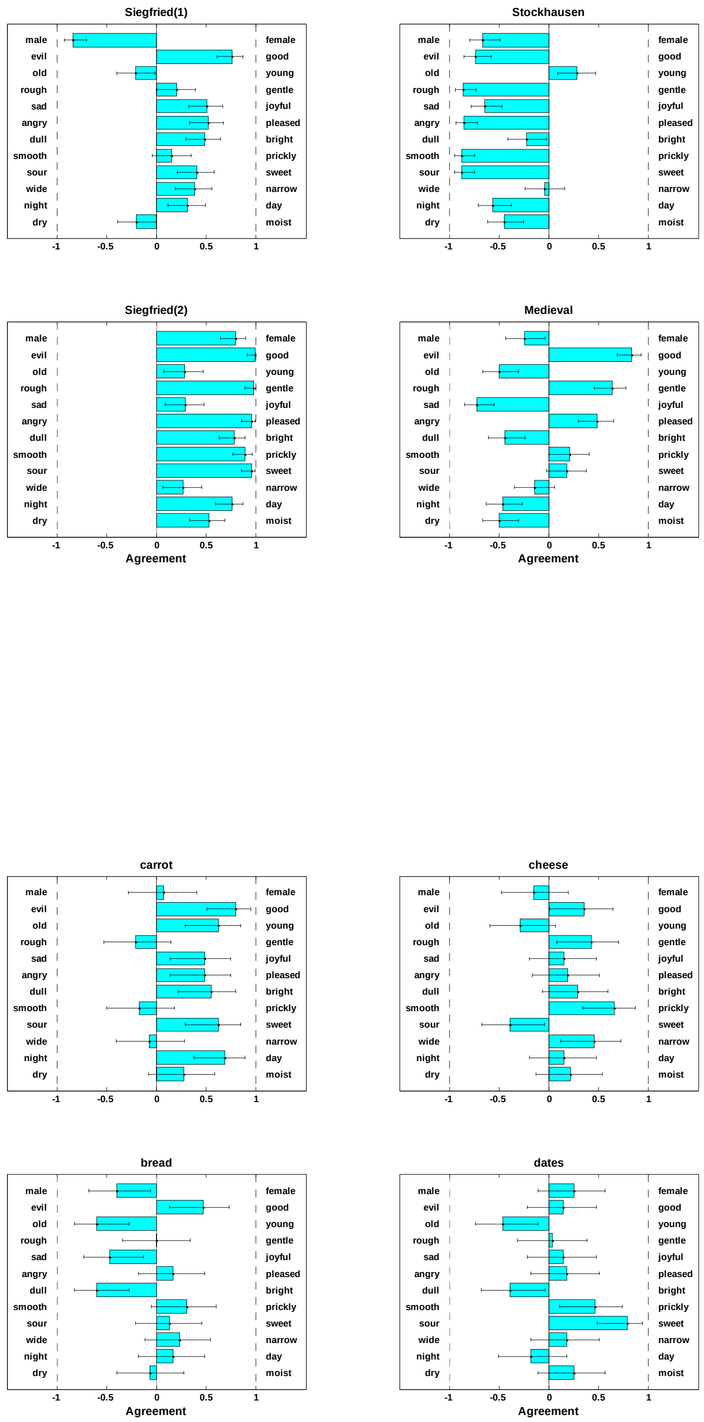


Figure 2: information levels for experiment 1

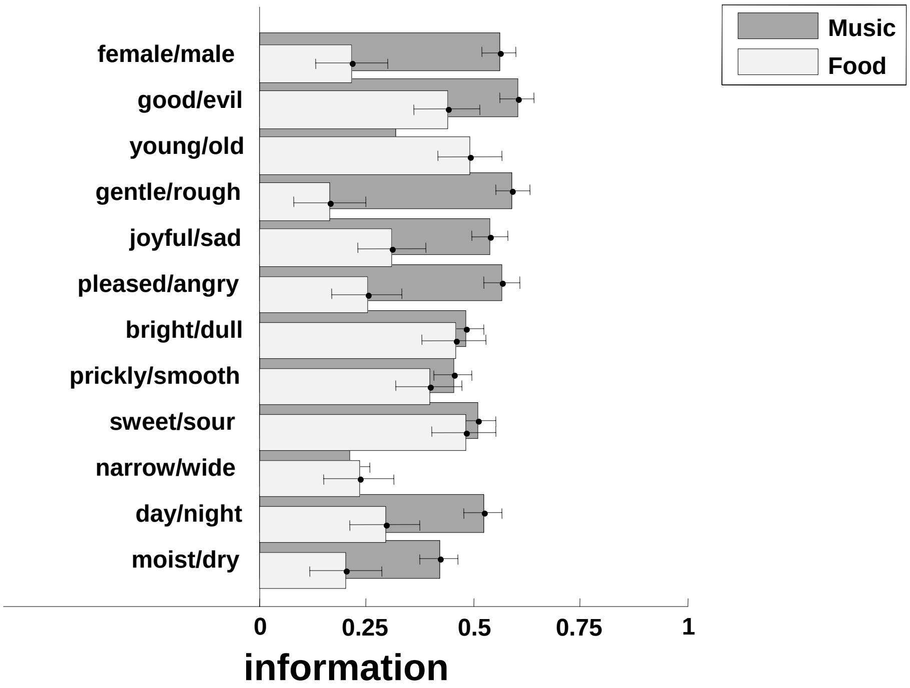
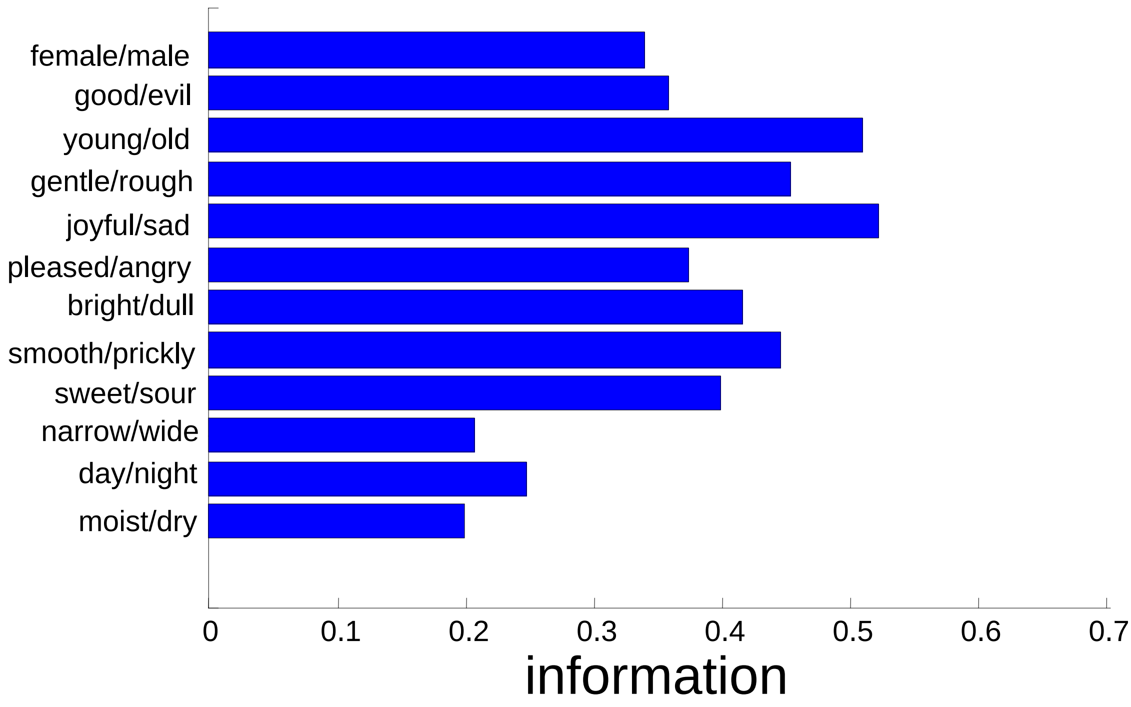


Figure 3: information levels for experiment 2



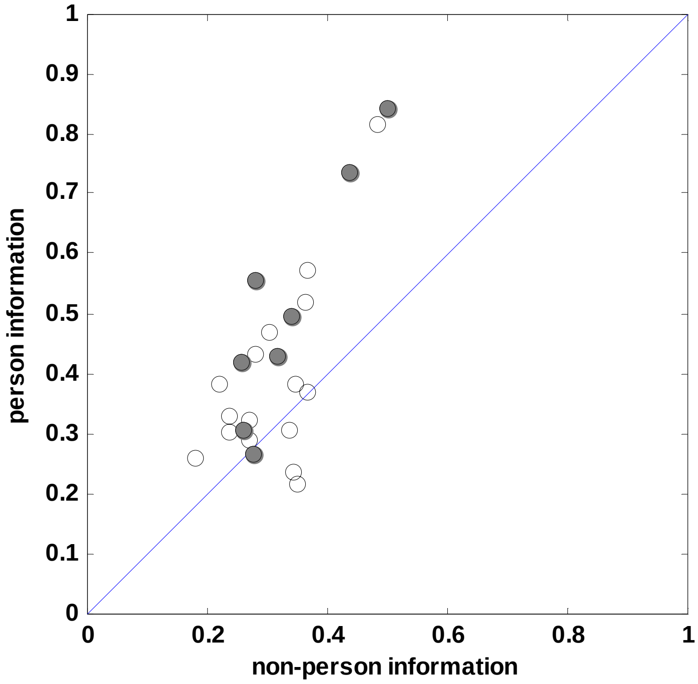
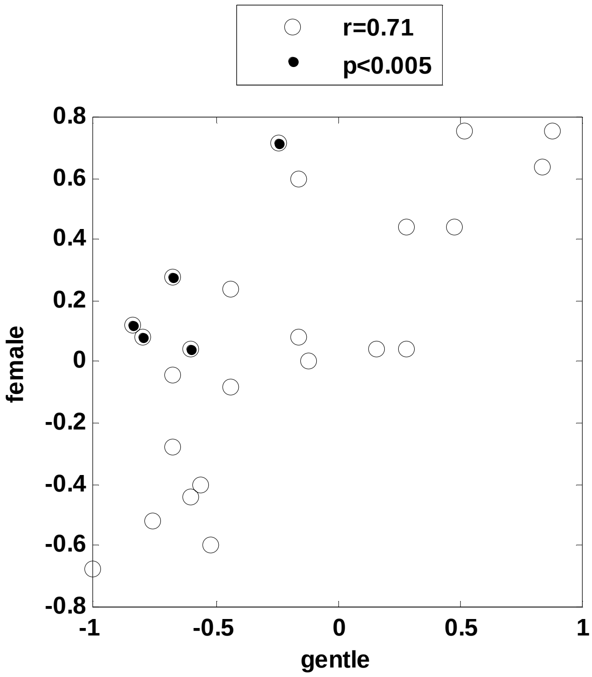


Figure 4: correlational structure in responses for experiment 2



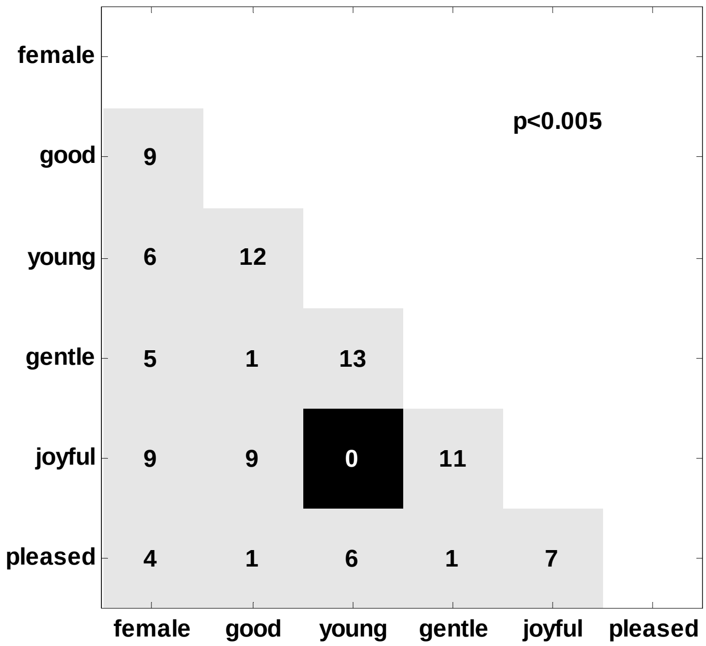


Figure 5: no of required factors in PCA

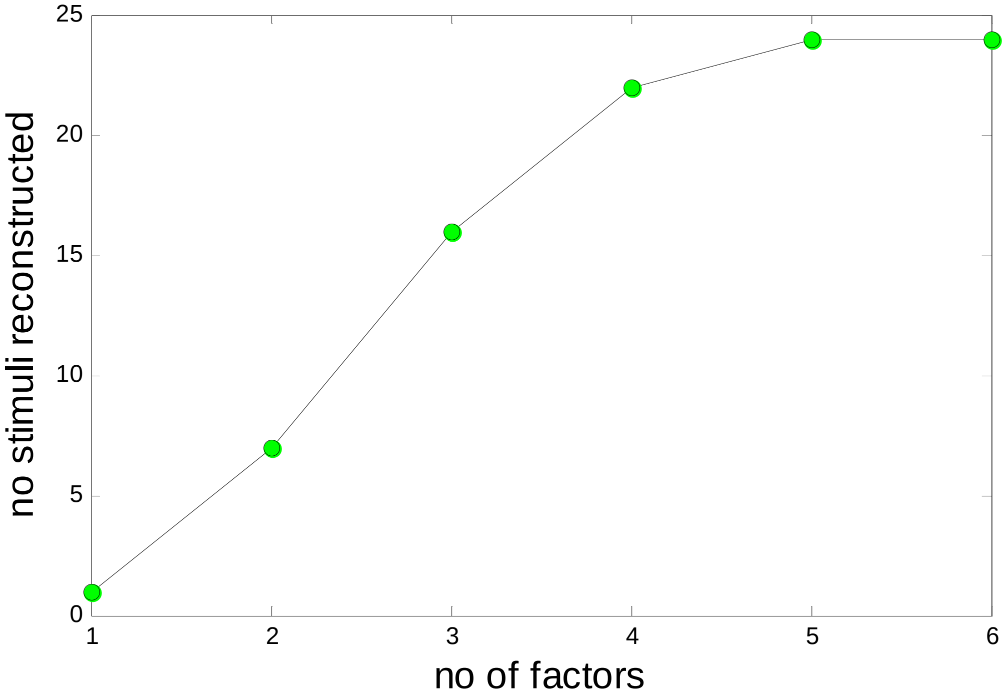

25
20
15
10
5
0
1
2
3
4
5
6
no stimuli reconstructed
no of factors

Figure 6: the 6 factors from PCA for the data of experiment 2