

Before we are born we are listeners in the womb. Even more so, we are musical listeners, taking in the pitch contours and amplitude structures of the world outside and around us as we softly dream ourselves into being. Listening takes time to develop though, anatomical structures first have to be built and neurological structures developed and tuned to the environment

Not everyone is a listener, some people do not hear and they are as fully human as any other, with intellectual, emotional and social lives as rich as any other. But this paper is about people who can hear – although much of the section on Infant directed talk applies to signed language as well. Similarly for people who may not have a practical visual capability. Our society is far from inclusive and, unfortunately, that is reflected in the empirical work and here as a consequence.

I am going to assume music is what we normally think it, the ‘we know it when we hear it’ argument that Wittgenstein made of games.

At about 7 months i.e. a couple of months preterm, rapid eye movements start and it appears the foetal brain cycles through REM and non-REM activity every half hour or so. What that actually means is up for grabs and, to be honest, is probably not dreaming. Recent work using fMRI indicates the eye movements are a measurable component of visuo-motor integration – the eye movements are linked with the occipital lobe (visual cortex, at the back of the head) and then out to motor areas. What is probably happening is that, at this very early stage, the eyes are being linked into the brain so that eye movements are integrated with head movements and gaze direction and all of that is then incorporated in to the visual cortex proper. We need to be able to tell if what we see is something eg a car moving from left to right or if that movement of the car shape across our retina is actually from a stationary car but our head / gaze is the thing that is moving (the retinal stimulus might be the same either way and only gets resolved visually by information about our head and eye movements being combined with the retinal information)

This listening takes time to develop so that say toward the end of the third trimester a foetus can hear sounds 20-30dB quieter than a foetus at the middle of the second trimester and is sensitive to a much broader frequency range. Backtracking, at about 16 weeks or so, when the foetus is only 10-15cms long and yet to make those first fluttering movements, the major anatomical structures underlying hearing have developed and can hear frequencies in the range of 250Hz to 500 Hz¹. The audible frequency range extends down to 100Hz by week 27, and up to 1000Hz at week 30 and 3000Hz by week 35². Sensitivity to the structure of the auditory environment also develops – discrimination between a pure tone and a broadband noise burst has been observed at 27 weeks and by 35 weeks the foetus is able to discriminate between different pure tones and speech sounds such as [ba] and [bi]³

Tuning our auditory system to the environment, i.e. learning what sounds mean - takes years to develop, for even though phenomenally we hear the noise of an aircraft far off in the distance what is really happening is that the aircraft sends out pressure waves, and those pressure waves eventually reach us, bounce off our shoulders, around our heads and the external ridges and curves of our ears, to be funnelled into the ear canal where they interact with the various structures of hearing. We don't actually hear something far away, we infer from the impact of pressure waves upon our auditory system the distance and identity of whatever it is making the noise. And that identification doesn't come about through hearing alone but through the integration of all the sensory systems, during development for sure, but also across our entire lifespan. For example, if at the age of 23 you attend a country fair and hear for the first time the distinctive sound of a vintage stationary engine you will, possibly and almost certainly for hours if not days, remember both the sound and the vision of that engine running. When presented with a picture of that engine you might bring to the mind's ear the sound of the engine running. Alternatively, hearing something chugging away on a subsequent visit to the fair you might recall and bring to the mind's eye an image of that engine. Vision and audition of the stationary engine are integrated and the presentation of one will naturally give rise to a recollection of the other. Of course the system isn't perfect, we forget and details are lost, but with sufficient repetition – say like a foetus has of the mother's voice – we will have a fairly total and interrelated set of memories from which the presentation of part will give rise to remembrance of the whole. There is some evidence that from the first few days after birth babies are sensitive

The foetal listener bathes in a relatively simple auditory environment. Like traffic in the suburban distance, the regular sounds of the mother's heartbeat, digestion and respiration sit quietly in the background. Less predictable are sounds coming from the outside world, which must first pass through the tissues and intrauterine fluids of the mother's body before they can be heard. The mother's body reduces their volume, particularly the volume of frequencies above 250Hz (around middle C and above). Voices, both male and female, are quite audible to the foetus, but the loudest and most distinctive sound the foetus will hear is the mother's voice⁵ as not only does her voice come via the air, like all other external sounds, but it is also conducted directly through her body. The mother's voice is perhaps twice the volume of all regular sounds reaching the foetus.

In utero, the mother's voice is heard more as sound than as speech due to the filtering of the body reducing the speech signal complexity to the point where even adult listeners cannot understand the speech that reaches foetal hearing⁵. Instead of speech, then, the foetus hears the modulated pitch of vocal phrases, presented with all the pauses, novelty and repetition of normal adult speech – but without the words and syllables.

External sounds have to pass through the air to the mother's body, they then have to transmit through the mother's body to the foetus. "Transmit through" is the trick, it really means that the energy of the sound pressure wave sufficiently vibrates the mother's tissues and fluids such that the waveform can continue. Fluids, like solids, are actually pretty good for that but the elasticity of the tissues absorbs then dissipates the energy of the pressure wave. Think of it like hitting a drum – if the drumskin is tight (i.e. not very elastic) then the drum makes a nice defined loud crack. But if the skin is loose then the sound is not as loud and the energy of the fast hit is smeared out. The sound is not so much a crack as a thump. And if the skin is really loose then there is not much of a sound at all.

The repetition of the mother's vocal dynamics in the complete absence of semantics perhaps presages the astonishing ability of music to be repetitive and listened to hundreds of times without loss of emotional impact. Margulis offers a number of other suggestions⁶ including repetition as a method for encouraging shifts of attention up and down temporal scales eg listening to micro variations in timing, taking note of phrasing. In one experiment she repeated parts of Berio and Carter pieces here and there within the complete works - untrained everyday listeners rated those versions with repetition as more likely to have been composed by a human than a computer. I would think one aspect of musical repetition that is distinct from language repetition is that language acts to use symbols to tell you something – symbols to carry meaning about the world. In contrast music is more about inducing a state that will exist in context with your previous states – and as your previous states will be novel to some extent the musically induced state will always bear a new relationship to you – if only to your attentional systems as they scan for new information from a repetitive input.

These wordless phrases have enough information for a near term foetus to want to listen longer to speech and speech like sounds than listen to unfamiliar music⁷, and for a newly born baby to prefer hearing human speech over synthetic speech⁸. Similarly, newborns who have heard their mother reading or singing to them when still in the womb, prefer hearing the mother reading and singing those stories again over and above hearing those same stories and songs from anyone else⁵⁹. From birth, then, a baby has preferences that carry emotional significance and those preferences have been developed with proto-musical vocal sounds. More broadly, at the very first awakening of consciousness, in utero and before vision, sound is the primary conduit for the emotional signification of the external world.

There is evidence from machine learning that learning is enhanced by starting with simplified training 'sets' and then gradually increasing the complexity of the training material – at the start one wants enough information for the learning system to pick out regularities but not so complex a system as to make the signal too difficult to get a handle on. But we don't really need machine learning experiments to teach us this – our entire lives as learners (and teachers) is to start simple ('with the basics') and increase complexity as the simpler material is learned.

The primacy of sound is not surprising when one know that the auditory system is reasonably mature at birth for basic hearing (although not so much for spatial localisation of sound sources¹⁰) whereas the visual system really requires exposure to the visual world for perhaps months before reasonable visual acuity and understanding is achieved¹¹.

During infancy the child begins the process of using all the different senses to create a coherent world view. This process will continue into early adolescence, with the brain having to make sense of different sensory information, received at different times, from different directions and from a body that changes size and shape¹².

Sound and vision begin to be linked quite early. Newborns look at their mother's mouth when she speaks¹³ and generally associate synchronised sound and vision to the same object. The combination of the fairly basic sound localisation and eyesight available at birth is enough for babies to recognise faces more easily if the face can be seen whilst talking (as against, say, seeing just the face smiling)¹⁰. As babies are often surrounded by multiple smiling faces all talking with various levels of overlap it is probably fairly important for recognizing individuals that they can combine sound localisation with vision to clarify identity.

As the relation between sound and vision develops, babies continue to learn statistical regularities in the environment so that by the age of two months or so they begin to understand rhythm as a form of pattern and start to tell if visual and auditory rhythms are the same or different. Soon they will know the rhythmic patterns of their own language well enough to differentiate language groups by rhythm alone. As development continues infants start to recognise when a rhythmic pattern they have previously only heard is then presented purely as a visual animation¹⁴ and can tell if a melody is new but the rhythm a familiar one¹⁵. By around one year infants start to understand the idea of synchrony across different sensory streams and so can tell, for example, if audio and visual rhythms are synchronised to each other as against just playing simultaneously^{14,16}

The entire field of studying babies really begins with the assumption that vision and sound are integrated. Tests to 'see' if a newborn can notice or understand stimuli are based around the child orienting to the stimulus – whether that is head turning or eye movements in the direction of the stimulus, or even sucking rates as a conditioned response are linking the stimulus modality, motor behaviour and taste.

There is little opportunity to develop spatial hearing in utero as the temporal difference between the sound arriving at each ear is too small to extract phase information and sound entering the womb is presumably heavily diffused.

This strength of association between the senses begins to show in various illusions where one sense influences what another sense seems to be telling us. The McGurk effect is a famous example where what we see overrides what we hear. In the McGurk effect seeing someone speaking a syllable will influence your interpretation of the sound they are speaking – if you see /va/ at the same time you hear /ba/, then you will hear /va/. It takes about 5 months before the visual and auditory systems are sufficiently integrated for a baby to experience the McGurk effect and it isn't really until about 8 or 9 years of age before a child will experience the effect much as an adult would¹⁷.

It is not the case that vision always dominates and over-rides what we hear. Sound can also influence what we see. If two animated dots travel toward each other from opposite sides of a screen they will look to an adult as if they collide and bounce back if there is the sound of an impact when they 'touch'. Without the sound they will appear to pass through each other without touching. Sound resolves the ambiguity in the visual image – we have to make sense of the information and the sound of an impact tells us the circles collided, otherwise the sound is just a very unlikely coincidence. By about 5-8 months babies will come to the same conclusion, the conjunction of vision and sound (and touch etc) will be assumed to make sense in probabilistic terms¹⁸. Things that happen together belong together.

There is little opportunity to develop spatial hearing in utero as the temporal difference between the sound arriving at each ear is too small to extract phase information and sound entering the womb is presumably heavily diffused.

Even if the start of multimodal perception begins in utero we should not overstate the abilities of newborns and infants. Development of multimodal perception continues for 12 years or more after birth. But even with their physical and mental capabilities still to reach maturity, it is not long before babies are toddlers, walking, talking, and making common household objects into toys that they can use in imaginary play alone and with friends.

What seems to be happening through development is we begin with broad and fairly basic capabilities. Those capabilities take time to mature, to become integrated and refined into the appropriate skills for the environment within which our development takes place^{18,20,21}. We learn 'our' way of speaking and hearing language then find it harder to learn to hear and speak the languages of others. We learn 'our' musical scales and rhythms and then find others strange and difficult to listen and dance to. Those refined abilities cut off some experiences but make other experiences far more efficient. They let us estimate what time it is based on the ambient light, to sing in tune, to feel the beat and to understand who might be a friend. And much of that development, particularly social and emotional development, grows out of our experience with sound and a special proto-musical form of language called motherese.

and if the languages have similar rhythms they cannot discriminate eg English and Dutch have similar rhythmic alternation durations whereas English and Japanese don't. Interestingly head movements – as against face movements – are very good independent indicators of prosody down to the syllable level for both adults and infants as young as 8 months

For audio-visual information babies are acting differently to adults. Adults have learned that the sound and vision of someone speaking are linked – there is no real need to learn that in an ongoing basis. Babies on the other hand need to learn that the synchrony is meaningful – they need to look for synchronous events. ERP studies seem to indicate that is what is happening¹⁹ This sort of info shows us we have to be careful in attributing infant behaviour to adult-like capabilities, it may be that babies are seeing similar differences in the world to adults, but for different reasons.

If the child grows up in a language community that privileges the gaze then they will tend to link vision and sound as they learn their language more than children who grow up in language communities that do not use direct gaze as much. Similarly people who are born deaf and first encounter spoken language only through visual channels will tend to greater dominance of visual channels even after cochlear implants – unless those implants are made before about 30 months¹⁷.

Actually infants are better than adults at many perceptual discriminations. As with most cognition we train up for our own environment- which means culture as far as music goes. Hence infants at 6 months can tell if someone mucks up a Bulgarian or a simple Rock rhythm, but adults from Nth America will struggle to tell if the Bulgarian rhythm gets an accent out of place. Bulgarian adults will have no problem with either as the simple rock rhythms can be understood as a subset of their possible rhythms whereas the Bulgarian rhythms are not at all subset of rock. I.e. the Bulgarians have a much broader musical environment as far as rhythm goes. By only 12 months the Nth American infants are enculturated enough to be losing discrimination between different Bulgarian rhythms²².

It is through the spoken language of adults that babies are helped into the world of being and belonging. But it is not the language that adults use with each other that babies first hear. Instead, caregivers across all cultures use a particular type of speech called motherese, babytalk or infant directed speech (IDS). In western cultures, people speak with babies using raised pitch, exaggerated pitch contours, longer pauses and drawn out their vowels. The effect is to give IDS a tuneful or proto-musical quality²⁶⁻²⁸ that weaves music into the very basis of our relations with the world of feeling and meaning. Babies prefer the high pitch of IDS compared to normally pitched adult speech^{29,30}. A positive emotional tone is also important, for while babies prefer comforting speech in their first few months, by 4 months they start preferring happy voices, whether speaking or singing^{31,32}.

IDS doesn't just have higher pitch but also has wider ranging pitch contours which people vary to fit the situation. In the very early months rising pitch is used to initiate and encourage attention and eye contact. Once attention is gained pitch will then rise and fall to maintain contact and a positive emotional tone. Structured pitch movement then is the basis through which interpersonal communication develops and it would appear that it is pitch relations rather than absolute pitch that infants and babies remember. Certainly by 6 months of age infants understand, just as adults do, that melodies involve relative pitch relations rather than absolute pitches³⁴.

Infant directed speech should not be understood as something an adult *does* to a child, or *at* a child either, but as something that develops in the context of the adult-child relationship¹³. Consequently IDS changes over time according to the developmental stage of the child and the adult's recognition of the appropriate way to behave to a child at that developmental stage³⁵. The behaviour of the infant tells the adult how successful their IDS is in achieving the desired and developmentally appropriate outcome. For example, the higher pitch of IDS peaks at about 4 - 6 months or so and then gradually returns to normal adult pitch by the time the child is about two. Raised pitch acts to draw the babies' attention to the language interaction and hold it there but by the time the baby is two it has learnt a few hundred words and is freely engaging through language anyway. There is no longer any need to flag "This is an important language moment" through using a higher pitched voice.

Interestingly in both Western and Carnatic music a greater interval range in melodies is associated with positive emotion²³

It seems that the stress patterns of a language influence the rhythm of material – at least for the French and English classical traditions.²⁴ There is also evidence for a difference in the way tone and non-tone languages use pitch in music as well²⁵. I am not so sure we should be that surprised by this as one would expect that the environment affects learning and through that artistic production. Different people born within the same language groups will share one very common part of their environment – language.

Babies in the first year tend to prefer the unaccompanied voice singing a song over listening to the same song with accompaniment. By 12-18months toddlers are beginning to preference instrumental music that is fast and loud compared to quieter and slower instrumentals³³. This is consistent with the idea that babies and toddlers become more and more interested in is being excited rather than soothed i.e. arousal is the main factor

Besides promoting and sustaining attention, pitch contours and other aspects of vocal tone draw infant attention to situationally appropriate emotions. Earliest infant directed speech is comforting in tone, moving to affectionate and approving when the baby is about 6 months old (Tickling the baby on the tummy and saying with happy excitement “Who’s a good baby? You’re a good baby”) and then becoming more informative by the time the child is 9 months (slowly speaking “Apple, Apple” whilst holding an apple and maybe jiggling it a bit). A gradual change from emotional support to picking out and naming objects of interest underpins ongoing development as the baby becomes firstly secure in its environment and then old enough to need, and to understand, more specific information about the world around it.

A similar pattern of developmentally appropriate interaction³⁸ is apparent in singing to infants - high pitch, sustained vowels, a slow tempo and gliding between pitches. As with speech, infant directed singing allows adult and infant to regulate and synchronise their emotional states and through that build social regulation and bonding⁴⁰. Genuinely responsive engagement between adult and baby is key for holding the babies’ attention so that recorded singing or singing by an unresponsive stranger will not capture and hold an infant’s attention for long⁴¹.

This highly interactive engagement between baby and caregiver creates functional divisions of pitched vocal sound, where pitched vocals are used as a method to direct attention, to label the things of the world, and to express goals, needs and wants. And, most importantly, pitched sound is the primary vehicle through which we learn to regulate our emotions and build our most important social relations.

Situationally appropriate emotions by definition will depend on the cultural context. Not every culture wants an expressive baby, some, such as the Nso of the Cameroons, will tell an infant they are bad for crying and basically scold them, whereas a German mother will ask a crying infant what is wrong, then try and intuit what is wrong and remove the cause or redirect attention – perhaps even trying to turn the crying around into laughter through some sort of play or display³⁶.

At around 6 months or so babies begin to look around more at the environment rather than at the mother – mothers correspondingly start to follow the baby’s gaze and talk about things the baby is looking at rather than initiating and determining the objects they are engaging with³⁷. Note, I am going to use ‘mother’ mainly but also ‘caregiver’ at times. Sometimes ‘mother’ is exactly what is intended but at other times caregiver can be substituted – no offence is intended.

There are different singing styles for infants compared to pre-schoolers. Higher pitch singing for infants is the main one although pronunciation is clearer to the pre-schoolers as well, indicating mothers are targeting the skill levels and needs – infants are being schooled in prosody segmentation, pre-schoolers schooled in words. As far as importance to a performance context, Bergeson and Trehub state that it is “likely that the primary function in both cases is to maximise the mood of the interactants”. Obviously viewing a recorded performance removes the interaction between performer and audience. It is not clear what effect or constraint the lack of interaction has on the audience experience although there is some evidence performers modify their behaviour in response to audiences. (again this is obvious to anyone who performs or watches live performance). Performers, at least in a live piano recital setting, may moderate their performance so that it sits in the middle of their idea of audience expectations. Basically they reduce variability in amplitude and duration in the live setting but don’t do that in recordings.³⁹ Putting it another way, when performers perform for themselves they are more idiosyncratic in the performance. Audiences like the recorded live versions better than the personal versions of the performer. Tiny study though.

In these first months a social relation between audience and performer is built as the child learns that behaviour elicits performance (response) and that different behaviours elicit different performances. From the very earliest age the child is beginning the process of individuation, of being the performer and being the audience, of turn taking and acting in concert. These early (and earliest) performer/audience interactions are built around fulfilling fundamental emotional and physical needs so that the internalisation of the role of performer and audience is deeply linked to foundational experiences of emotion, individuation, nurture and sociability. This foundation underpins the meaning of music into adulthood, where people primarily use music in their everyday lives for “self-awareness, social relatedness and arousal and mood regulation”⁴².

We can see that adult vocal character and speech content changes in a developmentally appropriate way. Heightened pitch modulation introduces language as an important and learnable aspect of the environment. As language is learned two major functions for sound emerge for the developing infant— sound for language and sound for social bonding and emotion. People still use sound quality to convey emotion through language, the same sentence shouted or whispered will carry different meanings, but language

This is really about turn-taking – or perhaps turn taking is about performer audience interaction. Either way turn taking and performer audience interaction have within them similarities with respect to communicative goals. Turn taking – of the call and response kind – has been reported for premature babies in the 32 to 34 week range by Wataanbe – reported in Trevarthen(2008)²⁸ An astonishing finding that brings to the foreground what it means for something to be innate – surely preterm presence would seem ‘innate’ yet there is sufficient information in the environment for foetal learning and what may be innate is imitation.

sound, movement, rhythm and beat

Music is essentially physical and inextricably linked to movement. It is through music that we synchronise ourselves with others and build a sense of belonging. Through music and dance we challenge what Freeman has called 'the solipsistic divide' between our interior life and the life of the community⁴⁴.

Music and dance are linked in all cultures. Simply listening to music will activate those areas of the brain associated with movement and it is commonplace for listeners to unconsciously "move to the groove" with much the same gestural velocity as the performers did when making it.⁴⁵ Our memories are geared for recalling tempo and rhythm so accurately that the American Heart Association (AHA) officially recommends that "if you see a teen or adult suddenly collapse, call 9-1-1 and push hard and fast in the center of the chest to the beat of the classic disco song "Stayin' Alive."⁴⁶

Movement in response to music begins in utero⁴⁷ and continues right from birth as parents calm, comfort, soothe and excite their babies with jiggling, rocking, talking and crooning. From the very beginning parents provide a supportive social world where sound, vision and movement are used to regulate emotion. As early as 2 months babies actively begin to engage in social rhythmic interplay⁴⁸ using speech sounds in responsive turn taking with a parent^{28,49}.

Both the neuroscientists Walter Freeman and Richard Davidson, who has spent years investigating human happiness, have separately recommended social dancing as an activity that ameliorates depression

I guess we could point out societies that deliberately and forcefully prevent music and dance – but they, through their puritanical pathology, are naturally flagging the power and extent of their appeal by what they forbid.

Other countries use other songs – for example one Japanese group used Ob-La-Di Ob-La-Da to train nurses.

It appears that multimodal integration is possibly the initial state and that disaggregating sensory channels into individuated sources happens with development rather than the other way round ie with babies starting off with unintegrated channels that then get put together¹¹ This reminds me of synaesthesia, where a lack of pruning keeps sensory flow between regions with consequent overlap of phenomenal response eg see yellow and hear C#

The social and physical aspects of rhythm also develop along with the more basic perceptual aspects to build an integrated system where sound, vision, movement and emotion come together to influence interpretation and social bonding. Bouncing a baby up and down on the parental knee in time with some ambiguously stressed music will influence how that baby hears the stress pattern. Bounce them in threes and they hear 3/4, bounce them in twos and they hear duple meter^{50,51}. In social development if you bounce an infant gently in time to music they will be more helpful to people they see who are bouncing in time with them. If you are not bouncing in time, you won't get as much help⁵⁴.

Very early on then babies and infants are beginning to decompose music into constituent elements of pitch and rhythm which are then reconstituted and recombined with information from the visual and vestibular systems to form abstract relations between sound, vision, movement and emotion. With further development babies extend this process and begin to make their own patterns in sound and movement. Before long children can move in time with an external pulse and when the integration of perception and production is sufficiently mature children have the building blocks needed to actively synchronise their own activity with others in a musical setting.

There's some caveats here – babies can pick up the meter (stress pattern) pretty well if it is at all obvious, but if you present them with ambiguously stressed music whilst bouncing them on the knee in classic style, they will interpret the meter according to the stress pattern you are bouncing them to – e.g. bounce every third beat as against bounce every fourth etc. will give them a feel for 3/4 as against 4/4. Similar results occur with adults but there are some additional experiments that are interesting. Adults who are rocked on a see saw – i.e. they are passively rocked – will also disambiguate a stress pattern according to the movement but – if only the legs are passively moved they do not disambiguate – it appears that the vestibular system must be involved to impose movement onto the interpretation of rhythmic sound⁵². The vestibular system develops at about the same time as the auditory system so those linkages may be quite fundamental – in fact the auditory system originally evolved from the vestibular system in evolutionary history⁵³. This may have some impact on viewing media when seated – it is unclear what seated performance impacts apply to rhythmic entrainment for people who have a history of listening across a range of environments. It is reasonable to assume that they 'map' the rhythm onto the same affective and perceptual circuits / representations that most closely approximate their experiences of similar signals whilst fully mobile. But we don't actually know whether that is true or what the extent of the induced response is – it may be that there is a thinner, paler affective response or it might be there is just a different quality to the affective response, affect working under different constraints when immobile compared to mobile

Think of dances like the punk Pogo (up and down like a pogo stick), which seems the simplest method to achieve group bonding amongst an outgroup that only comes together at dances – no learning required, little variation amongst the dancers, maximal differentiation from other dance forms, rejection of skill based performance by people who felt they were rejected by having inappropriate or even no skills valued by the dominant socio-economic grouping.

Coordinated rhythmic movement with others is arguably the most powerful method for uniting people into an emotionally bound collective⁵⁵. Even something as simple as tapping together for a few minutes will give children the feeling they are similar to each other⁵⁶ and similar experiments in adults show that combining synchronous movement and shared goals leads to greater cooperation on future tasks⁵⁷. The goal need not be explicit and the cooperation can be extended outside the original task. Dancing for example enhances cooperative bonds in future non-dance activities and drumming together encourages being helpful on completely unrelated tasks. It appears this emotional binding is served through an overlap in the brain regions activated by synchronised activity between people and those regions responsible for feeling rewarded⁶⁰.

Movement does not just help us bond with others, it also communicates emotion. When people are asked to dance 'Happy' they dance faster and with broader, more impulsive gestures. When dancing 'Sad' they are slower with less movement⁶¹. And making those sad and happy movements has a corresponding emotional influence that in turn influences our perception of music. If children dance 'happy' or 'sad' to a piece of emotionally neutral music their interpretation of the music will be correspondingly influenced, possibly for days after⁶². Children will also respond to hearing music with movements that reflect parameters of the music they hear: fast music leads to fast movement and changing pitch encourages vertical movement⁶³.

It seems quite natural that if people have a shared goal and cooperate to achieve that goal they are more inclined to cooperate on achieving other goals⁵⁷. As so often, 'natural' is where the interest lies.

Eventually, when we are adults, just watching an intense film in the same room as others will increase our feelings of social cohesion with the rest of the audience. You don't even have to share your experience through talking – just being together is enough⁵⁸.

Interestingly the chance of 'catching' an earworm – (Involuntary Musical Imagery, or those annoying tunes you just can't stop singing in your head) is increased if you tap or move along with the tune when you hear it⁵⁹. In cultures where song and dance are deeply linked within group activities – which is pretty much every cultural group ever – recollection of song through the earworm will further reinforce group cohesion

Kokal et al⁶⁰ make the interesting point that the synchronised task – in their case drumming – has to be easy for pro-social bonding to take place. If the tasks are difficult then people concentrate too much on the task demand and basically ignore those around them.

Even if we are not moving ourselves we can recognise the emotion a dancer has tried to express just by seeing their movements, even without any sound⁶⁴. Similarly, we can identify quite accurately the emotional intension of instrumentalists just from vision of their performance (even if we can only see their torso). Not surprisingly, hearing the performance and both seeing and hearing give much more accurate results - as we saw for multi-modal perception in babies.

Both production and perception of movement are therefore linked to interpretation of the emotional valence of music from an early age and that linkage continues throughout life.

(and in experiments not even a dancer is used but rather a faceless androgynous avatar is evaluated so it really is just movement alone that gives us enough information to deduce the dancer's emotional expression. Interestingly the avatars whose motion was derived from female dancers seemed to express more obviously the relevant emotion – i.e. people could guess the emotion of the female based avatars more easily)

Visual information provides indications of the performer approach that is stronger in some cases than that provided by sound alone⁶⁵

Quite possibly this same system is used for visualisation training, where people imagine their performance on a task to then improve on the actual task later. Passive listening to something we have learned to play will improve our playing later – this must involve some sort of circuit that hears the piece and maps it to the motor circuits we use to play and strengthens those circuits in some way ie there is some sort of covert rehearsal happening⁶⁶.

linking movement, music and emotion in the brain

Seeing others dance and perform, and understanding and feeling music, activates common resources in our brain. One view is that these resources are linked through a system of 'mirror neurons' in the fronto-parietal cortex. It is thought that the mirror system maps our perceptions of purposeful action in others, through sight or sound, to our own motor systems and through that to the intentional and affective states we would be in if we were to be making similar actions^{68,69}.

We can hear someone sneaking up quietly as the sound and rhythm of their footsteps is intimately linked to their 'malign' intention. And we know what their intention is because if our footsteps were making that sound, in that rhythm, then that would be our intention too. In other words we simulate the other to find the meaning behind their actions and the richness of the meaning we derive from that simulation reflects the richness of our own experience^{70,71}.

And so it is with our experience with art.

Emotional responses are complex – music that evokes sadness for example is not necessarily to be avoided, and happy music can be annoying. For some Bayreuth is a pilgrimage, for others a nightmare. That music can be annoying to the point where someone can't stand hearing it, that it can be used as a vehicle for torture for example, is testament to the deep power music has. Visual aesthetics can have a similar power, perhaps noticed when interior decorating violates our taste to the extent we "just could not live in a room like that". This has nothing to do with practicalities and may have no obvious link to painful childhood memories of 'Shabby Chic' or "Hampton Style.

And so it is with our experience with art.

Quite possibly this same system is used for visualisation training, where people imagine their performance on a task to then improve on the actual task later. Passive listening to something we have learned to play will improve our playing later – this must involve some sort of circuit that hears the piece and maps it to the motor circuits we use to play and strengthens those circuits in some way ie there is some sort of covert rehearsal happening⁶⁶.

There is good evidence that both speech and song use the same audioparameters to indicate emotion. It follows from the mirror neuron hypothesis that hearing sung emotions and translating them to felt emotion is possible through a mapping of audible song parameters to normal spoken emotion⁶⁷.

It seems likely that the mirror neuron system provides another channel through which people with musical experience can understand performance – a channel that links their own history of sound production through movement with their emotional systems. People without practical music experience do not have this channel as their mirror neuron systems have no history with personal musical production⁷². EEG activation patterns show that musicians will respond more strongly to musical sounds compared to non-musicians, but have similar responses to non-musicians for non-musical sounds – like a child crying⁷³.

People may listen to self identified sad music to feel a connection at that moment with the music, to trigger memories, to hear something good or beautiful, or to hear a message they can relate to⁷⁴. Retrieving memories is often actually about feeling a social connection through memory. People may be looking to clarify their current mood, once clarified they can look to change it. Some indigenous Australians deliberately set phone ringtones to music that will make them remember their family when they are away. These ringtones are set knowing they will make them feel very sad, but this sadness 'is a feature, not a problem'.

Emotional responses are complex – music that evokes sadness for example is not necessarily to be avoided, and happy music can be annoying. For some Bayreuth is a pilgrimage, for others a nightmare. That music can be annoying to the point where someone can't stand hearing it, that it can be used as a vehicle for torture for example, is testament to the deep power music has. Visual aesthetics can have a similar power, perhaps noticed when interior decorating violates our taste to the extent we “just could not live in a room like that”. This has nothing to do with practicalities and may have no obvious link to painful childhood memories of ‘Shabby Chic’ or “Hampton Style.

The complexity in personal aesthetic preference makes understanding the neural circuitry underlying aesthetic response difficult. Experimental protocols and technologies in neuroimaging often lack support for what might be considered genuinely moving aesthetic experiences. Instead, neuro experiments tend to operationalise aesthetic experience as “unexpected chord”, or “10 second segments of popular tunes”. At best we get self-selected playlists of music listened to whilst in an fmri machine – an experience akin to listening to your favourite music with headphones whilst in a steel drum that someone is shaving with a lawnmower. Or perhaps visual art becomes forty paintings resized in Photoshop and saved as 500px square jpegs looked at while lying flat on your back in that same claustrophobic space⁷⁷.

It does appear though that there is a broad network of regions in the brain involved in our responses to music and visual art that are differentially activated according to the response we have--. For example, when listening to music that activates cortical regions known to be involved in introspection, emotion, and autobiographical memory, people report feelings of nostalgia^{79–81}.

People may listen to self identified sad music to feel a connection at that moment with the music, to trigger memories, to hear something good or beautiful, or to hear a message they can relate to⁷⁴. Retrieving memories is often actually about feeling a social connection through memory. People may be looking to clarify their current mood, once clarified they can look to change it. Some indigenous Australians deliberately set phone ringtones to music that will make them remember their family when they are away. These ringtones are set knowing they will make them feel very sad, but this sadness ‘is a feature, not a problem’.

Concert settings are social settings and we know that emotional responses to music and emotional responses to social situations both activate the anterior insular cortex⁷⁵ so there is at least an anatomical reason to think that there may be enhancement to the strength of emotional states when both social and musical emotions are involved. Nonetheless, some people consider classical concerts more an individual than a collective experience and can be irked by the presence of others, who they see as a distraction⁷⁶ whereas the communal aspect of cinema or of spectacles such as fireworks and sport can add to the experience.

This refers to one of two studies that implicate the caudate nucleus in aesthetic preference for visual images. The other study covers visual and musical beauty operationalised as forced choice ratings of 16 sec presentations of musical excerpts or projected visual images of paintings⁷⁸.

It even seems that music, and probably temporal art in general, invokes a broader reward mechanism than many other pleasures by activating dopamine release to both the anticipation and the conclusion of emotional musical moments that are known to the listener⁸². This finding is consistent with other reports that familiarity increases pleasure in listening and that we delight in hearing pieces over and over again^{83,84}. By activating two distinct reward pathways in the brain, music can be rewarding across a very broad range of circumstances, and given that music also activates all the core regions of the brain associated with emotion⁸⁵ the corresponding emotions are also very broad in scope.

Whatever the neural mechanism underlying the aesthetic experience we gain from the arts it is clear that we develop associations across all our experiences, so that each experience is a bundle of all the sensory, emotional and cognitive activity concurrent with it. Re-exposure to any part of that activity in the future will reactivate the corresponding associative network. A song heard as we walk in company along the foreshore, a smile, the turning of a head, a hand reaching out, will in some way reactivate all the other associated thoughts, sensations and feelings we have previously felt when even just one aspect of that experience is repeated. The strength of those recalled associations will be proportionate to the frequency of their previous conjunction and the emotional intensity of their past experience⁸⁶.

And sometimes, alone or with others, gazing at a screen or lost in the performance before us, that conjunction of the past with the present will be sublime.

1. Lopez-Teijon, M., Garcia-Faura, A. & Prats-Galino, A. Fetal facial expression in response to intravaginal music emission. *Ultrasound* (2015). doi:10.1177/1742271X15609367
2. Hepper, P. G. & Shahidullah, B. S. Development of fetal hearing. *Arch. Dis. Child.* **71**, 81–87 (1994).
3. Shahidullah, S. & Hepper, P. G. Frequency discrimination by the fetus. *Early Hum. Dev.* **36**, 13–26 (1994).
4. Gervain, J., Macagno, F., Cogoi, S., Peña, M. & Mehler, J. The neonate brain detects speech structure. *Proc. Natl. Acad. Sci.* **105**, 14222–14227 (2008).
5. Querleu, Denis;Renard, Xavier;Versyp, Fabienne;Paris-Delru, Laurence;Crepin, G. Fetal hearing. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **29**, 191–122 (1988).
6. Margulis, E. H. Repetition and Emotive Communication in Music Versus Speech. *Front. Psychol.* **4**, (2013).
7. Granier-Deferre, C., Ribeiro, A., Jacquet, A.-Y. & Bassereau, S. Near-term fetuses process temporal features of speech: Fetuses process temporal features. *Dev. Sci.* **14**, 336–352 (2011).
8. Vouloumanos, A. & Werker, J. F. Listening to language at birth: evidence for a bias for speech in neonates. *Dev. Sci.* **10**, 159–164 (2007).
9. DeCasper, A. J. & Spence, M. J. Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behav. Dev.* **9**, 133–150 (1986).
10. Freigang, C., Richter, N., RübSamen, R. & Ludwig, A. A. Age-related changes in sound localisation ability. *Cell Tissue Res.* **361**, 371–386 (2015).
11. Walker-Andrews, A. Infant's Perception of expressive behaviors: differentiation of multimodal information. *Psychol. Bull.* **121**, 437–456 (1997).
12. Burr, D. & Gori, M. Multisensory integration develops late in humans. (2012).
13. Saint-Georges, C. *et al.* Motherese in Interaction: At the Cross-Road of Emotion and Cognition? (A Systematic Review). *PLoS ONE* **8**, e78103 (2013).
14. Trehub, S. E. & Hannon, E. E. Infant music perception: Domain-general or domain-specific mechanisms? *Cognition* **100**, 73–99 (2006).
15. Hannon, E. E. & Trainor, L. J. Music acquisition: effects of enculturation and formal training on development. *Trends Cogn. Sci.* **11**, 466–472 (2007).
16. Lewkowicz, D. J. Learning and discrimination of audiovisual events in human infants: The hierarchical relation between intersensory temporal synchrony and rhythmic pattern cues. *Dev. Psychol.* **39**, 795–804 (2003).
17. Soto-Faraco, S., Calabresi, M., Navarra, J., Werker, J. F. & Lewkowicz, D. J. in *Multisensory Development* (2012).
18. Lewkowicz, D. J. in *Multisensory Development* (2012).
19. Hyde, D. C., Jones, B. L., Flom, R. & Porter, C. L. Neural signatures of face-voice synchrony in 5-month-old human infants. *Dev. Psychobiol.* **53**, 359–370 (2011).
20. Plantinga, J. & Trehub, S. E. Revisiting the innate preference for consonance. *J. Exp. Psychol. Hum. Percept. Perform.* **40**, 40–49 (2014).
21. Cohen, A. J. Development of tonality induction: Plasticity, exposure, and training. *Music Percept. Interdiscip. J.* **17**, 437–459 (2000).

22. Kitamura, C., Guellai, B. & Kim, J. Motherese by Eye and Ear: Infants Perceive Visual Prosody in Point-Line Displays of Talking Heads. *PLoS ONE* **9**, e111467 (2014).
23. Bowling, D. L., Sundararajan, J., Han, S. 'er & Purves, D. Expression of Emotion in Eastern and Western Music Mirrors Vocalization. *PLoS ONE* **7**, e31942 (2012).
24. Patel, A. D. & Daniele, J. R. An empirical comparison of rhythm in language and music. *Cognition* **87**, B35–B45 (2003).
25. Han, S. er, Sundararajan, J., Bowling, D. L., Lake, J. & Purves, D. Co-Variation of Tonality in the Music and Speech of Different Cultures. *PLoS ONE* **6**, e20160 (2011).
26. Phillips-Silver, J. & Keller, P. E. Searching for Roots of Entrainment and Joint Action in Early Musical Interactions. *Front. Hum. Neurosci.* **6**, (2012).
27. Cross, I. Music and evolution: Consequences and causes. *Contemp. Music Rev.* **22**, 79–89 (2003).
28. Trevarthen, C. The musical art of infant conversation: Narrating in the time of sympathetic experience, without rational interpretation, before words. *Music. Sci.* **12**, 15–46 (2008).
29. Fernald, A. Four month old infants prefer to listen to Motherese. *Infant Behav. Dev.* **8**, 181–195 (1985).
30. Fernald, A. & Kuhl, P. Acoustic determinants of infant preference for motherese speech. *Infant Behav. Dev.* **10**, 279–293 (1987).
31. Corbeil, M., Trehub, S. E. & Peretz, I. Speech vs. singing: infants choose happier sounds. *Front. Psychol.* **4**, (2013).
32. Ilari, B. & Sundara, M. Music Listening Preferences in Early Life: Infants' Responses to Accompanied Versus Unaccompanied Singing. *J. Res. Music Educ.* **56**, 357–369 (2009).
33. Lamont, A. Toddler's musical preferences: musical preference and musical memory in the early years. *Ann. N. Y. Acad. Sci.* **999**, 518–519 (2003).
34. Plantinga, J. & Trainor, L. J. Long-term memory for Pitch in six month old infants. *Ann. N. Y. Acad. Sci.* **999**, 520–521 (2003).
35. Gogate, L., Maganti, M. & Bahrack, L. E. Cross-cultural evidence for multimodal motherese: Asian Indian mothers' adaptive use of synchronous words and gestures. *J. Exp. Child Psychol.* **129**, 110–126 (2015).
36. Keller, H. & Otto, H. The cultural socialization of emotion regulation during infancy. *J. Cross-Cult. Psychol.* **40**, 996 (2009).
37. Nakata, T. & Trehub, S. E. Infants' responsiveness to maternal speech and singing. *Infant Behav. Dev.* **27**, 455–464 (2004).
38. Bergeson, T. & Trehub, S. E. Mother's singing to infants and pre-schoolchildren. *Infant Behav. Dev.* **22**, 51–64 (1999).
39. Shoda, H. & Adachi, M. Why live recording sounds better: a case study of Schumann's TrÅmerei. *Front. Psychol.* **5**, (2015).
40. Milligan, K., Atkinson, L., Trehub, S. E., Benoit, D. & Poulton, L. Maternal attachment and the communication of emotion through song. *Infant Behav. Dev.* **26**, 1–13 (2003).
41. de l'Etoile, S. K. Infant behavioral responses to infant-directed singing and other maternal interactions. *Infant Behav. Dev.* **29**, 456–470 (2006).
42. SchÅfer, T., Sedlmeier, P., StÅtler, C. & Huron, D. The psychological functions of music listening. *Front. Psychol.* **4**, (2013).

43. Koelsch, S. *et al.* Music, language and meaning: brain signatures of semantic processing. *Nat. Neurosci.* **7**, 302–307 (2004).
44. Freeman, W. *Societies of brains. A study in the neuroscience of love and hate.* (Lawrence Erlbaum, 1995).
45. Leman, M., Desmet, F., Styns, F., Van Noorden, L. & Moelants, D. Sharing Musical Expression Through Embodied Listening: A Case Study Based on Chinese Guqin Music. *Music Percept. Interdiscip. J.* **26**, 263–278 (2009).
46. Two steps to staying alive with hands on CPR.
47. Kisilevsky, B. S., Hains, S. M. J., Jacquet, A.-Y., Granier-Deferre, C. & Lecanuet, J.-P. Maturation of fetal responses to music. *Dev. Sci.* **7**, 550–559 (2004).
48. Ruzza, B., Rocca, F., Lenti Boero, D. & Lenti, C. Investigating the musical qualities of early infant sounds. *Ann. N. Y. Acad. Sci.* **999**, 527–529 (2003).
49. Stormark, K. M. & Braarud, H. C. Infants' sensitivity to social contingency: A 'double video' study of face-to-face communication between 2–4 month-olds and their mothers. *Infant Behav. Dev.* **27**, 195–203 (2004).
50. Phillips-Silver, J. & Trainor, L. J. Hearing what the body feels: Auditory encoding of rhythmic movement. *Cognition* **105**, 533–546 (2007).
51. Phillips-Silver, J. & Trainor, L. J. Feeling the Beat: Movement Influences Infant Rhythm Perception. *Science* **308**, 1430 (2005).
52. Phillips-Silver, J. & Trainor, L. J. Vestibular influence on auditory metrical interpretation. *Brain Cogn.* **67**, 94–102 (2008).
53. Koelsch, S. Brain correlates of music-evoked emotions. *Nat. Rev. Neurosci.* **15**, 170–180 (2014).
54. Cirelli, L. K., Wan, S. J. & Trainor, L. J. Fourteen-month-old infants use interpersonal synchrony as a cue to direct helpfulness. *Philos. Trans. R. Soc. B Biol. Sci.* **369**, 20130400–20130400 (2014).
55. Phillips-Silver, J., Aktipis, C. A. & A. Bryant, G. The Ecology of Entrainment: Foundations of Coordinated Rhythmic Movement. *Music Percept.* **28**, 3–14 (2010).
56. Rabinowitch, T.-C. & Knafo-Noam, A. Synchronous Rhythmic Interaction Enhances Children's Perceived Similarity and Closeness towards Each Other. *PLOS ONE* **10**, e0120878 (2015).
57. Reddish, P. & Bulbulia, J. Let's Dance Together: Synchrony, Shared Intentionality and Cooperation. *PLoS ONE* **8**, (2013).
58. Rennung, M. & Göritz, A. S. Facing Sorrow as a Group Unites. Facing Sorrow in a Group Divides. *PLoS One* **10**, e0136750 (2015).
59. McCullough Campbell, S. & Margulis, E. H. Catching an Earworm Through Movement. *J. New Music Res.* **44**, 347–358 (2015).
60. Kokal, I., Engel, A., Kirschner, S. & Keysers, C. Synchronized Drumming Enhances Activity in the Caudate and Facilitates Prosocial Commitment - If the Rhythm Comes Easily. *PLoS ONE* **6**, e27272 (2011).
61. Van Dyck, E., Maes, P.-J., Hargreaves, J., Lesaffre, M. & Leman, M. Expressing Induced Emotions Through Free Dance Movement. *J. Nonverbal Behav.* **37**, 175–190 (2013).
62. Maes, P.-J. & Leman, M. The Influence of Body Movements on Children's Perception of Music with an Ambiguous Expressive Character. *PLoS ONE* **8**, e54682 (2013).
63. Kohn, D. & Eitan, Z. Musical parameters and children's movement responses. (2009).77. Vartanian, O. & Goel, V. Neuroanatomical correlates of aesthetic preference for paintings. *Neuroreport* **15**, 893–897 (2004).

63. Kohn, D. & Eitan, Z. Musical parameters and children's movement responses. (2009).
64. Van Dyck, E., Vansteenkiste, P., Lenoir, M., Lesaffre, M. & Leman, M. Recognizing Induced Emotions of Happiness and Sadness from Dance Movement. *PLoS ONE* **9**, e89773 (2014).
65. Davidson, J. W. Visual perception of performance manner in the movements of solo musicians. *Psychol. Music* **21**, 103–113 (1993).
66. Lahav, A. The Power of Listening: Auditory-Motor Interactions in Musical Training. *Ann. N. Y. Acad. Sci.* **1060**, 189–194 (2005).
67. Scherer, K. R., Sundberg, J., Tamarit, L. & Salomão, G. L. Comparing the acoustic expression of emotion in the speaking and the singing voice. *Comput. Speech Lang.* **29**, 218–235 (2015).
68. Molnar-Szakacs, I. & Overy, K. Music and mirror neurons: from motion to 'e'motion. *Soc. Cogn. Affect. Neurosci.* **1**, 235–241 (2006).
69. Ocampo, B., Kritikos, A. & Cunnington, R. How Frontoparietal Brain Regions Mediate Imitative and Complementary Actions: An fMRI Study. *PLoS ONE* **6**, e26945 (2011).
70. Schiavio, A., Menin, D. & Matyja, J. Music in the flesh: Embodied simulation in musical understanding. *Psychomusicology Music Mind Brain* **24**, 340–343 (2014).
71. Overy, K. & Molnar-Szakacs, I. Being Together in Time: Musical Experience and the Mirror Neuron System. *Music Percept.* **26**, 489–504 (2009).
72. Chapin, H., Jantzen, K., Scott Kelso, J. A., Steinberg, F. & Large, E. Dynamic Emotional and Neural Responses to Music Depend on Performance Expression and Listener Experience. *PLoS ONE* **5**, e13812 (2010).
73. Krause, A. E. & North, A. C. Music listening in everyday life: Devices, selection methods, and digital technology. *Psychol. Music* 0305735614559065 (2014).
74. Van den Tol, A. J. & Edwards, J. Listening to sad music in adverse situations: How music selection strategies relate to self-regulatory goals, listening effects, and mood enhancement. *Psychol. Music* **43**, 473–494 (2015).
75. Petrini, K., Crabbe, F., Sheridan, C. & Pollick, F. E. The Music of Your Emotions: Neural Substrates Involved in Detection of Emotional Correspondence between Auditory and Visual Music Actions. *PLoS ONE* **6**, e19165 (2011).
76. Dobson, M. C. New Audiences for Classical Music: The Experiences of Non-attenders at Live Orchestral Concerts. *J. New Music Res.* **39**, 111–124 (2010).
77. Vartanian, O. & Goel, V. Neuroanatomical correlates of aesthetic preference for paintings. *Neuroreport* **15**, 893–897 (2004).
78. Ishizu, T. & Zeki, S. Toward A Brain-Based Theory of Beauty. *PLoS ONE* **6**, e21852 (2011).
79. Trost, W., Ethofer, T., Zentner, M. & Vuilleumier, P. Mapping Aesthetic Musical Emotions in the Brain. *Cereb. Cortex* **22**, 2769–2783 (2012).
80. Janata, P. The Neural Architecture of Music-Evoked Autobiographical Memories. *Cereb. Cortex* **19**, 2579–2594 (2009).
81. Janata, P. Brain Networks That Track Musical Structure. *Ann. N. Y. Acad. Sci.* **1060**, 111–124 (2005).
82. Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A. & Zatorre, R. J. Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nat. Neurosci.* **14**, 257–262 (2011).
83. Margulis, E. H. *On repeat. How music plays the mind.* (Oxford University Press, 2016).

84. Pereira, C. S. *et al.* Music and Emotions in the Brain: Familiarity Matters. *PLoS ONE* **6**, e27241 (2011).
85. Koelsch, S. Towards a neural basis of music-evoked emotions. *Trends Cogn. Sci.* **14**, 131–137 (2010).
86. Maes, P.-J., Leman, M., Palmer, C. & Wanderley, M. M. Action-based effects on music perception. *Front. Psychol.* **4**, (2014).

