Joint Attention to Music

Tom Cochrane

This paper contrasts individual and collective listening to music, with particular regard to the expressive qualities of music. In the first half of the paper a general model of joint attention is introduced. According to this model, perceiving together modifies the intrinsic structure of the perceptual task, and encourages a convergence of responses to a greater or lesser degree. The model is then applied to music, looking first at the silent listening situation typical to the classical concert hall, and second the noisy listening situation typical to rock or jazz concerts.

Philosophical and psychological studies of musical expression typically focus on the solitary listener. Yet most of our experiences of music are socially orientated. Cross-culturally, music is used to accompany all manner of social rituals from weddings to funerals, festivals, religious services, healing ceremonies, games, and storytelling. In addition, all societies use music to dance to, either as part of rituals or for its own sake. My aim in this paper, then, is to look at the ways in which listening to music as a group differs from listening to music on one's own, particularly in regard to the expressive properties of music.

The most relevant issue in the philosophical literature here concerns the nature of joint attention. Joint attention is initially defined as when two or more people are mutually aware that both are attending to some object in the environment. This is typically signalled by co-ordinated orientation towards the stimulus or communicative gestures such as pointing. Research is also mainly focused on infant joint attention, because engaging in this behaviour is often regarded as a landmark in the infant's ability to understand other minds. It is claimed, for instance, that joint attention provides the foundation for all manner of co-operative activities involving mind reading, most notably the acquisition and use of language.[1]

However, it is the mature form of joint attention that is most applicable to music appreciation. So whilst it is instructive to see the roots of mature joint attention in its infant form, I devote more space to the analysis of a general theory of joint attention. In particular, I am interested in John Campbell's claim that we should analyse joint attention as not merely attending to something, plus being aware of the other person, but that we share the experience of seeing in a more fundamental way.[2] I qualify Campbell's account somewhat, and then apply this idea to the case of listening to music. This paper is thus divided into two halves. In the first half I discuss joint attention in general, and in the second half I apply my account to the case of music. Overall I want to examine whether joint attention to music can cause a convergence of responses, or a mutual fixing of expressive properties. Yet I am not just interested in whether we can agree in our perceptions so much as whether we can listen to the music *as a group*. That is, to what extent when listening together our perceptual activities are integrated or interdependent.

---

1 For example, Michael Tomasello, *The Cultural Origins of Human Cognition* (London: Harvard U.P., 1999).

2 John Campbell, *Reference and Consciousness* (Oxford: Clarendon Press, 2002). See also 'Joint Attention and Common Knowledge' in N. Eilan, C. Hoerl, T. McCormack, and J. Roessler (eds), *Joint Attention: Communication and Other Minds* (Oxford: Oxford U.P., 2005), pp. 287–297.

British Journal of Aesthetics Vol 49 | Number 1 | January 2009 | pp. 59–73 DOI:10.1093/aesthj/ayn059

Joint Attention

Perhaps the most interesting feature of infant joint attention is that it reveals just how *minimal* the requirements for joint attention are. Joint attention is believed to be established in infants between the time they are able to follow another's gaze or pointed finger towards some third object in the room (typically around nine months), and when they later initiate this kind of interaction (by around twelve months).[3] This demonstrates a degree of sensitivity to the mental intentions of the other person as well as a willingness to co-ordinate their experiences of the world. Yet in its initial passive stage at least, it is both unnecessary and unlikely that infants understand the distinction between their own intentions and the intentions of others. Instead infant joint attention seems to rely on a more automatic imitation of the adult's behaviour given a context in which they are somehow 'tuned in' to the adult's goals.[4] Then, as they get older, infants learn to check and direct the adult's behaviour in ways that gradually take on the sophistication of the mature case. Despite these developments, however, it seems likely that mature joint attention relies on essentially the same mutual stance that infants are able to establish. Even in adult life we have a strong automatic tendency to follow other people's gazes when they suddenly look around.

We must note, however, that the joint attention an infant has with an adult is not like that between equals. Joint attention is a means by which infants learn from the adult how to attend to the world to begin with, both in terms of when and what they attend to, as well as which properties of that object they are interested in. The way we attend to objects and people then fundamentally shapes the way we experience them. For example, one of the most interesting 'applications' of joint attention is social referencing, when infants look to adults when meeting unfamiliar objects or strangers in order to determine what kind of emotional reaction is appropriate. Infants will avoid objects or people that their mothers show fear towards, not just whilst the mother is present but also when they are on their own later on. Infants show a high degree of specificity in how they correlate the adult reaction to the particular object or person, indicating that the infant is focused on the properties of the object, rather than simply feeling a general sense of unease.[5]

Psychologist Peter Hobson even makes the claim that the main purpose of joint attention is to share emotional reactions towards objects. He claims that one of the reasons autistic children do *not* engage in joint attentional behaviour is because they are generally not responsive to emotion sharing in dyadic interactions.[6] In a similar vein, Johannes Roessler argues that the most significant aspect of this emotion sharing in reference to objects is that it is a form of predication. That is, objects come to acquire properties other than just being

---

3 It is also necessary that the infants return eye contact with the other, ostensibly to confirm that both are indeed attending to the same object. This confirmatory behaviour is important because it proves that joint attention is a truly shared activity, as opposed to just two people looking at the same thing at the same time.

4 See for instance, Johannes Roessler 'Joint Attention and the Problem of Other Minds' and Naomi Eilan, 'Joint Attention, Communication and Mind', both printed in Eilan *et al.* (eds), *Joint Attention*.

5 Dare Baldwin, 'Understanding the Link Between Joint Attention and Language', in C. Moore and P. J. Dunham (eds), *Joint Attention: Its Origins and Role in Development* (Hillsdale, NJ: Erlbaum, 1995), pp. 131–158.

6 Peter Hobson, 'What Puts the Jointness into Joint Attention?', in Eilan *et al.* (eds), *Joint Attention*, p. 190.

worthy of attention, such as being funny or scary. The importance of this is that emotional predication can be effectively right or wrong.[7] The infant may have one emotional reaction to an object that the adult transforms, or 'corrects' with a different emotional reaction. This may be the earliest case of a sense of objectivity, which Roessler consequently claims that joint attention helps to establish.

As such, joint attention allows the infant to begin cohabiting the cultural world, understanding the purpose and meaning that certain objects have to offer on top of whatever natural ways they can be manipulated. Moreover, when infants learn to manipulate the adult's attention in return, it signals their capacity to *negotiate* the meaning of their common environment. Adults understand that we need not automatically agree about what we are attending to. Yet in the first few years of life, infants are not yet able to understand that adults may have false beliefs about the situation or the infant, and that they in turn could have false beliefs about the adult. In contrast, the adult can appreciate the ways in which two people can fail to attend to the same thing in the same way. The adult can shift his attention not just to different objects within the environment, but also to different aspects of those objects, as well as bring to bear a variety of background knowledge that will enable him to conceptualize the object and recognize its functional affordances. So the adult can understand that even if an object lies within another's perceptual field, the other will not necessarily see the same object as they do, or focus on the same aspect of it, or understand it in the same way unless specifically directed to do so.

Adults are more sensitive to what kinds of conditions can upset the mutual awareness involved in joint attention. Yet in both the infant and adult case, jointly attending to something is still a matter of having, or setting up, a framework for attending to the environment.[8] The nature of this framework in the mature case is just more complicated. For instance, adults can manipulate each others' attention in various ways, such as which sensory organ is to be brought to bear on the world, or the degree of urgency involved. More significantly, mature joint attention is infused with a sense of the *normality* of the situation, where objects have conventional ways in which they are attended to, and certain conventional reactions are implied. For instance we normally look at clocks to check the time, and plan our actions accordingly. In this way normality helps to fix the content of what is jointly attended to.

On the other hand, if the object of joint attention is more individual, then a more sophisticated process of mutual alignment towards the object is required. As well as directing your gaze, I may verbally describe various aspects of the object to you, for example saying 'Look at the way the light reflects from that building', perhaps implying an emotional response as well. By reciprocal behaviour on your part, we can then build up a *specific* framework for that object, the mutual negotiation of which gradually fixes the nature of the object, and confirms that we are jointly attending to the same thing.

In this respect the framework is similar to the notion of a shared cognitive environment that Dan Sperber and Deirdre Wilson appeal to as the background for successfully disambiguating the content of communicative intentions.[9] It involves gradually building up a

---

7 Johannes Roessler, 'Joint Attention and the Problem of Other Minds', in Eilan *et al.* (eds), *Joint Attention*, pp. 245–246.

8 Cf. Greg Currie, 'Framing Narratives', in D. Hutto (ed.) *Narratives and Persons* (Cambridge: Cambridge U.P., 2007).

9 Dan Sperber and Deidre Wilson, *Relevance: Communication and Cognition* (Oxford: Blackwell, 1993).

picture about what kinds of information are mutually available (and mutually known to be mutually available) to both of us. This can include not just ordinary facts about the physical world, but also obvious cases of natural meaning such as black clouds indicating the coming of rain, or that certain behaviours indicate certain emotional states. Of course in most cases there will be a degree of uncertainty as to what facts we can reasonably assume other people to possess. Yet on a day-to-day scale, as we continue to interact with other people, we mutually gain all kinds of information about what facts are mutually available. Moreover being able to identify someone as belonging to a particular social group, such as a culture or a profession, will immediately entail all kinds of information about what kinds of facts are shared between normal members of these groups. Particularly relevant to my case, amongst listeners familiar with the Western Classical tradition, certain generic styles of music should immediately signify certain generic emotional states. Hence given similar cultural backgrounds, even listeners who have never met before should find it fairly easy to communicate ideas about the expressive content of a work.

The shared framework is a way for the content of our attentional states to be mutually fixed. Yet establishing a framework is also a matter of mutually structuring the *activity* of attending to the world. Although I may be motivated to attend to the object because it is an interesting stimulus, I am attending to it at this particular time, for this duration, with regard to this or that particular feature (that may be socially meaningful) because you are attending to it as well. For example, when I direct your attention towards the clock, I may be interested in the time because I want to catch the train, but also because I want to synchronize my behaviour or emotional feeling with yours, or to help explain to you the reason why I am rushing around.

In order to mutually structure the activity of attending, some form of mutual monitoring is required that causally influences the way we attend to the object. Monitoring the other can involve periodically exchanging eye contact or any other form of mutually reciprocal behaviour such as verbal exchange or touch, yet it can also be less overt. In normal conditions, I can track another's attention simply by being disposed to notice when he is no longer attending with me, for instance if he walks away or changes the subject. Hence joint attention is characterized by an ongoing *preparedness* to alter the way I attend to something should you direct me to it, as well as being self-consciously aware of the publicly available aspects of my own behaviour as they have the potential to lead your attentional focus. This preparedness and openness should be mutual.

Different occurrences of joint attention will involve different degrees of monitoring, which will accordingly affect what facts are mutually available. The notions of preparedness and openness can also help us to identify a minimal kind of mutual monitoring. This is perhaps most effectively demonstrated with a crude example. Suppose that I happen to be walking along a quiet country lane alongside a complete stranger with whom I never exchange eye contact, nor react to in any discernible way . If I were to then suffer a burst of flatulence, I would feel instant embarrassment (whereas if I was alone I would not). The other person could similarly anticipate an embarrassed reaction on my part (though he may feel no sympathy). The fact of my flatulence, and the appropriate emotional response would be mutually available. I would not even have to actively recall the presence of the other person prior to my embarrassment being triggered. This is because we can both assume that we are both aware of any reasonably salient event in the environment. Given no reason to think that one's

behaviours are unperceivable to the other (that is, they are not using a white stick or listening to headphones), the default position is to anticipate a degree of mutual sensitivity.

In general, when confronting very sudden or salient events, particularly where they impact on everybody around, it is possible to assume joint attention without needing to engage in overt monitoring.[10] The pervasive sense of normality also means that the joint attentional attitude is *always on* to a minimal extent when we are out in public. It is when the situation is more unusual or subtle, or of a more individual impact (for instance, if one finds money in the street) that we need to check each other's reactions more explicitly. We also expect certain normal reactions to situations, which when violated (for instance, if I continued to pass flatus loudly) motivates closer monitoring of joint attention. Yet in any case there is still a preparedness to adjust our responses to the situation according to the responses of others and an assumption that others can respond in kind to us.

The point of all of this is that since this monitoring behaviour is embedded within the activity of attending to world, we should say that it is a *constituent* of the task of attending to the object rather than an additional factor. The mutual awareness involved in joint attention entails basic differences in our perceptual attitudes in terms of the preparedness and openness that I have described. So it is not the case that when jointly attending to *x* that we each have (i) perception of *x* plus (ii) mutual awareness of (i), but rather a perceptual state which is *of both of us perceiving x*.

This is comparable to the view of John Campbell who argues that,

> [J]oint attention is a primitive phenomenon of consciousness. Just as the object you see can be a constituent of your experience, so too it can be a constituent of your experience that the other person is, with you, jointly attending to the object.[11]

One consequence of this view is that it would not be possible for me to have this perceptual state if the other person were not in fact attending to the object. So, for example, suppose that I thought we were both sitting by the lake watching the ducks when in fact you slipped away some time ago. It might be argued here that if the visual experience of attending to the ducks was the same whether you were there or not, then your presence could not partly constitute the perceptual state. However, Campbell argues that we should take a disjunctivist perspective on this case, in the same way as disjunctivist accounts of perception distinguish veridical and hallucinatory perceptions. On this account, when veridically perceiving, the actual object is necessary to determine the content of the perceptual state. Hallucinations in contrast, make the perceiver *believe* that an object is part of the content of the state when in fact it is not. They are two fundamentally different sorts of state. In the same way when I falsely believed that we were jointly attending to the ducks, I took you to be part of the content of my perceptual state when in fact you were not. I was simply not having an experience of joint attention.[12]

---

10 Due to an undeveloped sense of normality, it is unlikely that infants could jointly attend in this minimal manner. Hence the sophistication of the adult case both allows more complex mutual alignments (that is, conversation about the object) as well as less overt monitoring.

11 Campbell, *Reference and Consciousness*.

12 Campbell also argues that only this view of joint attention can provide the common knowledge necessary to engage rationally in co-operative actions.

It is worth clarifying this point about perceptual experience because Campbell's point is slightly ambiguous. We are assuming that I do not literally see the other person, so he is not part of my perceptual experience in this way; he is not one of its objects. Rather he is part of the subject, the person to whom the perceptual experience is happening. But again this is an odd thought. That person is presumably having a perceptual experience of his own, which may be qualitatively similar to mine, but presumably numerically distinct. So how could he be part of the *subject* of *my* experience?

I think we can best characterize it like this: I am having a perceptual experience, and my experience represents a relation between the subject and the object. So part of it is an awareness of myself as relating to a thing in the world. However, in this case, my awareness of the subject is also an awareness of the person sitting next to me. The other person helps to shape the framework of my visual perception. This literally affects the way I take myself to be approaching the object because, first of all, when visually jointly attending, my awareness is structured by a different spatial field. I am aware of myself, the other, and the object as three points on a triangle. This is not purely a spatial experience, however. I also have ideas about what kinds of information are available to the other person, and to some extent what kind of person he is, the kind of information I could equally self-consciously think about myself. So my thoughts about the other person, combined with my awareness of what facts about me are available to him, are a *filter* through which the visual perception is interpreted. Thus I have a different experience of the object when I see it with others.

Then, according to Campbell's general disjunctivist views, I could not be having this experience unless the other person was actually there, in space, next to me. We might concede this much at least. Yet I have only really referred to *thoughts* about the other person. This is all quite compatible with a thoroughly internalist conception of the whole state. And as long as we keep talking about other people being constituents of my *experiences*, the internalist can continue to talk about the thoughts of other people, not the people themselves.

Campbell's quite radical views about experience require considerable independent justification, and so for the sake of this discussion I will adopt a more modest functionalist stance. Let us assume, then, that my visual experience supervenes on my brain state. All information that I experience must be first processed by my brain. This is what allows the internalist to make his claims. However, we can instead say that the *task of seeing* the object is mediated by the other person's task of seeing. So to the extent that I am monitoring the other person, the direction of his gaze, his responses to the object and to me, and so on, my task of seeing is sensitive to him, and affected by him. I look at something because he looks at it; I identify certain features because he does. I have certain emotional responses to it because he does and in all these respects he is equally affected by me. And if he were not there at all (if I was hallucinating, or if he had slipped away unnoticed) I would not really be directing my seeing via his seeing. I would not really be having what is called a joint attentional state. I would be having something else, a regular attentional state, though one that might seem a lot like a joint attentional one. In this way, joint attention is constituted by the actual system of interactions, the shared framework. Adopting Margaret Gilbert's terminology, we could say the two form a 'plural subject' of attention.[13]

13 Margaret Gilbert, *Sociality and Responsibility: New Essays in Plural Subject Theory* (Lanham, MD: Rowman & Littlefield, 2000).

Gilbert's idea here is that the two try to integrate as much as possible the way that they approach the world. In this case, the perceivers integrate their attention to something, both in terms of when and what they attend to, as well as the sorts of properties of that object they are interested in.

The *extent* of monitoring is an important condition here. It allows the tasks of seeing to be more or less integrated. This factor will be important when we compare different sorts of joint attention to music, since depending on the situation, the mere fact of jointly attending (with no additional overt monitoring) can have various affects on the perceptual state. But first let us emphasize that the intrinsic content of my visual state is not partly constituted by the other person. Rather my task of seeing is partly constituted by the other person, what we might call the structure of the visual state.

Sometimes philosophers talk as if the task of seeing were equally the visual state, and since perception is a dynamic process, continually tracking features of the world, filling out details, solving the binding problem, focusing, recognizing, interpreting, and so on, it is often fair to say that this whole process is the visual state. Given this proviso, we could then say the other person partially constitutes the visual state. Yet when we talk about visual *experiences*, the distinction becomes clearer. One could not be engaged in the same task of seeing if one were a brain in a vat. But perhaps one could have the same experience if one were a brain in a vat (or at least one would need to make a separate argument to dispute that claim).

To summarize then, joint attention is defined as mutual awareness of our attending to something. We are then interested in how this state is set up, and how it affects the experiences of the participants. With regards to how it is set up, I claim that joint attention always occurs in a minimal fashion whenever one is in the (obvious) presence of other people—for example, on the street in daylight. With regards to how it affects one's experiences, there is always a minimal sense in which one is aware that one's behaviour could be perceived by others, and that we can assume mutual awareness about certain facts in the environment. I then note that joint attention can vary in intensity, once basic conditions have been satisfied, as a product of how much we monitor each other. This monitoring determines which features of one's experience (of the object) are mutually co-ordinated—one's behavioural response, emotional response, aesthetic response, and so on. Such mutual co-ordination entails that we share the task of perceiving together, filtering our experiences through our awareness of the other, such that the experience is intrinsically altered. Joint attention involves establishing a plural subject of attention, in which a framework for perceiving the world is generated, and in which the actual interactions involved define that state.

Furthermore, as was true in the infant case, a central function of joint attention is to objectively fix the content of attention. So to jointly attend to a clock is to fix one clock that constitutes our two perceptual states and in this way to converge our perceptual states. At the same time, however, various details of our perceptual experiences may differ, such as the specific orientation towards the clock, or whether one of us attends to the second hand and the other does not. But joint attention does not require the *matching* of perceptual experiences beyond a basic level. Rather we can both be attending to the same thing whilst 'filling out' the details of our experiences in different ways. What matters is that we are able accommodate these differences within the shared framework of the joint attentional

experience. Where these differences are expressed we can recognize that they apply to the same thing. This means that we can jointly keep track of the shared object even if whilst one person is looking at the clock, the other is monitoring my gaze or vice versa. At the same time, joint attention can imbue the object with additional social meaning, as something that signifies our relation to each other, as something that affords a co-operative action or as something that arouses a social emotion in us. It is with this in mind that we can now turn to the case of jointly attending to music.

Joint Attention to Music—the Silent Case

To explore the consequences of this view of joint attention on our social experiences of music, I will start with case of silent joint attention. Silent joint attention can be characterized by the following common sort of scenario. In a prestigious city concert hall, an audience of several thousand people chatter quietly amongst themselves. After a while the musicians of the orchestra make a ramshackle entrance, take their seats, and begin to tune up. Finally the conductor strides in and the audience applauds him to his podium. The conductor ritualistically shakes the hand of the leader of the orchestra and the musicians lift their instruments in preparation. The audience gradually becomes completely silent and then at a moment of exactly his own choosing the conductor begins the performance. The disorder of life suddenly coalesces into perfect order. The actions of the orchestra are exquisitely balanced, beyond most other social activities. The audience too has become ordered, uniformly focused on the movements of the conductor and orchestra. During the performance they remain silent and immobile. They hardly even look at each other until the music is finished.

If when participating in audience listening of this kind you take a mental step back from the music and think instead about those around you, you may appreciate the enormous concentration of attention focused on the stage. Concert halls of this kind are designed to direct all eyes onto the conductor (or soloist) at the centre. The audience is still visible, yet any kind of audience noise is considered a nuisance. So the listener is encouraged to ignore the other listeners and concentrate on the music as much as possible. Yet this is certainly a case of joint attention, since the listeners are aware of their mutual participation in a listening experience. However, the minimal degree of mutual monitoring of each other's reactions throughout the performance will limit the extent to which the aural perception of the music is integrated.

Yet even without the freedom to openly comment on the music during its performance, the audience is at least directed towards the same event and may nonetheless be having extremely similar experiences. As in other cases of joint attention there will be *normal* ways in which the music is perceived. At the most basic level the music is to be treated as a piece of music, performed largely for its own sake rather than any other practical purpose. In addition, within the classical repertoire usually performed in concert halls, the audience can commonly expect that there will be a theme to follow, that the music will have a familiar large-scale form and that it will be emotionally expressive.

The problem of course is that pieces of music are multifaceted objects, capable of sustaining all manner of different perspectives beyond a fairly superficial level. One of the most significant ways in which listeners can differ concerns the different levels of expertise

that they can bring to bear on their listening experiences. Some listeners will be familiar with the cultural context in which the piece was composed, some will have heard the same piece before and some will have greater theoretical knowledge about how the music is produced. The result of this expertise is that some listeners will actually be able to hear elements of the music that others cannot.

Yet this consideration should not lead us to abandon the possibility of a shared response to music. Several theories of musical expression argue that music relies on its natural resemblance to emotions for its expressive character.[14] We also see a high degree of agreement concerning expressive content in empirical studies, at least when very generic emotion labels are used.[15] As a result, most expression theorists agree that music possesses definite dispositional properties to express certain emotions to suitably sensitive listeners.[16] So the different responses that listeners have should for the most part be intelligible to each other and grounded in real features of the music.

I also described above how the shared framework for the perception of objects that joint attention generates is one that can accommodate different ways in which different people fill out the details of the experience. So even if one cannot eliminate the different levels of expertise, joint attention still entails that there is a common target at which the listeners all aim. There is a general preparedness to mutually structure the activity of perceiving the object, including the perception of expressive qualities.

Despite this preparedness, exactly how much the *content* of our listening experiences can be co-ordinated depends on how much a listener can be made aware of the reactions of other listeners. The problem of course is that within a silent joint attentional situation, there is neither the opportunity, nor the motivation, to allow overt responses to be expressed and thus negotiated. Yet even in the silent case of joint attention, there is still a sense in which the awareness of being part of a situation like this will impress itself in common ways upon the individual listeners within an audience. When it comes to the emotional content of the music it is important to note that silent joint attention implies a tacit *acceptance* of that emotional content. Sometimes this can be a particularly uncomfortable experience. For instance, imagine listening to a highly sentimental and romantic piece of classical music in a room alone with your boss. Since the emotional content of the music must be commonly assumed to be equally obvious to both listeners, it generates a palpable sense of an emotional 'atmosphere', which given your background knowledge about each other may or may not seem appropriate. This is comparable to the embarrassing flatulence I described earlier. Due to its mutual availability, the emotional content of the music now gains social dimensions.

---

14 For example, Peter Kivy, *The Corded Shell. Reflections on Musical Expression* (Guildford, Surrey: Princeton U.P., 1980); Stephen Davies, *Musical Meaning and Expression* (Ithaca, NY and London: Cornell U.P., 1994); Malcolm Budd, *Values of Art. Pictures, Poetry and Music*. (London: Penguin Books, 1995).

15 Patrik Juslin, 'Emotional Communication in Music Performance: A Functionalist Perspective and Some Data', *Music Perception*, vol. 14, no. 4 (Summer 1997), pp. 383–418.

16 For example, Davies, *Musical Meaning and Expression*; Kendal Walton, 'Listening with Imagination: Is Music Representational?', *Journal of Aesthetics and Art Criticism*, vol. 52, no. 1 (Winter 1994), pp. 47–62; Derek Matravers, 'The Experience of Emotion in Music', *Journal of Aesthetics and Art Criticism*, vol. 61, no. 4 (Fall 2003), pp. 353–363; Jerrold Levinson, 'Musical Expressiveness as Hearability-As-Expression', in Matthew Kieran (ed.), *Contemporary Debates in Aesthetics* (London: Blackwell, 2005), ch. 12.

Furthermore, sitting in a huge concert hall surrounded by thousands of other people immediately lends intensity to any event that occurs in that space. Apart from the vast range of sonic forces that can be achieved, part of the drama inherent in live performance is to witness an extraordinary act of human skill, comparable to watching a tightrope act with no safety net. It is unlikely that the same atmosphere could be achieved if there was only one listener present. The greater sense of tension is dependent on the sheer numerical concentration of attentive states that the mass audience generates. Every event in that space has massively increased social consequences.

In some respects music has an especially high potential for uniting listeners in experience. Just how great this potential is can be appreciated when we note that the important difference between joint attention to music as opposed to joint attention to most other events is that music is so richly expressive of the inner character of emotion. For this reason to involve oneself in the music and allow it to dominate one's sense of stress and flow is to locate a source for the character of one's inner life that is common to the thousands of listeners around you. It is based on this consideration that phenomenologist Alfred Schutz claims that music has a special capacity to align listeners' sense of 'inner time'. Schutz here is drawing on Bergson's notion of inner time or 'durée', in contrast to measured clock time (consider, for instance, the subjective difference between a minute waltz and one minute of a funeral dirge).[17] He states that the reason that music can structure inner time is because, unlike a mathematical proof, music cannot be grasped all at once, but must always be experienced as a gradual revelation stretched out in time.

Moreover, Schutz claims that the structure of the music captures the subjective stream of consciousness as 'an interplay of recollections, retentions, protensions, and anticipations' whereby the listener is continuously reorganizing the sounds that he has heard previously as well as anticipating what is to come. This view is similar to Meyer's theory that music arouses emotions by generating and resolving expectations.[18] And just as in Meyer's theory we may worry that listeners' relative familiarity with the work will significantly influence what expectations they have. Nevertheless, the point is that by being drawn into the same sequence of sounds, the listeners gain a sense of simultaneity with other listeners, in Schutz's terms of 'growing older together',[19] even if they have differing perspectives on that experience. As such, Schutz draws a particular connection to the sharing of inner time and the state of joint attention:

> [T]his sharing of the other's flux of experiences in inner time, this living through a vivid present in common, constitutes what we called in our introductory paragraphs the mutual tuning-in relationship, the experience of the 'we,' which is at the foundation of all possible communication.[20]

Adorno similarly claims that great symphonic performances can 'annihilate . . . the contingencies of the listener's private existence', enabling the communal elation of an audience. Accordingly he complains that playing symphonies on the radio has 'atomized' the audience

17 Alfred Schutz, 'Making Music Together: A Study in Social Relationship', in *Collected Papers II: Studies in Social Theory*, ed. A. Brodersen (The Hague: Martinus Nijhoff, 1971), pp. 159–178.

18 Leonard B. Meyer, *Emotion and Meaning in Music* (London: University of Chicago Press, 1956).

19 Schutz, 'Making Music Together', p. 175.

20 *Ibid.*, 173.

and destroyed the traditional collective aspect of the symphony.[21] Schutz has a different opinion here. He believes that radio listening provides just the same sense of simultaneity as the concert hall, though there may be 'variations of intensity, intimacy and anonymity'.[22] Schutz is mainly interested in how the content of the music must be experienced in a particular temporal way. However, he does not appreciate the extent to which private listening can significantly affect the sense of control over the music. The listener to a radio programme can certainly imagine all the millions of other listeners that might be tuned in to the same music, yet that fact does not impress upon him to such an extent that he will not happily turn down, change stations, or switch off the music, whistle or sing along, or walk into another room for a while. The important difference with the concert hall experience is that a basic level of co-ordination is retained such that we are committed to focusing on the music as long as everybody else is. In contrast, a private listener has far greater control over exactly what he listens to and how he listens to it.[23]

Finally, the concert hall listener will be self-consciously aware of his emotional responses in a way that the radio listener need not be. Joint attention involves a preparedness to notice the responses of others, whether or not such responses are welcome. Certain micro-signals of the emotional arousal of others will be available that may well pass below the radar of conscious awareness. Others will be more obvious, such as the difference between someone sitting in a tense, alert position and another in a more languorous, detached way. Such signals can lead to emotional contagion, in which as a result of unconsciously imitating the expressive behaviour of others, the corresponding emotional state is aroused. We could argue that this is not a case of genuinely shared response, since the participants are unaware of the mechanisms by which they are aroused. Yet the mere fact that such an event would occur within the context of a joint attentional experience entails that the object of any significant response is by default the most salient event in the common environment. So even if listeners are not consciously monitoring the emotional responses of other listeners, they can nevertheless share the task of deciding their emotion, just as they share the task of perceiving the object. In the absence of a distinct private reason for the emotion, the joint attention experience automatically provides a reason, or object for the emotional response, and this is especially true given the normal expectation that musical works are emotionally expressive.

The same is true if one is more consciously aware of the responses of others. And it is the awareness that others can observe one's reactions, and the taboos against any behaviour that might be considered distracting (including exaggerated facial expressions) that could inhibit the concert hall listener's emotional responses, because they inhibit his expressive

---

21 Theodore Adorno, 'The Radio Symphony', in Richard Leppert (ed.), *Essays on Music* (London: University of California Press, 2002), pp. 256–257. Adorno would probably have been mortified by the further 'atomization' that the internet allows.

22 Schutz, 'Making Music Together', p. 174.

23 Although there remain notable experiential differences between listening to a work on the radio and the very same work on a CD. Even joint attention with some imagined mass of anonymous listeners, and its consequential sense of social intensity, could explain why people are sometimes motivated to tune in to the radio to listen to works that they could listen to at anytime, at their own convenience. Cf. the watching of films on television that one already owns on tape. (I am indebted to Andy Fisher for this example.)

behaviours.[24] This inhibition is then compounded if other listeners do not show signs of arousal. If, on the other hand, other listeners *do* show arousal, we can expect the listener to feel more validated in his own responses. In either case we can anticipate a feedback effect. If other listeners show arousal, then the probability that the individual listener will as well is increased, which goes on to encourage the response of others. Likewise the lack of arousal shown by others can dampen the individual's response. Due to the conventional constraints of the concert hall listening scenario, we might expect the dampening effect to occur more often. Nevertheless there is also potential for greater arousal, due to factors such as the tension of the social situation or one's sense of community with other listeners. This would be worth empirically testing, for example by comparing the difference between the emotional responses of listeners when made to listen to music whilst directly facing each other as opposed to facing the source of the sound. Either an inhibition or an intensification of responses could result, depending on the listeners' mutual familiarity or the degree to which they identify with each other. In either case, however, we should predict a greater alignment of responses.

So my claim, then, is that even in this silent case of joint listening, there are a number of factors concerning the normality of the situation and its social conventions which will mutually structure the task of listening to the music. I do not claim that the expressive content of the music is *necessarily* shared, particularly where that content relies on perceiving details that some may lack to expertise to perceive. Nevertheless empirical and philosophical studies suggest that some common recognition of the emotional content of the work is probable. More importantly, there is a *preparedness* to monitor the emotional responses of other listeners and an ability to accommodate differing reactions within the shared cognitive environment, as applicable to a single objective, or rather intersubjective, event. Hence, a significant effect of silent joint attention to music is to intensify the drama of performance, which may accordingly intensify one's reaction to the expressive properties of the music, especially since they gain additional social meaning, generating an almost palpable sense of emotional atmosphere. In these ways our responses to the music's expressive qualities can be interdependently structured.

Thus this situation could never ordinarily be described as a case of mass individual attention to the music. Just as in ordinary cases of joint attention we should not separate out the listening to the music into (i) perception of the music plus (ii) mutual awareness of (i), but rather a perceptual state of us all listening to the music. However, the greatest potential for the integration and convergence of responses cannot be fully realized whilst the audience is unable to openly respond to what they are hearing. It is thus in what I call the 'noisy' case of joint attention that the opportunity for music to generate a deep sense of community in an audience becomes most apparent.

Joint Attention to Music—The Noisy Case

The situation of the concert hall immediately changes when the performance finishes and the audience bursts into applause. Now a listener can fully appreciate the excitement (or

24 The fact that adopting/inhibiting expressive behaviour arouses/inhibits the corresponding emotion has been convincingly demonstrated by over a hundred empirical experiments. See James Laird, *Feelings*: *The Perception of Self* (Oxford: Oxford U.P., 2007).

apathy) of his fellow listeners. If he has found the performance to be of high quality, he will enjoy the sense of agreement in the rapturous applause of thousands. We have at least the beginnings here of the overt negotiation of the experience of music. Yet because the applause is separated from the experience of the work, it cannot track the moment-to moment-fluctuations of music itself. So any new perspective that the other listeners provide can only be treated like a single filter through which the experience of the work is affected in its entirety. For this reason I am mostly interested here in reactions that occur during the performance itself. This includes noisy reactions like applause, cheering, booing, whistling, singing or humming along with the music, finger clicking, emotional exclamations, and verbal commentary. But in addition I consider clear observable behaviours such as foot tapping, nodding, dancing, and explicitly communicative behaviours such as exchanging looks, smiles, grimaces, or placing one's hands over one's ears.

An interviewee in a paper by music psychologist Alf Gabrielson describes a particularly strong example of the kind of noisy joint attention I am talking about:

> The music began before the curtain rose, and you just stood there as semi-paralysed and screaming. . . . Everybody in the audience is exciting each other to a stage next to a climax, and when the artist at last comes on stage he does not have to say more than 'hi' to trigger that climax. It is very much the atmosphere in the audience that gives this concert feeling. . . . One feels so free somehow. At concerts one can dance, jump, scream and sing as much as one wants. You are like a part of it all, not just a spectator. Throughout the whole concert the audience was in total ecstasy. It was the only thing that mattered: the music! . . . You don't think about what you are doing. You do what you feel like without even thinking about it.[25]

There is clearly a massive difference between this situation and the concert hall ritual I presented earlier. When emotional responses are unconstrained to this extent, we are unlikely to get the sense of tension that silent joint attention generates. What we have instead is an orgy of emotional abandonment. It would be hard *not* to get caught up in such a scene. So it is virtually guaranteed that the perceived emotional content and impact of the music will be mutually recognized by the audience.[26] In less extreme situations, it is still the case that enthusiastic or disapproving responses to particular moments in the performance will focus the attention of other listeners onto those passages in order to ascertain the cause of the response (assuming it is not obvious already). Hence an awareness of the responses of other listeners can allow joint attention to particular aspects of the music and its perceived content.

Note also that joint attention is grounded in a general *willingness* to coordinate our experiences of the world with one another. Noisy joint attention not only enables mutual recognition of what others perceive in the music, it also encourages the 'endorsement' of those reactions, that is, mutual agreement about the effectiveness of the music's expression of emotion. In this respect, adult listeners have not moved significantly beyond infant social referencing. There are still strong social pressures to emotionally conform. In fact what is less certain about the noisy

---

25 Interviewee quoted in Alf Gabrielson, 'Emotions in Strong Experiences with Music', in Juslin and Sloboda (eds), *Music and Emotion: Theory and Research* (Oxford: Oxford U.P., 2001), ch 19, p. 437.

26 It could be argued that much of the audience reaction is directed towards the social situation itself. Yet the music is still the foundation and justification for all responses.

case of joint attention is how much it could allow disagreement about the content of the music. Certain responses will be socially endorsed where contrary responses are silenced. The noisy reaction of others will make it abundantly clear which emotion is socially endorsed. So even if one did not agree about character of the music, one would certainly be disinclined to voice such a sentiment. The most likely result would be a sense of alienation from the group.

As mentioned above, engaging in expressive behaviour tends to intensify arousal levels, though the degree to which this occurs will depend on the emotional personality of the listener.[27] Clearly those listeners who already happen to have similar emotional personalities will have closer emotional reactions to such a situation. In many ways, these similarities will be self-selected for: those who enjoy a particular genre of music and its attendant performance situations are already liable to have similar emotional personalities. We can expect alignments along the introvert–extrovert divide, for instance.[28] Because music has such immediate connections with the character of emotions, with particular types of music tending to express distinctive temporal profiles and intensities of emotions, we can expect that those listeners who already enjoy or aspire towards such emotional profiles will be most likely to seek out such performance situations.

Given that in the noisy case of joint attention to music, people express their opinions of the music, we can expect not just agreement about the character of the music, but a strong mutual awareness of this agreement. This sense of agreement will both intensify the listener's enjoyment of the music and their sense of community with the other listeners. Then, because the audience are converging on their interpretation of the music, and expressive behaviour encourages an aroused response, the audience are also more likely to converge on aroused emotional states. The mutual awareness of this intensifies the emotional atmosphere of the situation. None of this is particularly counterintuitive. If two people are cheering loudly in response to a piece of music, the chances are that they both feel the same way about the music, or more precisely, that they feel the same way about the music as it is performed in that particular social context. What is interesting is that the structural features of jointly attending to the situation (as well as the general properties of music) are contributing to this emotional coordination. Moreover, the members of the audience are not simply undergoing similar emotions at the same time. They are using the music to mutually determine their emotional states.

Yet although the perception of the music is mutually structured, and this perception typically involves emotional responses, we must be careful to qualify that the audience are not sharing their emotional states in a strict sense. The *task* of determining one's emotional state is to some extent interdependent, yet this is mostly a matter of either affirming or denying the expressive character of the music. It does not involve detailed control over the character of one's emotion. And so long as the model of sharing involved is that of

---

27 For a review see J. T. Cacioppo, B. N. Uchino, S. L. Crites, M. A. Snydersmith, G. Smith, G. G. Berntson, and P. J. Lang, 'The Relationship Between Facial Expressiveness and Sympathetic Activation in Emotion: A Critical Review, with Emphasis on Modelling Underlying Mechanisms and Individual Differences', *Journal of Personality and Social Psychology*, vol. 42, no. 1 (1992), pp. 110–128.

28 Stelmack and Campbell have shown that, in response to music, introverts are more sensitive to low auditory stimulation and show progressively less sensitivity to higher levels of auditory stimulation (corresponding with their tendency to inhibit higher levels of arousal). Extraverts reveal the opposite trend. R. M. Stelmack and K. B. Campbell, 'Extraversion and Auditory Sensitivity to High and Low Frequency', *Perceptual and Motor Skills*, vol. 38 (1974), pp. 875–879.

reproducing the emotional state of the music, or the emotions of other listeners, there are numerous ways in which such reproduction can be partial or distorted, for the same reasons of differing levels of sensitivity, background knowledge, and background mood that were mentioned above. Hence, although clear audience reactions can go a significant way towards overcoming the subjectively variable nature of individual responses to music, there can be no guarantee that listeners are in fact enjoying exactly similar emotions.

To finish, it is worth noting that one of the best ways for an audience to become intimately involved in the music and thereby with each other is through dancing together. This is because dancing not only allows the listener to physically synchronize with the music, but also with other listeners, ultimately co-ordinating responses towards the musical event as a whole. One particularly interesting example that music anthropologist John Blacking describes is the 'possession dance' of the Venda tribe.[29] Here dancers can apparently achieve trance-like states of absorption in the music by physically co-ordinating their movements with the drumming. Interestingly, however, it can only occur when dancers are surrounded by members of their own cult, presumably people with whom they identify and trust. This highlights how a sense of community with others can enhance, as well as be enhanced by, joint participation in a musical event.

Blacking emphasizes the role of physical movement in experiencing one's relation to others. He states, 'I do not say that we can experience exactly the same thoughts associated with bodily experience; but to feel with the body is probably as close as anyone can ever get to resonating with another person.'[30] Again, Blacking's idea of emotional resonance is based on the reproduction of feelings, in this case an additional reproduction of the dancers' proprioceptive states. Hence although it is guided by a communal source—the rhythm of the drums—we should not claim that an emotional state is shared in any radical sense. The arousal of each listener, though more transparent, is still individual to each. As in the case of joint visual perception, the intrinsic content of each listener's emotional state is not constituted by the other listeners' arousal.

Yet we can still recognize that the more listeners can actively respond to the music, the more their individual responses should converge. By interdependently structuring listeners' perceptual activities, joint attention provides a solid foundation for such convergence. Moreover, when listeners can clearly observe, negotiate, and agree on their emotional responses, a significant part of their experience will involve an awareness that it is the same for others. This will help to validate their emotional reactions, which given overt expression will lead to an intensification of arousal. At the same time, by providing a basic 'we' perspective, joint attention allows many individuals to think of themselves in terms of a group identity. An individual listener in these situations can legitimately think of himself as part of an emotionally bonded group. So as long as he remains tuned in to the attitude of the crowd, he is entitled to prefix his descriptions of his behaviour and attitudes with 'we did', 'we felt'. Although their *intrinsic* emotional states are not shared, joint attention to music defines a plural subject, which listens and responds to the music as a group.

Tom Cochrane
Swiss Center for Affective Sciences
thomas.cochrane@gmail.com

29 John Blacking, *How Musical Is Man?* (London: Faber & Faber, 1976).

30 *Ibid.*, 111.