Mimesis and Mechanical Coupling in the Bodily Response to Sound: An Oscillatory Account

Flyxion
Independent Researcher

November 2026

Abstract

Arnie Cox's mimetic hypothesis holds that listeners understand music partly by covertly simulating the actions that produce it. Extended to impacts, bass, and engineered film sound, the hypothesis meets bodily responses that begin before any sound source has been identified. This essay distinguishes five partly independent sources of bodily response to sound: mechanical coupling to tissue, the acoustic startle reflex, entrainment to periodic structure, covert mimetic simulation, and learned anticipation. The organism is treated as regulating these sources through antagonist activation and opposed rhythmic units, and the visual route is added because observed action engages overlapping motor circuitry. Associations between sound and bodily state can become shared by common exposure or by copying between individuals, and only copying is called memetic. A compact formal model with five proved propositions shows, among other results, that bracing changes the body's transfer function without cancelling the stimulus, that a feedback loop between tension and arousal runs away only above unit gain, and that the startle drive cannot be recovered from a gesture projection that discards level and onset rate. Seven experiments are specified together with outcomes that would count against the account. No measurements are reported, and the central hypothesis is not established.

Keywords: embodied music cognition; mimetic hypothesis; acoustic startle; sensorimotor entrainment; mechanical impedance; central pattern generators; sound symbolism; film sound

1 Introduction

Sound reaches an organism first as a mechanical event and only afterwards as information. Most cognitive accounts of listening begin after that transformation, at the point where a signal has already been segmented, categorised, and assigned a source. Embodied accounts of music cognition move the starting point earlier by insisting that the listener's own body participates in perception, and the most developed version of this position is Arnie Cox's mimetic hypothesis, according to which listeners understand music in part by imagining or covertly performing the actions that would produce it (Cox, 2001, 2016). The hypothesis has been productive in music theory and in the psychology of music, and it has found support in neuroimaging and stimulation studies showing that heard sounds engage motor systems (Gazzola et al., 2006; Fadiga et al., 2002; Zatorre et al., 2007).

An obvious temptation is to extend the hypothesis beyond music. Environmental sounds, sudden impacts, vocal bursts, and the engineered soundtracks of film and television all appear to produce bodily responses, and a mimetic account would unify these with the response to music. The extension is less straightforward than it looks. A violinist's bowing is a recognisable action that a listener might covertly reproduce, whereas a gunshot is not an action the listener performs, and the bodily response to it begins within tens of milliseconds, before any identification of its source can have been completed (Brown et al., 1991; Koch, 1999). The response is a contraction of facial and neck musculature, a transient change in autonomic activity, and a shift in the felt character of the scene, and none of this seems to require imitation of anything. A further observation sharpens the difficulty. A viewer who understands an action sequence perfectly well by sight may nonetheless be driven into a state of muscular and autonomic activation by its soundtrack, and may reduce that state sharply by attenuating the sound while keeping the picture. The ordinary experience of muting a loud scene separates comprehension from forcing: what the sound does to the body is not simply a by-product of what it tells the viewer.

The thesis developed here is that the bodily response to sound has several partly independent sources, that the organism regulates their combined effect rather than merely receiving it, and that mimetic simulation is one component of the system and not the whole of it. Section 2 fixes terminology and states the central claims together with their evidential status. Five sources are distinguished in Section 4: direct mechanical coupling, reflexive defensive response, entrainment to periodic structure, covert mimetic simulation, and learned anticipation. Section 5 adds the visual route, since observed movement recruits motor circuitry in ways that overlap with the auditory route, and the two can converge on one bodily state. Section 6 takes up a further step. A correspondence between a sound parameter and a bodily parameter that is acquired by individual learning can come to be shared by two routes that are kept apart: common exposure to the same stimuli, and copying between individuals. The second is the step from mimetic to memetic in the literal sense of Dawkins's coinage, in which the word *meme* was built from *mimeme*, a unit of imitation (Dawkins, 1976). Section 7 proposes that parameters of this kind, and in particular intensity realised as the contrast between a clenched and an unclenched state, are stored not as symbols but as the tuning and coupling of opposed rhythmic units of the kind long studied in motor control. Section 8 gives a compact formal model that makes

the proposal precise enough to be wrong. Section 9 states which parameters the proposed measurements can and cannot identify, Section 10 specifies experiments and the outcomes that would count against the account, Section 11 compares the account with its closest alternatives, and Section 12 considers the objections most likely to be pressed.

The contribution is conceptual and formal. No measurements are reported, the numerical values in the text are order-of-magnitude illustrations rather than estimates fitted to data, and the experiments of Section 10 are proposals. Where the literature is contested, as in the case of mirror neurons, facial feedback, and memetics, the account is built so that it does not depend on the contested claim, and the dependence that remains is stated.

2 Definitions and Central Claims

A few terms need fixed meanings if the account is to remain testable. *Forcing* refers to the physical and neural influence that a sound exerts on the body independently of what the listener takes the sound to mean, and *response* to the resulting bodily state, including muscular activation, autonomic change, and the associated experience. *Comprehension* refers to the listener's identification of the sound's source, the event it indicates, and the action that produced it. The central observation is that comprehension and forcing can dissociate: the same event understood through vision alone is accompanied by a different bodily state from the same event accompanied by its sound. A *source* of bodily response is a causal pathway from sound to bodily state that can be separated from the others by a manipulation of the stimulus or of the listener, as in Table 2. *Regulation* refers to any change in the transmission or expression of the forcing that is produced by the organism, whether by altering impedance through antagonist activation, by joining or resisting a rhythm, or by redirecting attention, and it is distinguished from *attenuation*, which is a change in the stimulus itself. A *pole pair* is a pair of mutually inhibiting units whose contrast represents a parameter, as in the contrast between a tense and a released state, and a *mapping* is the set of weights that connect features of a stimulus to the units of a pole pair. Two routes by which a mapping can come to be shared are distinguished in Section 6, common-exposure convergence and copying between individuals, and the word *memetic* is reserved for the second.

The account makes eight claims of unequal strength, listed in Table 1 with their evidential status and the place where each is tested. The grading matters because the claims do not stand or fall together. The first four are structural or rest on findings established in the component literatures, whereas the last four depend on evidence that is limited or absent, and a failure of the later claims would leave the earlier ones intact.

3 The Mimetic Hypothesis and Its Boundaries

Cox's central claim is that musical meaning is grounded in the listener's imagined or covertly enacted participation in the sound-producing gestures of the music. The mechanism includes mimetic motor imagery, in which the listener imagines performing the action, and subvocalisation,

Claim	Statement	Evidential status	Tested by
C1	The bodily response to sound has at least five partly separable sources.	Each pathway is documented in its own literature; joint separability in one paradigm is not shown.	Experiments 1 and 3
C2	Regulation through antagonist activation changes the body's transfer function without cancelling the stimulus, and the sign of the effect above the relaxed resonance depends on the relative gain of stiffness and damping.	Mathematical consequence of the plant model; parameter values are unmeasured.	Proposition 1; Experiment 2
C3	Visual and auditory routes can converge on one bodily state, and sound can acquire through association the capacity that picture or reflex first had.	Supported for observation-execution matching and for associative learning; transfer from sound to tension is not shown.	Experiments 1 and 4
C4	The startle drive cannot be recovered from a gesture projection that discards level and onset rate.	Formal, conditional on an invariance hypothesis about the projection.	Proposition 4
C5	Intensity is the most transferable dimension of the response to sound.	Modest cross-cultural support from a small number of studies.	Experiment 5
C6	Muscular tension is the privileged physical realisation of intensity.	Conjecture.	Experiment 5
C7	Cultural transmission preferentially preserves the mapping between sound and that realisation.	Conjecture.	Experiment 5
C8	Parameters of this kind are stored in oscillator dynamics and coupling weights and not as symbols.	Analogy with motor control; no direct evidence.	Experiments 4 and 5, indirectly

Table 1: Central claims of the account, with evidential status and the test that bears on each.

in which the listener covertly sings or articulates the heard line, and it may recruit motor systems without any overt movement (Cox, 2016). The hypothesis connects to the broader programmes of motor-mimetic music cognition developed by Godøy and of embodied music cognition developed by Leman (Godøy, 2003; Godøy and Leman, 2010; Leman, 2007), and it draws on the image-schematic theories of meaning of Lakoff and Johnson, in which abstract concepts are structured by patterns of bodily experience (Lakoff and Johnson, 1980).

The empirical case for the covert engagement of motor systems by sound is substantial though not decisive. Gazzola and colleagues found that listening to sounds produced by hand actions and by mouth actions activated premotor regions in a somatotopically organised way, and that the same regions were active when the listener performed those actions (Gazzola et al., 2006). Kohler and colleagues identified neurons in the monkey premotor cortex that respond to the sound of an action as well as to its sight and its execution (Kohler et al., 2002). Fadiga and

colleagues showed that listening to speech modulates the excitability of the listener's tongue muscles in a manner specific to the articulatory demands of the heard words (Fadiga et al., 2002). Zatorre, Chen, and Penhune reviewed the auditory-motor interactions that accompany both the perception and the production of music (Zatorre et al., 2007), and Janata and Grafton described shared neural substrates for sequencing and for music (Janata and Grafton, 2003). Jeannerod's account of covert action simulation supplies a general framework in which such engagement is a normal part of motor cognition and need not be accompanied by conscious imagery (Jeannerod, 2001). The last point matters, because it means the hypothesis does not depend on listeners reporting mental images of movement.

The interpretation of these findings is contested. Hickok catalogued problems with the mirror neuron account of action understanding, including the dissociation between motor deficits and comprehension deficits (Hickok, 2009). Heyes proposed that the matching properties of mirror neurons are the product of associative learning from correlated experience of observing and executing actions and are not an evolved specialisation for understanding others (Heyes, 2010), and a later review concluded that mirror-neuron areas contribute to low-level processing of observed actions while the stronger claims about action understanding remained unsupported (Heyes and Catmur, 2022). The account developed here is deliberately neutral on this dispute. What it requires is the weaker and less controversial claim that auditory and visual inputs can engage motor and autonomic systems through connections that are either innate or learned, and the associative account of those connections is in fact the more congenial to the learning rule of Section 8.

Two boundaries of the mimetic hypothesis deserve explicit statement. The first is that it concerns understanding, which presupposes some identification of the action or gesture that the sound is taken to express. A response that precedes identification, such as the startle reflex, is therefore outside its natural scope. The second is that it is one proposal among several about how music and sound produce bodily and emotional effects. Juslin and Västfjäll distinguished brain stem reflexes, rhythmic entrainment, evaluative conditioning, emotional contagion, visual imagery, episodic memory, and musical expectancy as separable mechanisms by which music induces emotion, and argued that research had suffered from the failure to consider them separately (Juslin and Västfjäll, 2008). Their taxonomy concerns emotion and does not treat direct mechanical coupling of sound to tissue as a distinct mechanism, but its methodological lesson applies here with full force: a theory that attributes every bodily response to one mechanism cannot be tested against its competitors, and a theory that names several must say how they can be told apart.

4 Five Sources of Bodily Response to Sound

4.1 Mechanical coupling

The auditory system is a mechanical system before it is a neural one. The ossicles transmit the motion of the tympanic membrane into the fluid of the cochlea, where a travelling wave moves along the basilar membrane, and the place of maximal displacement depends on frequency: high

frequencies peak near the base, where the membrane is narrow and stiff, and low frequencies peak near the apex, where it is wide and compliant (von Békésy, 1960; Robles and Ruggero, 2001). The sensory hair cells then transduce this motion into neural signals. The frequency selectivity of the cochlea is therefore a consequence of a gradient of mechanical properties, and it is not accurately described as small structures resonating at high frequencies and large ones at low frequencies. This is the best-understood case of a general point: the first stage of the response to sound is the response of a mechanical structure.

Mechanical coupling is not confined to the cochlea. Acoustic pressure acts on the entire body surface, and structure-borne vibration enters through the floor, the seat, and the skeleton. The skin contains mechanoreceptors that respond to vibration over a wide range, and psychophysical work has shown that distinct channels mediate sensitivity to different vibration frequencies, with maximal sensitivity near a few hundred hertz and separate channels serving lower frequencies (Bolanowski et al., 1988). The vestibular system also responds to loud low-frequency sound. Todd and Cody recorded acoustically evoked electromyographic responses in the sternocleidomastoid muscle to techno music at intensities above about 90 dB(A) in a sample of ten subjects, and argued that the response is vestibularly mediated because earlier work had obtained it in subjects with lost cochlear function but not in subjects with lost vestibular function (Todd and Cody, 2000). The whole body also has mechanical resonances at low frequencies, with trunk resonances in the range of a few hertz to about ten hertz under whole-body vibration (Griffin, 1990). The experience of bass at a loud performance is therefore partly auditory and partly somatosensory in a literal sense, because tissue and skeleton are being moved.

The magnitude of the direct acoustic force on the body is easily overstated, and a short calculation clarifies it. Sound pressure level L in decibels corresponds to a root-mean-square pressure $p = p_0\,10^{L/20}$ with $p_0 = 20\,\mu\mathrm{Pa}$. At $L = 110$ dB the pressure is about 6.3 Pa, and at $L = 120$ dB it is 20 Pa. Acting on a frontal torso area of the order of $0.3\,\mathrm{m}^2$, these pressures correspond to forces of roughly 2 N and 6 N rms. The weight of a 70 kg adult is about 690 N, so the direct pressure force is below one percent of body weight even at very high levels, although it is applied cyclically and at frequencies where tissue is sensitive. The figures assume a plane wave in free field and ignore standing-wave and diffuse-field effects, so they indicate only the order of magnitude. The conclusion is that the direct mechanical route matters most for sustained low-frequency content and structure-borne vibration, which can drive tissue near its resonances, and matters little for brief transients, whose effect on the body is carried predominantly by neural pathways.

4.2 The startle reflex and its autonomic accompaniments

A sudden loud sound with a rapid onset elicits the acoustic startle response, a stereotyped pattern that Landis and Hunt described in detail in the 1930s as beginning with an eyeblink and spreading to flexion of the neck, shoulders, and trunk (Landis and Hunt, 1939). The circuit is short. The obligatory pathway runs from the cochlear root neurons through the caudal pontine reticular nucleus to motoneurons, and its brevity accounts for latencies of tens of milliseconds (Davis, 1984; Koch, 1999). Brown and colleagues documented the pattern of muscle

recruitment and the habituation of the response in human subjects (Brown et al., 1991), and the standard methodological guidelines for laboratory work specify stimuli of high intensity and near-instantaneous rise time because the abruptness of onset is the critical feature (Blumenthal et al., 2005). The reflex is accompanied or followed by changes in skin conductance and heart rate, and the acoustic features that elicit it, notably intensity and abruptness, are also those that produce unpleasant and arousing ratings in studies of affective response to sound (Bradley and Lang, 2000).

Two properties of the reflex bear directly on the present account. The first is modulation. The reflex is not a fixed gain. A weak stimulus presented shortly before the startling stimulus reduces the response, a phenomenon known as prepulse inhibition (Graham, 1975; Hoffman and Ison, 1980), and the response is potentiated during aversive states and attenuated during pleasant ones, so that the affective state of the organism changes the gain of a brainstem reflex (Lang et al., 1990). The second is habituation. Repeated presentation reduces the response, and a classical account describes this as a process separate from sensory adaptation (Groves and Thompson, 1970). The reflex is therefore a gain-controlled, adapting element embedded in the regulation of the organism, which is the form in which it enters the model of Section 8. It is also by definition independent of source identification, which is why it is outside the natural scope of the mimetic hypothesis. That the reflex precedes interpretation does not mean that interpretation and motor simulation cannot follow it within the next few hundred milliseconds. The sources overlap in time without being identical, which is one reason Section 10 relies on contrasts between conditions and not on the response to a single event.

4.3 Entrainment to periodic structure

When sound has periodic structure, bodily activity can become coupled to it. Large and Jones proposed that attention is itself rhythmic, with internal oscillations that synchronise with the temporal structure of events (Large and Jones, 1999). Nozaradan and colleagues showed with electroencephalography that neural activity is selectively tagged at the frequencies of the beat and the meter of a rhythmic stimulus, including frequencies that are weak or absent in the acoustic input (Nozaradan et al., 2011). The behavioural literature on sensorimotor synchronisation shows that people tap and move in time with a beat with an accuracy and flexibility that depends on the structure of the rhythm (Repp and Su, 2013), and Phillips-Silver and Trainor showed that the way infants are moved while listening to an ambiguous rhythm determines which metrical structure they later prefer, which is direct evidence that movement shapes rhythm perception (Phillips-Silver and Trainor, 2005). The mathematical theory of synchronisation supplies the general framework: a self-sustained oscillator driven weakly by a periodic force locks to it when the frequency mismatch is smaller than a quantity proportional to the coupling strength (Pikovsky et al., 2001). Entrainment differs from reflex in its time course, because it builds over several cycles and does not occur on a single transient, and it differs from mimetic simulation in that it requires periodicity but does not require identification of any sound-producing action. A listener entrained to a drone or to a steady pulse has no particular action to simulate.

4.4 Covert mimetic simulation

The fourth source is the one that the mimetic hypothesis addresses: covert engagement of the motor system in a manner that reflects the action thought to produce the sound, or the gesture-like contour of the sound itself. It requires some recognition of the sound's source or shape, builds over a time course of hundreds of milliseconds to seconds, can be partial or inhibited, and depends on the listener's experience with the relevant actions. A violinist's bowing, a singer's phrasing, and the tapping of a drum are cases in which the correspondence between sound and action is clear. There is also a looser case in which the listener embodies not the production of the sound but its temporal and energetic character, so that a slow flowing passage is accompanied by reduced tension and slow breathing, and a sudden abrupt event by a transient contraction. In this looser case the bodily response is an analogue of the sound's dynamics and does not reproduce its production. Staccato material is then reconstructed as short motor impulses, legato material as smooth extension, and vibrato as oscillation, a mapping formalised in Section 8. The distinction between imitating how a sound is made and embodying how it unfolds in time and energy is central to what follows, because it is the second and not the first that extends naturally to sounds that no listener produces.

4.5 Learned anticipation

A fifth source is the learned anticipation of an event before it occurs. A viewer who has seen a scene before, or who has learned the conventions of a genre, may tense the neck and trunk and show raised skin conductance during the preceding swell or silence, so that the bodily state precedes the acoustic event for which it prepares. Huron's account of expectation in music treats prediction, and its fulfilment or violation, as a basis of affect (Huron, 2006), and the startle literature shows that the gain of the reflex depends on the state and context of the organism (Graham, 1975; Lang et al., 1990). Anticipation differs from the reflex in that it precedes onset and can develop with familiarity with the cue. It differs from entrainment in that it needs no periodicity, only a reliable contingency between a cue and an event. It differs from mimetic simulation in that what is simulated is the arrival of the event and not the action that produces the sound. It is also the point at which regulation and forcing interact most clearly, because raising co-contraction before a known impulsive event changes the plant before the force arrives, in the manner analysed in Section 8. Whether anticipatory tension adds to or reduces the response to the event itself is an empirical matter, and the reflex literature suggests that the answer depends on context and may go either way.

4.6 Comparison and dissociation

Table 2 summarises the distinctions. The five sources are not mutually exclusive, and a single event can engage all of them: a loud drum hit has an abrupt onset, a periodic context, a recognisable action, and low-frequency content. The point of separating them is that they make different predictions. Repeating an identical gunshot recording should produce habituation of the startle component and no corresponding habituation of the mimetic component, which

Source	Principal trigger	Approximate timescale	Requires source identification	Voluntary modulation
Mechanical coupling	Sustained low-frequency pressure; structure-borne vibration	Immediate; follows the waveform	No	Changes transmission through posture and tone, but not the stimulus
Startle reflex	Abrupt onset at high level	Tens of milliseconds; habituates over repetitions	No	Gain modulated by affect, anticipation, and prepulse; not abolished
Entrainment	Periodic structure	Builds over several cycles	No	Strong: can be joined, exaggerated, or resisted
Mimetic simulation	Recognisable action, gesture, or contour	Hundreds of milliseconds to seconds	Yes, at least partially	Strong; overt expression can be inhibited
Learned anticipation	Reliable cue that precedes an expected event	Precedes the event; develops with exposure	No, but requires a learned cue	Strong; shaped by attention and context

Table 2: Five sources of bodily response to sound. Timescales are approximate and are given only to indicate how the sources can be separated experimentally.

depends on the content and not the onset. Playing the same low-frequency tone through loudspeakers and through a seat transducer should differ in the airborne component but not in the structure-borne component. A steady drone should engage entrainment and mechanical coupling but leave the startle reflex quiescent. Repeated viewing of the same scene should reduce the startle component at the bang while increasing activity in the interval before it, if anticipation is a separate source. An account that cannot say which of these patterns it expects has not yet said anything, and the experiments of Section 10 are built on exactly these contrasts.

5 The Visual Route and the Cross-Modal Case

Observation of movement engages the motor system. Fadiga and colleagues used transcranial magnetic stimulation to show that watching another person's hand and arm movements facilitates motor-evoked potentials in the observer's corresponding muscles (Fadiga et al., 1995), and imaging work found somatotopically organised premotor and parietal activation during observation of actions made with different effectors (Buccino et al., 2001). The behavioural counterpart is automatic imitation, in which observing a movement speeds the execution of the same movement and slows the execution of a different one (Brass et al., 2001; Heyes, 2011).

Everyday mimicry is a further manifestation: people unconsciously adopt the postures and mannerisms of those they interact with (Chartrand and Bargh, 1999), and exhibit rapid facial reactions to emotional faces even when the faces are presented too briefly to be consciously seen (Dimberg et al., 2000). Neonatal imitation of facial gestures has also been reported (Meltzoff and Moore, 1977), though its interpretation remains contested.

These findings supply the familiar case that the mimetic hypothesis generalises: when a person sees a waving motion, the motor system engages in a manner related to waving, and the overt expression of this engagement may or may not follow. Hurley's shared circuits model gives the engagement a structure. It proposes layered control mechanisms in which mirroring, covert simulation, and inhibition of overt imitation are built from a common architecture, so that the observer's motor system can be activated by observation without producing overt movement, and the same architecture supports deliberate imitation when inhibition is released (Hurley, 1998, 2008; Hurley and Chater, 2005). The relevant feature for the present account is the inhibitory layer. It explains why watching a wave does not usually produce a wave, and it supplies the covert-to-overt boundary that the regulation model of Section 8 treats as a continuous control variable. Tversky's *Mind in Motion* places this engagement at the foundation of thought more broadly, arguing that bodily action and the spatial understanding built from it shape abstract thinking, and that gesture and diagram externalise that thinking so that it can be shared (Tversky, 2019). The waving case is then not a curiosity of imitation but an instance of a general relation between action, representation, and communication, and it is the relation that the step from mimetic to memetic in the next section extends from the individual to the group.

Two cautions apply. The first concerns the direction of causation between bodily state and experience. The facial feedback hypothesis, which holds that adopting an expressive posture changes emotional experience, received an early demonstration (Strack et al., 1988), but a coordinated registered replication effort did not reproduce the original effect (Wagenmakers et al., 2016). Claims that muscular activity causes emotional or autonomic change must therefore be stated with restraint, and the present account asks only that bodily states modulate subsequent processing and be measurable, which is a weaker claim than that they generate experience. The second caution concerns the interpretation of mimicry as imitation of an intended action. Associative accounts hold that the connections between seeing and doing arise from correlated experience (Heyes, 2010, 2011), which means that the visual route is learned in the same way as the sound route, and that learned links between sound and motor state can be built by the same mechanism.

The auditory and visual routes converge in media. It has long been known that sound and picture integrate at the perceptual level, as in the demonstration that seeing a speaker's lips changes what is heard (McGurk and MacDonald, 1976), and film theory describes the converse effect of sound on what is seen as added value (Chion, 1994). The sound design of film is not neutral with respect to the mechanisms of Section 4. A survey of more than a hundred films found that soundtracks contain what the authors called nonlinear analogues, noisy and abruptly shifting sounds of the kind produced by animals under distress, with a distribution across genres that matches their emotional function (Blumstein et al., 2010), and an experimental follow-up

found that adding such features to neutral music changes ratings of arousal and valence, with the effect depending on context (Blumstein et al., 2012). These features are precisely those that engage the reflex and entrainment pathways, and they are used where the picture alone would be less intense.

This yields a decomposition that is useful for the muted-viewing observation. Watching an action sequence without sound preserves the visual route, including the motor engagement elicited by observed action, and removes the airborne and structure-borne mechanical route, the acoustic startle reflex, and the entrainment and mimetic components that depend on the sound. A silent image of a gunshot does not trigger the acoustic startle reflex, because the trigger is the acoustic transient and not its meaning. Comprehension of the scene is preserved because it is mediated by vision and by the viewer's knowledge of what is occurring. What differs between the two viewing conditions is therefore the part of the bodily state that sound contributes independently of comprehension. This part is not the consequence of understanding, and it is not necessarily what the viewer wants to be experiencing, which is why an attentive viewer may regard it as something imposed by the production.

Finally, learned association connects the routes. If sound and picture co-occur repeatedly, as they do in film and in ordinary experience, then on an associative account sound features should acquire the capacity to evoke motor and autonomic states that originally required the picture or the reflex. This is the higher-level learned response: after sufficient pairing, a sound alone can elicit a pattern that was originally elicited by a visual event, a reflex, or an overt action. Section 8 states the learning rule and derives what it predicts.

6 From Mimetic to Memetic

The term *mimetic* concerns imitation within an individual, as in the engagement of the motor system by observation. The term *memetic* concerns the transmission of patterns among individuals, and the two are connected etymologically and conceptually. Dawkins constructed *meme* from the Greek root of *mimesis* to name a unit of cultural transmission that is copied by imitation (Dawkins, 1976), and Blackmore developed the proposal that imitation is the characteristic mechanism of such transmission in human culture (Blackmore, 1999). Hurley and Chater assembled the case that imitation is an important and under-studied process linking neuroscience and social science (Hurley and Chater, 2005), and Tversky's account of gesture and diagram as shared externalisations of action-based thought supplies the corresponding picture of how bodily patterns become public (Tversky, 2019). Donald proposed that human cognitive evolution passed through a mimetic stage, in which gesture, pantomime, rhythm, and non-linguistic vocalisation provided a medium for the intentional representation and transmission of skills before language (Donald, 1991). The memetic framework is contested as a general theory of culture, and the contest is not entered here. The term is used descriptively and in a restricted sense. Two routes by which a correspondence between a sound parameter and a bodily parameter can come to be shared must be kept apart. In common-exposure convergence, individuals who never influence one another acquire similar mappings because they learn from the same stimulus environment. In transmission by copying, one individual's behaviour is the

stimulus from which another learns, so that the mapping passes from person to person. Only the second is memetic in the strict sense used here, and the first can occur without it.

There is good reason to think that some such correspondences are widely shared even before cultural transmission is considered. The crossmodal correspondence literature documents robust associations between auditory dimensions and dimensions in other modalities, including pitch and size, pitch and elevation, and loudness and brightness (Spence, 2011). The classic demonstration that people systematically match rounded and jagged shapes with particular nonsense words (Köhler, 1947) has been interpreted as reflecting correspondences between articulatory and visual features (Ramachandran and Hubbard, 2001). In animal communication, Morton argued that harsh, low-frequency calls are used in hostile contexts and tonal, higher-frequency calls in appeasing contexts, a pattern he called motivation-structural rules (Morton, 1977), and Ohala extended the idea to human use of fundamental frequency as a signal of size and dominance (Ohala, 1984). In music, listeners from a culture with no exposure to Western music were able to recognise happiness, sadness, and fear in Western excerpts at above-chance levels (Fritz et al., 2009), and a study comparing Canadian listeners with Mbenzélé Pygmy listeners found that subjective and physiological arousal responses to music tracked arousing features similarly in both groups while valence responses differed (Egermann et al., 2015). The last result is the most relevant to the present account. It suggests that the axis of intensity, which can be realised in the body as the degree of tension, is the most transferable dimension, while valence depends more heavily on learned convention. Both studies compared a small number of groups, and the claim of universality should be held loosely.

Three hypotheses should be separated here. The first is that intensity, as distinct from valence, is the most transferable dimension of the response to sound, which the cross-cultural studies above support modestly. The second is that muscular tension is the privileged physical realisation of that dimension, so that a clenched or unclenched state is the natural bodily medium for it, as opposed to breathing, vocal effort, gesture size, or autonomic arousal. The third is that cultural transmission preferentially preserves the mapping between sound and that realisation. The cross-cultural studies bear on the first hypothesis only, and the second and third are conjectures that Experiment 5 is designed to address separately. If all three held, the contrast between a clenched and an unclenched state would be a good carrier for other parameters. A bodily contrast of this kind is available to every listener, can be produced and recognised in others by sight and sound, and has a direct mechanical interpretation, so that urgency, threat, effort, resistance, proximity, and excitement could be mapped onto it by correspondence, and sound could serve as the carrier because it can evoke and modulate that state directly. On this view the sound would not be a signal interpreted by a separate symbolic system but a physical influence on the same variable that the organism uses to represent the parameter.

The notion of a private interlanguage requires care. Selinker introduced *interlanguage* for the intermediate system of a language learner, with its own regularities, lying between the native and target languages (Selinker, 1972). The analogy used here is that the mapping between acoustic parameters and bodily parameters is an intermediate code that is partly idiosyncratic and partly convergent, because individuals acquire it from exposure and correct

it by feedback. The word *private* should not be read in the sense of Wittgenstein's private language argument (Wittgenstein, 1953). The mapping is not a language whose meanings are fixed by inner ostension, and it does not need to be, because it is a regulatory coupling that becomes public through behaviour: a listener's tension is visible to others, audible in breathing, and measurable. It is private only in the weak sense that it does not require public symbols to operate, and it becomes memetic when the couplings of different individuals converge.

Section 8 states when common-exposure convergence occurs. Under the linear-response assumptions stated there, an associative rule applied to a shared stimulus ensemble gives learners equilibrium mappings that lie within a bounded distance of one another, the bound depending on differences in bodily gain and unconditioned drive and growing as learners approach the stability boundary of the learning loop. Similarity of equilibria is therefore a conditional result and not a general consequence of shared exposure, since differences in mechanics, prior experience, attention, and feedback can leave persistent divergence outside those conditions. Film and television sound design show both routes. Audiences converge by common exposure to a repertoire of sound features paired with intensity states, while sound designers and composers learn and reproduce one another's conventions by copying, which is memetic in the strict sense. Whether audience convergence is also supported by copying, as when viewers imitate the reactions of others in the room, is a further question that Experiment 5 separates from common exposure.

7 Dichotomous Poles and Oscillatory Storage

The proposal that intensity is stored as a contrast between tense and released states raises the question of what kind of storage this is. The motor-control literature offers a precedent in the organisation of opposed pairs. Sherrington described reciprocal innervation, in which the activation of an agonist is accompanied by the inhibition of its antagonist (Sherrington, 1906). Brown showed that rhythmic alternation can arise from the mutual inhibition of two neural populations together with fatigue, so that a pair of opposed units, a half-center, produces rhythm without rhythmic input (Brown, 1911). The subsequent literature on central pattern generators established that such circuits underlie locomotion, respiration, and other rhythmic acts across vertebrates and invertebrates (Grillner, 1985; Marder and Bucher, 2001). Matsuoka gave a compact mathematical model of two mutually inhibiting neurons with adaptation that sustains oscillation and can be entrained by periodic input (Matsuoka, 1985), and this model is used in Section 8.

The opposition between poles also has a long history in the theory of affect. The opponent-process theory of motivation proposed that a stimulus-driven affective state is followed by an opposed state, with dynamics that change with repeated exposure (Solomon and Corbit, 1974). It is important that the autonomic nervous system is not itself a single bipolar scale. Berntson, Cacioppo, and Quigley showed that sympathetic and parasympathetic activity can vary reciprocally, independently, or coactively, and that the autonomic space is therefore at least two-dimensional (Berntson et al., 1991). The poles proposed here are accordingly not a single global tense-versus-relaxed dimension but a family of pole pairs, organ-specific and

effector-specific, that can be coordinated. This observation also corrects a simple feedback story in which tension causes sympathetic activation which in turn increases tension: the relations are heterogeneous, and they include negative feedback, adaptation, and context-dependent gain.

At the mechanical end of this hierarchy, the pole pair has a direct physical meaning. Hogan showed that coactivation of antagonist muscles sets the mechanical impedance of a joint, so that the same pair of muscles that produces torque also determines stiffness (Hogan, 1984), and Burdet and colleagues showed that the nervous system learns to adjust this impedance to stabilise unstable dynamics (Burdet et al., 2001). Co-contraction is therefore a variable by which the organism changes how it transmits external forces, and not only a means of producing movement. The impedance variable is the bridge between the oscillator pair and the mechanical plant of Section 8.

Different features of an oscillator pair can encode different parameters. The period can carry tempo, the phase relative to a stimulus can carry timing, the amplitude and the duty cycle, meaning the fraction of each cycle spent in the tense phase, can carry degree of intensity. The coordination of two oscillators, whether the listener's tension waxes with the stimulus intensity or opposes it, can carry the stance of the listener toward the stimulus. This last feature has a formal counterpart in the Haken-Kelso-Bunz model of coordination, in which the relative phase φ of two coupled oscillators obeys $\dot{\varphi} = -a \sin \varphi - 2b \sin 2\varphi$, with both in-phase and anti-phase coordination stable when $b/a > 1/4$ and anti-phase coordination losing stability below that ratio (Haken et al., 1985; Kelso, 1995). The relevance is that bistability gives two discrete coordination modes within a continuous system, and that a change of coupling parameter can switch between them, which is a plausible analogue of the voluntary shift between joining a rhythm and resisting it. Hierarchies of oscillators at different time scales, from cardiac and respiratory cycles to the cycles of locomotion and of musical phrases, are a general feature of neural and bodily organisation (Buzsáki, 2006), and they provide the substrate in which such pole pairs would operate.

The claim that such parameters are stored in oscillator dynamics is thus a claim about the form of storage. A parameter is stored in the input weights that connect sound features to the units of an opposed pair, in the intrinsic parameters that determine period and duty cycle, and in the couplings among pairs. It is not stored as a symbol that is looked up. This has the consequence that retrieval and expression coincide: the stored mapping is exercised whenever the sound drives the pair, and the expression of the stored parameter is the bodily state that results. It also has the consequence that the storage can be modulated by regulation and context, since the same weights can yield different outputs depending on the state of the oscillator and the impedance of the body.

8 A Formal Sketch

The model combines a mechanical plant whose impedance is set by antagonist commands, two oscillator pairs that generate those commands, a startle pathway with habituation, and an associative rule that shapes the drive weights. Everything in this section is illustrative. The model is intended to show that the claims of the preceding sections can be stated precisely and

Symbol	Meaning
x, M, C, K	Displacement, effective mass, damping, and stiffness of a lumped body segment
F_a, F_m	Acoustic and structure-borne forcing; net force of the antagonist pair
w, w_{co}	Total antagonist activation; co-contraction, the simultaneous part of the activation
K_0, C_0, k, c	Relaxed stiffness and damping; stiffness and damping gained per unit activation
a_i, b_i, y_i	Activation, adaptation, and output of oscillator unit i
q, z, u	Tension pole variable; movement pole variable; common activation set by the tension pair
e_f, e_e	Flexor and extensor commands
ϕ_m, g_{mi}	Stimulus features; weights from feature m to unit i
E, θ, G	Sound level envelope; onset-rate threshold; startle gain
η, λ, ρ	Learning rate; weight decay; small-signal gain of a unit
Σ_ϕ, Σ_u	Covariance of the features; covariance of the features with the unconditioned response
μ	Loop gain of the bodily feedback path
Γ, $\mathcal{G}$, $\mathcal{P}_{\mathrm{mim}}$	Prosodic trajectories; motor gesture space; mimetic projection

Table 3: Principal symbols used in Section 8.

can be wrong, and the parameters are not fitted to any dataset.

8.1 Notation

8.2 Plant

Let $x(t)$ denote a lumped displacement of a body segment, for example the chest wall, with effective mass M. Let $w(t) \geq 0$ denote the total antagonist activation, defined in Section 8.3 as the summed activation of an antagonist muscle pair, which sets stiffness and damping. The plant is

$$M\ddot{x} + C(w)\,\dot{x} + K(w)\,x = F_a(t) + F_m(t), \tag{1}$$

with impedance modulated by antagonist activation according to

$$K(w) = K_0 + k\,w, \qquad C(w) = C_0 + c\,w, \qquad k > 0,\ c \geq 0,\ 0 \leq w \leq w_{\max}. \tag{2}$$

The forcing $F_a(t) = A_{\mathrm{eff}}\,p(t) + F_{\mathrm{sb}}(t)$ comprises the pressure force on an effective area A_{eff} and any structure-borne term, and $F_m(t)$ is the net force produced by the antagonist pair. The natural frequency and damping ratio at co-contraction w are

$$\omega_0(w) = \sqrt{K(w)/M}, \qquad \zeta(w) = \frac{C(w)}{2\sqrt{M\,K(w)}}. \tag{3}$$

For sinusoidal forcing at angular frequency ω the steady-state displacement gain is

$$|H(\omega; w)| = \left[\big(K(w) - M\omega^2\big)^2 + \big(C(w)\,\omega\big)^2 \right]^{-1/2}. \tag{4}$$

At resonance, $\omega = \omega_0(w)$, the gain is $1/(C\,\omega_0) = 1/(2\zeta K)$.

Proposition 1 (Impedance regulation changes the transfer function and does not cancel the stimulus)**.** *Under Eqs.* (1)–(4)*: (i) $\omega_0(w)$ is strictly increasing in w; (ii) for every $\omega \le \omega_0(0)$ the gain $|H(\omega; w)|$ is strictly decreasing in w for $w > 0$; (iii) the logarithmic sensitivity $\partial \ln |H|/\partial w$ is $O(\omega^{-2})$ as $\omega \to \infty$; (iv) the damping ratio increases with w if and only if $c/C(w) > k/(2K(w))$; and (v) for any ω let $w^*(\omega) = \big[k(M\omega^2 - K_0) - cC_0\omega^2\big]/(k^2 + c^2\omega^2)$; then $|H(\omega; w)|$ increases with w on $[0, w^*)$ and decreases on $(w^*, w_{\max}]$ when $w^* > 0$, and decreases on the whole range when $w^* \le 0$. In particular $w^* > 0$ requires $kM > cC_0$ and $\omega^2 > kK_0/(kM - cC_0)$, so that if $kM \le cC_0$ bracing lowers the gain at every frequency.*

Proof. (i) follows from $\omega_0^2 = K(w)/M$ and $k > 0$. For (ii), write $D(w) = (K - M\omega^2)^2 + C^2\omega^2$, so that $|H| = D^{-1/2}$ and

$$\frac{\partial D}{\partial w} = 2(K - M\omega^2)\,k + 2\,C\,c\,\omega^2.$$

If $\omega \le \omega_0(0)$ then $M\omega^2 \le K_0 \le K(w)$, so the first term is non-negative for all $w \ge 0$, and the second is non-negative because $c \ge 0$. For $w > 0$ we have $K(w) > K_0 \ge M\omega^2$, so the first term is strictly positive and $\partial D/\partial w > 0$, hence $|H|$ strictly decreases. For (iii), $\partial D/\partial w = 2\omega^2(Cc - kM) + 2Kk$ and $D = M^2\omega^4 + (C^2 - 2KM)\omega^2 + K^2$, so $\partial \ln |H|/\partial w = -\frac{1}{2}(\partial D/\partial w)/D = O(\omega^{-2})$. For (iv), $\partial \ln \zeta/\partial w = c/C(w) - k/(2K(w))$, and the sign gives the stated condition. For (v), $\partial D/\partial w = 2k\,f(w)$ with $f(w) = K(w) - M\omega^2 + (c/k)\,C(w)\,\omega^2$, which is affine in w with slope $k + c^2\omega^2/k > 0$ and root w^*. Where $f < 0$ we have $\partial D/\partial w < 0$, so D falls and the gain rises, and where $f > 0$ the gain falls. The root is positive if and only if $k(M\omega^2 - K_0) > cC_0\omega^2$, that is, $\omega^2(kM - cC_0) > kK_0$, which requires $kM > cC_0$. □

The proposition has a direct interpretation, and the qualification in (v) matters. Bracing, that is, raising antagonist activation, reduces the displacement produced by forcing below the relaxed resonance, because the response is stiffness-dominated there. Far above resonance the response is inertia-dominated and nearly independent of stiffness, so bracing has little effect. In the band between the relaxed resonance and the raised resonance, however, bracing moves the resonance toward the forcing frequency and can increase the response. Voluntary bracing is therefore not a general attenuation mechanism. It changes the body's frequency-dependent transfer function, sometimes beneficially and sometimes adversely, depending on where the dominant energy of the stimulus lies relative to the body's resonance. The increase in the band above the relaxed resonance is conditional on damping. It occurs only if $kM > cC_0$, so that the stiffness gained by bracing outweighs the damping gained, and if damping grows sufficiently with activation, bracing lowers the response at every frequency. This makes the prediction more discriminating, because the sign of the effect in the band depends on measurable properties of the body. In no case does regulation alter the incoming acoustic pressure, which is why actual attenuation of the sound is a different operation from regulation of the response to it.

8.3 Oscillator pairs and antagonist commands

Four units $i \in \{1, 2, 3, 4\}$ follow the Matsuoka model, with activation a_i and adaptation b_i, arranged in two mutually inhibiting pairs: a tension pair $(1, 2)$ representing the tense and released poles, and a movement pair $(3, 4)$ representing opposed directions of rhythmic movement. For each unit let $j(i)$ denote its partner within the pair. Then

$$\tau_r \, \dot{a}_i = -a_i - \beta \, b_i - \gamma \, \left[a_{j(i)}\right]_+ + c_0 + s_i(t), \tag{5}$$

$$\tau_a \, \dot{b}_i = -b_i + [a_i]_+ \, , \tag{6}$$

with outputs $y_i = [a_i]_+$. The pole variables are $q = y_1 - y_2$ for the tension pair and $z = y_3 - y_4$ for the movement pair. The tension pair sets a common activation $u = \tanh([q]_+)$, so that the released pole reduces u directly and not only through inhibition of the tense unit. Commands to an antagonist flexor-extensor pair are

$$e_f = [u + z]_+ \, , \qquad e_e = [u - z]_+ \, , \tag{7}$$

with net force $F_m = \kappa \, (e_f - e_e)$ and total antagonist activation $w = e_f + e_e$. This follows the separation in Hogan's analysis, in which the difference of antagonist activations produces torque and their sum sets stiffness (Hogan, 1984). Total activation and co-contraction proper must be distinguished. Define the co-contraction as $w_{\mathrm{co}} = \min(e_f, e_e) = [u - |z|]_+$, the part of the activation that is simultaneous in both muscles. Then

$$w = |e_f - e_e| + 2 \, w_{\mathrm{co}} = |F_m|/\kappa + 2 \, w_{\mathrm{co}}, \tag{8}$$

so that stiffness and damping, which depend on total activation, receive one contribution from the net command, because a muscle that is producing force is also stiffer, and a second from co-contraction, which is the part attributable to bracing. When $|z| \leq u$ both muscles are active, $F_m = 2\kappa z$, $w_{\mathrm{co}} = u - |z|$, and $w = 2u$, so that the tension pair alone fixes the total activation while the movement pair sets how it divides between movement and bracing. Reciprocal activation without bracing ($u = 0$) gives force κz with $w = |z|$ and $w_{\mathrm{co}} = 0$, and bracing without movement ($z = 0$) gives zero net force with $w = 2u = 2w_{\mathrm{co}}$. A person who braces against a stimulus without moving in time with it is thus represented by $u > 0$, $z \approx 0$, and a person who moves with the stimulus without bracing by z following the stimulus with $u \approx 0$. Outside the region $|z| \leq u$ one muscle is silent and the two are coupled.

The drive to unit i is a weighted sum of feature channels,

$$s_i(t) = \sum_m g_{mi} \, \phi_m(t) + s_i^{\mathrm{st}}(t), \tag{9}$$

where $\phi_m(t)$ are features extracted from sound (for example a loudness envelope, an onset-rate feature, rhythmic features, and cue features such as a swell or a silence that precede events), from picture (for example a motion-energy feature), and from interoceptive channels, and s_i^{st} is the startle term, which is nonzero only for the tension pair. For features that signal intensity

one expects $g_{m1} > 0$ and $g_{m2} \leq 0$, so that the feature raises the tense unit and lowers the released unit. Rhythmic features drive the movement pair. Features of mimetic origin enter through the mimetic projection of the author's earlier work on embodied prosody, $\mathcal{P}_{\mathrm{mim}} : \Gamma \to \mathcal{G}$, which maps a prosodic trajectory Γ to a point of a motor gesture space $\mathcal{G}$ parameterised by direction, amplitude, periodicity, and continuity, carrying prosodic curvature and coherence onto reconstructible motion primitives, so that staccato material evokes short repeated motor impulses, legato material continuous smooth extension, and jitter and vibrato oscillatory motion (Flyxion, 2026). Its proposition on mimetic equivalence states that perception of a trajectory is mediated by covert reconstruction of the corresponding motor trajectory, up to equivalence in curvature and coherence. Writing $\phi^{\mathrm{mim}}(t)$ for the components of $\mathcal{P}_{\mathrm{mim}}(\Gamma)(t)$, the weights on these channels are the learned coupling from a reconstructed gesture to the antagonist commands. On this reading the present model is one possible dynamical realisation of the projection when it is carried out by a body that has impedance and can oppose the stimulus. Whether the two routes commute, that is, whether projecting a sound to a gesture and then driving the pair gives the same result as driving the pair directly from the sound's features, is taken up in Section 8.7.

For a wide range of parameters the movement pair has a stable limit cycle with an intrinsic frequency ω_c set by τ_r, τ_a, β, and γ. Under the standard assumptions of phase reduction, namely a stable limit cycle, forcing weak enough that the amplitude is unperturbed, and an interaction function dominated by its first harmonic, a periodic input at frequency Ω with coupling amplitude ε gives the Adler equation for the phase difference ψ,

$$\dot{\psi} = \Delta - \varepsilon \sin \psi, \qquad \Delta = \Omega - \omega_c, \tag{10}$$

which has a stable fixed point, and hence locking, if and only if $|\Delta| \leq \varepsilon$ (Pikovsky et al., 2001). The Matsuoka pair does not by itself guarantee a sinusoidal interaction function. In general the reduction takes the form $\dot{\psi} = \Delta - \varepsilon \Lambda(\psi)$, where Λ is the interaction function determined by the phase response of the pair, and locking occurs when Δ/ε lies in the range of Λ, with stability at points where $\Lambda' > 0$ (Pikovsky et al., 2001). The Adler form is therefore a conditional simplification. What survives in general is that the locking range scales linearly with ε at weak coupling, which is the standard Arnold tongue structure and states in quantitative form why a stronger rhythmic drive entrains over a wider range of tempi.

8.4 Startle pathway and habituation

Let $E(t)$ be the sound level envelope. The startle term is triggered by onset rate and not by level alone:

$$s_i^{\mathrm{st}}(t) = \sigma_i \, G(t) \, (h * \left[\dot{E} - \theta\right]_+)(t - \tau_{\mathrm{st}}), \qquad i \in \{1, 2\}, \tag{11}$$

where θ is an onset-rate threshold, h is a short pulse kernel, τ_{st} is a latency of tens of milliseconds, and σ_i are signs ($\sigma_1 > 0$ for the tense unit, $\sigma_2 < 0$ for the released unit). The gain obeys a habituation equation,

$$\tau_h \, \dot{G} = (G_\infty - G) - \alpha \, G \, \left[\dot{E} - \theta\right]_+, \tag{12}$$

so that each supra-threshold onset reduces G and the gain recovers with time constant τ_h. Prepulse inhibition can be represented by multiplying G by $1-\iota$ for a short window after a weak preceding stimulus, and affective potentiation by letting G_∞ depend on a slowly varying state of the pair. These are the minimal elements that the cited literature on startle requires (Graham, 1975; Hoffman and Ison, 1980; Lang et al., 1990; Groves and Thompson, 1970).

8.5 Associative learning in the closed loop

For each unit i of the tension pair let $g_i = (g_{mi})_m$ be its vector of drive weights, adjusted by a covariance rule with decay,

$$\dot{g}_{mi} = \eta\,(\phi_m - \bar{\phi}_m)(y_i - \bar{y}_i) - \lambda\, g_{mi}, \tag{13}$$

where overbars denote running means and $\eta, \lambda > 0$. The output y_i depends on the weights, so the learning loop is closed. To make this tractable assume the following. The features $\phi(t)$ are exogenous, meaning that they are not influenced by the learner, and stationary with covariance matrix Σ_ϕ whose largest eigenvalue is $s_{\max}$. The unit responds approximately linearly and quasi-statically to its learned drive, $y_i = y_i^u + \rho\, g_i \cdot \phi$, where y_i^u is the unconditioned response (reflex, visual, or mechanical drive) and $\rho \geq 0$ is a small-signal gain that depends on the learner's bodily and neural properties. Let $\Sigma_u = \mathrm{Cov}(\phi, y_i^u)$.

Proposition 2 (Closed-loop learning and the similarity of equilibria)**.** *Under these assumptions the expected dynamics of Eq.* (13) *are linear,*

$$\dot{g}_i = \eta\,\Sigma_u + (\eta\rho\,\Sigma_\phi - \lambda I)\, g_i.$$

(i) The dynamics are stable if and only if $\eta\rho\, s_{\max} < \lambda$, and then the unique fixed point is

$$g^* = \eta\,(\lambda I - \eta\rho\,\Sigma_\phi)^{-1}\Sigma_u.$$

(ii) Let two learners a and b be exposed to the same feature ensemble, with the same η and λ, gains $\rho_a, \rho_b \geq 0$ and unconditioned covariances Σ_u^a, Σ_u^b, and let both be stable with margins $\mu_a = \lambda - \eta\rho_a s_{\max} > 0$ and $\mu_b = \lambda - \eta\rho_b s_{\max} > 0$. Then

$$\|g_a^* - g_b^*\| \;\leq\; \frac{\eta\,\|\Sigma_u^a - \Sigma_u^b\|}{\mu_a} \;+\; \frac{\eta^2\,|\rho_a - \rho_b|\, s_{\max}\,\|\Sigma_u^b\|}{\mu_a\,\mu_b}.$$

Proof. With $y_i = y_i^u + \rho\, g_i \cdot \phi$ we have $\mathrm{Cov}(\phi, y_i) = \Sigma_u + \rho\,\Sigma_\phi\, g_i$, and taking expectations in Eq. (13) gives the stated linear system. Its matrix $\eta\rho\Sigma_\phi - \lambda I$ is symmetric with eigenvalues $\eta\rho s_k - \lambda$ for the eigenvalues $s_k \in [0, s_{\max}]$ of Σ_ϕ, so it is negative definite if and only if $\eta\rho s_{\max} < \lambda$, which gives (i) and the fixed point. For (ii) write $A = \lambda I - \eta\rho_a\Sigma_\phi$ and $B = \lambda I - \eta\rho_b\Sigma_\phi$, which are symmetric with smallest eigenvalues at least μ_a and μ_b respectively, so $\|A^{-1}\| \leq 1/\mu_a$ and $\|B^{-1}\| \leq 1/\mu_b$. Then

$$g_a^* - g_b^* = \eta A^{-1}(\Sigma_u^a - \Sigma_u^b) + \eta(A^{-1} - B^{-1})\Sigma_u^b,$$

and the resolvent identity $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$ with $B - A = \eta(\rho_a - \rho_b)\Sigma_\phi$ and $\|\Sigma_\phi\| = s_{\max}$ gives the bound. □

The proposition replaces a statement that held only when the learned weights did not influence the outputs. It remains a conditional result, and it shows what the conditions are. The fixed point is the earlier covariance mapping $(\eta/\lambda)\Sigma_u$ amplified by the loop gain, which reduces to it when $\rho = 0$. Shared stimulation yields nearby equilibria only to the extent that learners are stable, similar in bodily gain, and similar in unconditioned drive, and the bound grows without limit as either learner approaches the stability boundary, where small differences are amplified. The proposition concerns equilibria. It does not show that the distance between arbitrary learners decreases over time, which would also depend on their initial weights. Learners with different Σ_ϕ, that is, with different exposure, have fixed points that differ in general. The assumptions also exclude two cases that matter for the argument of Section 6: interactive settings in which the features include other learners' behaviour, which is where copying between individuals enters, and entrained oscillating units, for which the quasi-static linear response is not valid. If the feature enters the rule through a low-pass filtered eligibility trace, the rule responds to the lagged covariance between a feature and the later state of the pair, so that cue features that precede an event acquire drive onto the tense unit and tension begins before the event. This is the model's account of learned anticipation.

8.6 Feedback from bodily state and the conditions for runaway

The model so far lets the pair drive the body without returning bodily state to the pair. Proprioceptive and interoceptive signals do return it, and this raises the question posed by the simple feedback story, in which sound produces tension, tension produces arousal, and arousal increases tension: whether such a loop must spiral into a saturated state. Let r be the activity of the tense pole, let I be the net external drive, and let bodily feedback add a term μr to the drive, where $\mu \geq 0$ is the loop gain of the feedback path. Ignoring adaptation for the moment, a reduced model is

$$\tau\,\dot{r} = -r + \tanh\big(I + \mu\, r\big) \;=:\; g_I(r). \tag{14}$$

Proposition 3 (Feedback gain and runaway)**.** *(i) If $\mu \leq 1$, then for every I Eq.* (14) *has a unique equilibrium, and it is globally asymptotically stable. For small $|I|$ the equilibrium is approximately $I/(1-\mu)$. (ii) If $\mu > 1$, then at $I = 0$ there are exactly three equilibria, the unstable state* 0 *and two stable states $\pm r^*$ with $r^* > 0$, and for all sufficiently small $|I|$ there are exactly three equilibria, two stable and one unstable.*

Proof. We have $g_I'(r) = -1 + \mu\,\mathrm{sech}^2(I + \mu r) \leq \mu - 1$. (i) If $\mu \leq 1$ then $g_I' \leq 0$ with equality at most at one point, so g_I is strictly decreasing, and since $g_I(r) \to \pm\infty$ as $r \to \mp\infty$ it has exactly one root. The scalar flow $\dot{r} = g_I(r)/\tau$ then converges to that root from every initial condition. For small I and r the approximation $\tanh x \approx x$ gives $r \approx I + \mu r$, hence $r \approx I/(1-\mu)$. (ii) For $I = 0$ the function g_0 is odd, $g_0'(0) = \mu - 1 > 0$, and g_0 is strictly concave on $r > 0$ because tanh is, with $g_0(r) \to -\infty$. Hence there is exactly one positive root r^*, with $g_0 > 0$ on $(0, r^*)$ and $g_0 < 0$ beyond, so $g_0'(r^*) < 0$. By oddness $-r^*$ is also a root with negative slope, and 0 is a

root with positive slope. The three roots are simple, so they persist for small $|I|$ by the implicit function theorem, and for large $|r|$ the sign of g_I is fixed by the term $-r$, so no further roots appear. □

The loop therefore spirals into a saturated state only if its gain exceeds one, and even then the result is a bistable pair of states and not a continuing escalation. For $\mu < 1$ the response is graded, returns to baseline when the drive is removed, and is amplified by the factor $1/(1-\mu)$ in the small-signal range. The units of Section 8.3 also carry slow adaptation, which erodes a high state and converts bistability into relaxation oscillation, as in the Matsuoka model (Matsuoka, 1985), so that even a supercritical loop does not hold the organism in the high state indefinitely. Regulation enters naturally as a reduction of μ, for example by descending inhibition or by redirection of attention. The proposition also yields a qualitative prediction that distinguishes the two regimes. If $\mu > 1$ the response to a ramp of drive should show hysteresis, so that the response at a given level differs between the upward and downward legs of the ramp, and if $\mu < 1$ it should not. The reduction ignores adaptation and the multi-unit structure of the pair, and its conclusions are conditional on that simplification.

8.7 Understanding and forcing do not commute

Two routes from a sound to a bodily response can be compared. In the direct route the features of the sound drive the pair through Eq. (9). In the mimetic route the sound is first projected to a gesture $\mathcal{P}_{\mathrm{mim}}(\Gamma) \in \mathcal{G}$, and only that gesture drives the pair. The drive is said to factor through the projection if there is a map T_i with $s_i(\Gamma) = T_i(\mathcal{P}_{\mathrm{mim}}(\Gamma))$ for all trajectories Γ.

Proposition 4 (Non-factorisation of the startle drive)**.** *Suppose $\mathcal{P}_{\mathrm{mim}}$ carries only curvature and coherence and is therefore invariant under scaling of the sound level, so that $\mathcal{P}_{\mathrm{mim}}(\Gamma_\nu) = \mathcal{P}_{\mathrm{mim}}(\Gamma)$ when Γ_ν has envelope νE with $\nu > 0$ and the same curvature and coherence. Let Γ have a continuously differentiable envelope with $\sup_t \dot{E}(t) \le \theta$ and $\sup_t \dot{E}(t) > 0$, and let $\nu > \theta / \sup_t \dot{E}(t)$. Then for $G > 0$ and $i \in \{1, 2\}$ the startle term of Eq. (11) satisfies $s_i^{\mathrm{st}}(\Gamma) \equiv 0$ and $s_i^{\mathrm{st}}(\Gamma_\nu) \not\equiv 0$, so the startle term does not factor through $\mathcal{P}_{\mathrm{mim}}$. If the remaining channels of Eq. (9) depend on the sound only through $\mathcal{P}_{\mathrm{mim}}$, the full drive does not factor either.*

Proof. Since $\sup_t \dot{E} \le \theta$ we have $\left[\dot{E} - \theta\right]_+ \equiv 0$, so $s_i^{\mathrm{st}}(\Gamma) \equiv 0$. For Γ_ν the envelope derivative is $\nu\dot{E}$, and $\nu \sup_t \dot{E} > \theta$, so by continuity $\left[\nu\dot{E} - \theta\right]_+ > 0$ on an open interval. Convolution with a non-negative, not identically zero pulse kernel h and multiplication by $\sigma_i G \neq 0$ give a drive that is not identically zero. Because $\mathcal{P}_{\mathrm{mim}}(\Gamma_\nu) = \mathcal{P}_{\mathrm{mim}}(\Gamma)$, any map T_i would have to assign both trajectories the same startle drive, which is impossible. The second statement follows because the other channels would then assign both trajectories the same value and the startle term would be the only difference. □

The invariance hypothesis is a property of the earlier operator that has to be checked, and if the projection retained level or onset rate the proposition would not apply. The result states what an operator that carries only curvature and coherence cannot account for, namely the response to abruptness at fixed shape. It formalises the distinction between the sound a

listener understands, which is represented by the reconstructed gesture, and the bodily response, which includes a component driven by features that the gesture discards. A second source of non-commutation lies downstream of the drive. Even when the drive does factor, the response depends on the gains and impedance state of the body, so by Proposition 1 the same gesture can produce different responses in different regulatory states.

8.8 Regimes

The model makes the following qualitative predictions, which correspond to the sources of Section 4. For sustained low-frequency forcing near a bodily resonance, the plant response dominates and is controlled by impedance, so that bracing and relaxing change the response in the manner of Proposition 1, lowering it below the resonance and possibly raising it in a band above. For an isolated transient with fast onset, the response is dominated by the startle term of Eq. (11), which has short latency, habituates under Eq. (12), and is only weakly dependent on the plant. For periodic input in the mid-frequency range, the response is entrainment of the movement pair, with a locking range that grows with drive amplitude as in Eq. (10). For cues that reliably precede events, learned weights on cue features raise the tense unit and the impedance before onset, so that activity precedes the event and can grow with exposure while the startle gain of Eq. (12) falls. For recognised actions and gestural contours, the response is given by the learned weights on the mimetic channels acting through Eq. (9), and it is the only component whose strength should depend on content and on experience. Feedback from bodily state returns to the pair with loop gain μ, which gives a graded response amplified by $1/(1-\mu)$ when $\mu < 1$ and bistability with hysteresis when $\mu > 1$, as in Proposition 3. The model therefore has the property that matters for an account of this kind: different manipulations are predicted to affect different components, and a component that fails to respond to its predicted manipulation is evidence against the corresponding part of the account.

9 Identifiability and Measurement

A model is useful for testing only if its quantities can be recovered from measurements, and a model whose parameters cannot be separated by any proposed experiment would be consistent with almost any result. This section states how the sources and model components correspond to measures, what each proposed experiment is intended to identify, and what it cannot identify.

9.1 Correspondence of sources, model components, and measures

9.2 Constructs and recordings

Each recording has limits that should be stated. The blink component of the startle response, recorded from orbicularis oculi under the standard guidelines, is a nearly direct peripheral readout of the reflex, although its amplitude is modulated by attention and affect, which is the gain modulation that the model represents (Blumenthal et al., 2005; Lang et al., 1990). Recordings from the sternocleidomastoid, trapezius, and masseter capture the spread of the

Source or process	Model component	Principal measures	Experiments
Mechanical coupling	Plant, Eqs. (1)–(4); Proposition 1	Accelerometer transfer function; neck-muscle activity	2 and 3
Startle reflex	Startle term and habituation, Eqs. (11)–(12)	Orbicularis oculi blink amplitude and latency; skin conductance	1 and 3
Entrainment	Movement pair; interaction function	Phase of rhythmic muscle activity or tapping relative to the stimulus	6
Mimetic simulation	Learned weights on mimetic channels; $\mathcal{P}_{\text{mim}}$	Effector-specific motor excitability for recognised action sounds against matched controls	7
Learned anticipation	Weights on cue features; lagged covariance rule	Pre-onset muscle and electrodermal activity in repeated sequences	1 (second phase) and 4
Regulation and feedback	Antagonist commands; loop gain μ; Proposition 3	Co-contraction from electromyography; hysteresis under ramped stimuli	2 and 3
Learning and sharing	Eq. (13); Proposition 2	Acquired response after pairing; convergence across chains	4 and 5

Table 4: Correspondence between the sources of Section 4, the model of Section 8, and the measures and experiments of Section 10.

defensive response, but the sternocleidomastoid also shows vestibularly evoked responses to loud low-frequency sound (Todd and Cody, 2000), so neck recordings cannot separate vestibular from startle contributions without the airborne and structure-borne manipulation of Experiment 3. Skin conductance indexes sympathetic arousal with a latency of seconds and responds to orienting as well as to defence (Boucsein, 2012). Heart rate and its variability reflect both autonomic branches and respiration and cannot be read as a single scale of sympathetic against parasympathetic activity, in line with the multidimensional structure of autonomic control (Berntson et al., 1991; Task Force, 1996). The accelerometer transfer function requires a measurement of the input and depends on posture and contact. Pre-onset activity in repeated sequences is confounded with general arousal and time on task and must be counterbalanced. Continuous subjective ratings are subject to demand effects and are used as convergent evidence and not as primary evidence.

9.3 Plant parameters

Proposition 1(v) turns on the sign of $kM - cC_0$, and the question is whether Experiment 2 can determine it without knowing the effective mass, the effective area, or the absolute scale of the electromyographic measure of activation.

Proposition 5 (Identifiability of the bracing condition)**.** *Let the input-response transfer*

function of the plant, Eq. (1)*, be measured at two or more distinct activation levels. Suppose that the force input is known only up to an unknown positive scale, and that activation is measured as* $w' = \alpha w$ *with unknown* $\alpha > 0$ *and with the relaxed state identified as* $w' = 0$*. Then* $\kappa_0 = K_0/M$*,* $c_0 = C_0/M$*,* $\tilde{k} = k/(\alpha M)$*, and* $\tilde{c} = c/(\alpha M)$ *are identified. The condition* $kM > cC_0$ *is equivalent to* $\tilde{k} > \tilde{c}\, c_0$*, and the location of the maximum of Proposition 1(v), expressed in the measured activation, is*

$$w'^*(\omega) = \frac{\tilde{k}\,(\omega^2 - \kappa_0) - \tilde{c}\, c_0\, \omega^2}{\tilde{k}^2 + \tilde{c}^2 \omega^2}.$$

Proof. The transfer function is $H(s) = (1/M)/(s^2 + (C/M)s + K/M)$, and an unknown positive scale on the force input multiplies it by a constant, so the monic denominator is identified at each activation level, giving $C(w)/M$ and $K(w)/M$. Since $K(w)/M = \kappa_0 + (k/M)w = \kappa_0 + \tilde{k}\, w'$ and $C(w)/M = c_0 + \tilde{c}\, w'$, regression on at least two distinct values of w' identifies the intercepts and slopes. Dividing $kM > cC_0$ by M^2 gives $k/M > (c/M)(C_0/M)$, that is, $\alpha\tilde{k} > \alpha\tilde{c}\, c_0$, which is $\tilde{k} > \tilde{c}\, c_0$. Substituting $k = \alpha M\tilde{k}$, $c = \alpha M\tilde{c}$, $K_0 = M\kappa_0$, and $C_0 = Mc_0$ into w^* of Proposition 1(v) and multiplying by α gives the stated expression. □

The proposition says that the sign condition and the location of the band of increased response are identified from the shape of the transfer function and a calibrated zero of activation, without the effective mass, the effective area, or the scale of the activation measure. The requirement of a calibrated zero is substantive: an unknown offset in the activation measure shifts the intercepts and can change the sign of $kM - cC_0$, so the relaxed state must be defined by a resting recording. If the force is known in absolute units then M is identified separately.

9.4 Startle and habituation parameters

The latency τ_{st} is read directly from the blink latency. The threshold θ is identified by varying the rise time at fixed level and finding the onset rate at which the blink first appears, using stimuli with near-instantaneous rise as in the standard guidelines (Blumenthal et al., 2005). The habituation parameters follow from Eq. (12). Let $P(t) = \left[\dot{E} - \theta\right]_+$ and $A = \int P \,\mathrm{d}t$ be the supra-threshold onset amount of a stimulus, which is computable from its waveform. During a brief onset the first term of Eq. (12) is negligible, so the gain is multiplied by $\exp(-(\alpha/\tau_h)A)$, and between onsets it relaxes as $G \to G_\infty - (G_\infty - G)e^{-\Delta t/\tau_h}$. A series of identical stimuli at varying interstimulus intervals therefore identifies τ_h from the recovery, α from the per-onset decrement given τ_h, and G_∞ from the asymptote at long intervals. The pulse kernel h and the signs σ_i are not separately identifiable from peripheral electromyography, because the response also depends on the downstream pair.

9.5 Learning parameters

Consider acquisition with a single feature of variance $s = \mathrm{Var}(\phi)$. In the setting of Proposition 2 the scalar weight obeys $\dot{g} = \eta\Sigma_u + (\eta\rho s - \lambda)g$, so the acquisition curve approaches its plateau $g^* = \eta\Sigma_u/\mu_s$ at rate $\mu_s = \lambda - \eta\rho s$. The rate and the plateau are identified from one curve, and

Σ_u is measurable from the unconditioned response before pairing, so $\eta\Sigma_u = \mu_s g^*$ identifies η. A single curve does not separate λ from ρ, because only the combination $\lambda - \eta\rho s$ appears. Since μ_s is affine in s with slope $-\eta\rho$ and intercept λ, varying the feature variance across conditions identifies both, and Experiment 4 is therefore specified with more than one level of feature variance.

9.6 What remains unidentified

The internal parameters of the oscillator units, τ_r, τ_a, β, γ, and the input gain κ, are constrained only coarsely by behaviour, through the intrinsic period and the locking range, and they are not individually identified by the experiments proposed. The loop gain μ is identified at best jointly with the input gain, and the hysteresis test discriminates $\mu > 1$ from $\mu < 1$ qualitatively and does not estimate μ. The interaction function of the movement pair is measured empirically in Experiment 6 and not derived from the unit equations. These limits are consequences of measuring a high-dimensional system through a few channels, and they bound the claims that can be made on the basis of the proposed work.

10 An Empirical Program

The experiments below are proposals, specified so that they can distinguish the sources of Section 4 and test the propositions of Section 8. All use within-participant designs analysed with mixed-effects models with random effects structures justified by the design (Barr et al., 2013). Electromyographic recording of startle follows the published guidelines (Blumenthal et al., 2005), electrodermal measurement follows standard practice (Boucsein, 2012), and heart rate and its variability are measured and reported according to the established standards (Task Force, 1996). All exposures should be held within recognised occupational noise limits (NIOSH, 1998), which for continuous exposure are well below the levels at which a startle response can be elicited by brief transients, so that the transient stimuli of Experiment 1 should be brief and infrequent and the sustained low-frequency stimuli of Experiments 2 and 3 should be at levels that respect the exposure limits for their duration.

Experiment 1: sound and picture, fully crossed. The core phase is a fully crossed 2×4 within-participant factorial. Picture is either present or replaced by a luminance-matched neutral field. Sound is one of four versions of the same sequence: the original soundtrack; an attenuated version in which all non-dialogue content is reduced by a fixed amount and dialogue is held at constant level; a version with abrupt effects removed and dialogue and music retained; and silence. This gives eight cells, and the two silent cells are genuine baselines, the picture-absent silent cell for the recording apparatus and arousal drift, and the picture-present silent cell for the visual route alone. Different action sequences, matched for duration and for the number and spacing of abrupt events, are assigned to cells by a Latin square, so that every sequence appears in every cell across participants and no participant sees a sequence twice in the core phase, which prevents prior viewing from contaminating the cell comparison. Dialogue

is level-matched across the three sound versions that contain it and followed by a comprehension probe, a probe about the visible action replaces it in silent cells, and a brief standard exposure to each picture condition precedes the block so that expectancy about the content is similar across cells. Measures are orbicularis oculi and neck-muscle electromyography, skin conductance, heart rate, and a continuous subjective rating of discomfort and arousal. The contrasts identify the sources separately. The main effect of sound relative to silence estimates the acoustic contribution, the main effect of picture estimates the visual contribution, and their interaction tests whether the two combine additively or superadditively. The comparison between original and effects-removed versions within each picture condition estimates the transient-specific contribution, and the comparison between original and attenuated versions estimates the contribution of non-dialogue sound level at constant dialogue. The account predicts the largest electromyographic and electrodermal responses at abrupt events in the picture-present original-sound cell, a reduction under attenuation that leaves a visual contribution, a larger reduction of the transient component than of the sustained component when effects are removed, and retention of transient responses in the picture-absent original-sound cell because they do not depend on the picture. A second phase repeats a subset of the original-sound sequences across blocks. If anticipation is a separate source, repetition should reduce the onset-locked startle component while neck-muscle and electrodermal activity in the interval before known events rises. If attenuating or removing the sound leaves the physiological response unchanged, the acoustic contribution is negligible and the account of Sections 4 and 5 fails in the case it was built for, and if repeated viewing produces no pre-onset activity, learned anticipation is not a separate source.

Experiment 2: impedance probe. Participants are exposed, while either bracing the trunk or relaxing it, to swept-sine or band-limited broadband low-frequency excitation delivered by loudspeaker. The input is measured by a calibrated microphone at the chest and an accelerometer on the platform, the response by accelerometers on the chest surface, and electromyography confirms the level of antagonist activation and of co-contraction. The transfer function between input and chest response is estimated separately under each bracing condition from magnitude and phase, and a second-order model of Eqs. (1)–(3) is fitted to give the resonance frequency and damping ratio in each condition, so that the resonance is estimated and not assumed from the surface displacement at a single level. Proposition 1 predicts that the fitted stiffness, and hence the resonance frequency, rises with bracing level and that the response below the relaxed resonance falls. If the fitted parameters satisfy $kM > cC_0$ it predicts a response that rises in a band above the relaxed resonance, with the band following the shifted resonance, and if $kM \leq cC_0$ it predicts a response that falls at every frequency. Far above resonance the effect of bracing should fall off approximately as the inverse square of frequency. A fitted stiffness that does not change with bracing, or a change in response whose sign is inconsistent with the fitted parameters, would falsify the plant model.

Experiment 3: airborne versus structure-borne stimulation. The same low-frequency content is delivered by loudspeakers, by a seat transducer with the ears protected, and by both.

The mechanical account predicts that the structure-borne route produces somatosensory and vestibular responses without an airborne contribution, that combination exceeds either alone for sustained low frequencies, and that none of these conditions engages the startle reflex unless there is an abrupt onset at high level. A startle-like response to a slowly ramped low-frequency stimulus would indicate that the reflex is being triggered by something other than onset rate. A ramp presented upward and then downward additionally tests hysteresis, which Proposition 3 predicts only for loop gain above one: at the same level the response on the way down should exceed the response on the way up only if bistability is present.

Experiment 4: acquisition of a sound-tension mapping. A novel sound feature that does not itself elicit startle, such as a particular spectral shape or modulation rate, is repeatedly paired in the paired group with a stimulus that raises co-contraction, such as a brief mild tactile event or a visually presented effortful action, while the control group receives the same stimuli unpaired. The learning rule of Eq. (13) predicts that after pairing the sound alone raises co-contraction and electromyographic activity in the paired group and not in the control group, with the effect proportional to the covariance between sound and the unconditioned stimulus and extinguishing as the covariance is removed. If arbitrary sound features acquire no such power under well-controlled pairing, the learned route proposed in Sections 5 and 6 is not supported. The paired groups differ in the variance of the sound feature across at least two levels, so that the acquisition rate identifies the learning parameters as set out in Section 9.

Experiment 5: common exposure and copying. Two kinds of chain are compared. In copying chains, in the diffusion-chain design used for cumulative cultural evolution (Kirby et al., 2008), each participant learns a mapping between sound features and a response from the previous participant's behaviour, with the first participant of each chain starting from a randomly assigned mapping. In common-exposure chains, participants receive the same stimulus stream and the same unconditioned pairings but never observe another participant's responses. Mappings are compared along an intensity axis, with sounds that vary in abruptness and roughness, and a valence axis, and for the intensity axis across response channels: muscular tension measured by electromyography, respiration, vocal effort, and gesture amplitude. The first hypothesis of Section 6 predicts that intensity mappings converge more quickly and to more similar forms than valence mappings in both kinds of chain. The contribution of copying is measured as the excess of convergence in copying chains over the common-exposure baseline, which separates the two routes. The second hypothesis predicts greater convergence when the response channel is muscular tension than for the other channels, and the third predicts that the excess due to copying is larger for intensity than for valence. Valence converging at least as quickly as intensity, no advantage for the tension channel, or no copying excess specific to intensity would count against the corresponding hypothesis.

Experiment 6: locking range and interaction function. Participants are exposed to periodic sound at non-startling levels on a grid of tempi around their spontaneous movement or tapping frequency and at several amplitudes, and the phase of rhythmic muscle activity or

tapping relative to the stimulus is recorded. The locking range, the set of stimulus frequencies for which the relative phase remains constant, is estimated at each amplitude. The account predicts that the range widens approximately linearly with amplitude at weak coupling, as for any phase-reduced oscillator. Single displaced beats are used to perturb the phase and estimate the interaction function Λ of Section 8.3, which tests whether the first-harmonic Adler form is adequate or whether the measured shape of Λ must be used. A locking range that does not grow with amplitude, or failure to lock at any comfortable amplitude, would count against the entrainment component, and an interaction function with strong higher harmonics would show that the Adler simplification is inadequate for the movement pair.

Experiment 7: recognised action against matched control. Pairs of sounds are constructed that are matched in level, onset rate, duration, and spectral envelope but differ in whether listeners recognise them as produced by a particular action, for example a hand-produced scraping sound and a control with the same spectral and temporal envelope that is not identified as action-produced. Participants listen while motor-evoked potentials elicited by transcranial magnetic stimulation, with standard screening for stimulation, are recorded from a hand muscle relevant to the action and from a control muscle, building on work showing that action sounds and speech engage motor systems (Fadiga et al., 2002; Gazzola et al., 2006). The mimetic component predicts effector-specific facilitation for recognised action sounds relative to matched controls, with no corresponding difference in startle measures because the pairs are matched for onset and level. Recognition is verified by a separate identification task, and participants who do not identify the action sounds are analysed separately. The absence of an effector-specific difference would count against the claim that recognition of an action source adds a motor component, and a difference in startle measures would indicate that the matching had failed.

11 Comparison With Alternative Accounts

A framework that claims to unify several mechanisms should say what its closest alternatives predict differently. Five are considered: the mimetic hypothesis taken alone, the multi-mechanism taxonomy of Juslin and Västfjäll, the shared circuits model, the vestibular account of loud low-frequency music, and a reflex-only account of sudden sounds. Table 5 records how each treats the three phenomena that organise the present account. The entries concern the formulations cited and not every possible extension of them.

The accounts make different predictions in specific places. Against the mimetic hypothesis taken alone, the present account predicts bodily responses to transients with no recognisable action source, such as synthetic bursts, and predicts that these habituate, whereas a purely mimetic account predicts neither a response nor habituation of that kind. Experiments 1 and 3 contain these contrasts. Against the multi-mechanism taxonomy, which is compatible with much of the account, the distinctive prediction is the frequency-dependent effect of bracing on transmitted vibration, including the conditional band of increased response of Proposition 1, which has no counterpart in a taxonomy of emotion-inducing mechanisms. Experiment 2 tests

Account	Sudden onset	Low-frequency mechanics	Learned association and sharing
Mimetic hypothesis (Cox, 2016)	Outside its natural scope, since the response precedes identification	Not addressed	Experience with actions shapes simulation; sharing not addressed
Multi-mechanism taxonomy (Juslin and Västfjäll, 2008)	Brain stem reflex	Not treated as a distinct mechanism	Evaluative conditioning, contagion, episodic memory
Shared circuits (Hurley, 2008)	Not addressed	Not addressed	Learned layered mirroring with inhibition of overt imitation; imitation as a route to sharing
Vestibular account (Todd and Cody, 2000)	Not addressed	Sacculus responses to loud low-frequency sound above about 90 dB(A)	Not addressed
Reflex-only account (Koch, 1999)	Gain-controlled and habituating reflex	Not addressed	Habituation only
Present account	Startle term with habituation and anticipation	Plant with impedance regulation	Associative rule; common exposure and copying separated

Table 5: Treatment of three phenomena by the present account and by five alternatives.

it. Against the vestibular account, the present account predicts that neck-muscle responses to loud low-frequency sound include both vestibular and startle contributions, which the airborne and structure-borne conditions of Experiment 3 separate. Against the reflex-only account, it predicts anticipatory activity before known events and the acquisition of responses by sounds that did not originally elicit them, which a pure reflex with habituation cannot produce, and Experiments 1 and 4 test these. Against the shared circuits model, the claims that go beyond it are the acquisition of a sound-to-tension mapping and the hypotheses about the intensity axis, tested in Experiments 4 and 5.

12 Objections and Limitations

The first objection is that the proliferation of mechanisms makes the account unfalsifiable, since any response can be assigned to some source. The reply is that each source in Table 2 has a different time course, a different dependence on source identification, and a different susceptibility to habituation and voluntary modulation, so that each is exposed to a distinct manipulation. The account is falsifiable to the extent that these manipulations have the predicted differential effects, and Section 10 states outcomes that would count against it. It would be unfalsifiable only if the sources were left unspecified, which is the defect of unitary accounts.

The second objection concerns mirror neurons and associated claims. The criticisms of the mirror neuron account of action understanding (Hickok, 2009) and the associative account of mirror properties (Heyes, 2010; Heyes and Catmur, 2022) bear on one reading of the mimetic hypothesis but not on the account developed here, which requires only that sound and sight can engage motor and autonomic circuitry through innate or learned connections. The dependence that remains is on the empirical finding that such engagement occurs, which is supported by stimulation studies and by behavioural work on automatic imitation, and which is independent of the origin of the connections.

The third objection concerns the direction of causation between bodily state and experience. The failure of the facial feedback effect to replicate (Wagenmakers et al., 2016) counsels against any claim that muscular tension generates emotional experience. The account does not make that claim. It proposes that bodily states modulate the gain of subsequent processing, as startle potentiation by affective state shows (Lang et al., 1990), and that they are measurable, and these weaker claims do not rest on the contested effect. The claim that autonomic changes are caused by muscular changes, or the reverse, is left open, and can be addressed only by the measurements of Section 10.

The fourth objection concerns the use of the autonomic framework. Theories that assign distinct evolutionary functions to branches of the vagus nerve have been criticised on physiological grounds (Grossman and Taylor, 2007), and the account does not rely on them. It uses only the general point, supported by the study of autonomic space (Berntson et al., 1991), that autonomic control is multidimensional and that the poles in question are plural. The interoceptive pathways through which bodily state is represented in the brain are a further relevant literature (Craig, 2002), though it is not developed here.

The fifth objection concerns memetics. The status of the meme as an explanatory unit is disputed, and the account does not require it. The claim of Section 6 is restricted: correspondences can become shared by common exposure, which requires no replicator, or by copying, which is the only route called memetic here, and the two are tested separately in Experiment 5.

Several limitations should be stated plainly. The model is a lumped, low-dimensional caricature of a high-dimensional system, and its parameters are not identified by any data. The learning result of Section 8 assumes exogenous stimuli and an approximately linear, quasi-static response of the units, so it does not cover interactive settings in which the stimulus includes other learners' behaviour, nor an entrained oscillating pair. The treatment of the startle reflex as a trigger on onset rate is a simplification of a response that depends on rise time, level, spectrum, and context. The cross-cultural evidence for the transferability of the intensity axis rests on a small number of studies with small numbers of groups. The proposal that parameters are stored in oscillator dynamics is an hypothesis about the form of storage that makes predictions only in conjunction with the learning and coupling assumptions, and it is at present supported by analogy with motor control and not by direct evidence. The essay does not establish the central hypothesis. Its aim is to state the hypothesis in a form that can be tested and to say what the tests are.

13 Conclusion

The extension of the mimetic hypothesis from music to the whole of the auditory environment is most defensible when the mimetic component is treated as one source among several. Mechanical coupling, the startle reflex, entrainment, covert simulation, and learned anticipation are separable by timescale, by dependence on source identification, and by susceptibility to habituation and voluntary regulation, and the organism regulates their combined effect through co-contraction and through opposed rhythmic units. Visual observation recruits overlapping circuitry, so that sound and picture can converge on a single bodily state, and associative learning allows sound alone to acquire the power that picture or reflex first had. Such associations can become shared by common exposure or by copying, and only the latter is memetic in the strict sense. The axis of intensity is a candidate carrier of a transmissible code because it appears to depend least on convention, while the further claims that muscular tension is its privileged realisation and that transmission preferentially preserves it remain conjectures. The formal sketch shows that these claims can be stated precisely, that bracing changes the body's transfer function without cancelling the stimulus and can raise the response in a band above the original resonance, that the proposed learning rule predicts nearby equilibrium mappings under shared exposure only within stated stability and similarity conditions, and that a gesture projection that discards level and onset rate cannot reproduce the startle drive, so that what a listener understands and what the sound does to the body are formally distinct. The claims of Section 2 are graded by evidential status, the identifiability analysis of Section 9 states which parameters the proposed experiments can and cannot recover, and the experiments of Section 10 convert the proposals into measurements, and the outcomes that would count against the account are specified so that the account can fail.

References

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random Effects Structure for Confirmatory Hypothesis Testing: Keep It Maximal. *Journal of Memory and Language*, 68(3), 255–278.

Berntson, G. G., Cacioppo, J. T., and Quigley, K. S. (1991). Autonomic Determinism: The Modes of Autonomic Control, the Doctrine of Autonomic Space, and the Laws of Autonomic Constraint. *Psychological Review*, 98(4), 459–487.

Blackmore, S. (1999). *The Meme Machine*. Oxford University Press.

Blumenthal, T. D., Cuthbert, B. N., Filion, D. L., Hackley, S., Lipp, O. V., and van Boxtel, A. (2005). Committee Report: Guidelines for Human Startle Eyeblink Electromyographic Studies. *Psychophysiology*, 42(1), 1–15.

Blumstein, D. T., Davitian, R., and Kaye, P. D. (2010). Do Film Soundtracks Contain Nonlinear Analogues to Influence Emotion? *Biology Letters*, 6(6), 751–754.

Blumstein, D. T., Bryant, G. A., and Kaye, P. (2012). The Sound of Arousal in Music Is Context-Dependent. *Biology Letters*, 8(5), 744–747.

Bolanowski, S. J., Gescheider, G. A., Verrillo, R. T., and Checkosky, C. M. (1988). Four Channels Mediate the Mechanical Aspects of Touch. *Journal of the Acoustical Society of America*, 84(5), 1680–1694.

Boucsein, W. (2012). *Electrodermal Activity* (2nd ed.). Springer.

Bradley, M. M. and Lang, P. J. (2000). Affective Reactions to Acoustic Stimuli. *Psychophysiology*, 37(2), 204–215.

Brass, M., Bekkering, H., and Prinz, W. (2001). Movement Observation Affects Movement Execution in a Simple Response Task. *Acta Psychologica*, 106(1–2), 3–22.

Brown, P., Rothwell, J. C., Thompson, P. D., Britton, T. C., Day, B. L., and Marsden, C. D. (1991). New Observations on the Normal Auditory Startle Reflex in Man. *Brain*, 114(4), 1891–1902.

Brown, T. G. (1911). The Intrinsic Factors in the Act of Progression in the Mammal. *Proceedings of the Royal Society of London, Series B*, 84, 308–319.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R. J., Zilles, K., Rizzolatti, G., and Freund, H.-J. (2001). Action Observation Activates Premotor and Parietal Areas in a Somatotopic Manner: An fMRI Study. *European Journal of Neuroscience*, 13(2), 400–404.

Burdet, E., Osu, R., Franklin, D. W., Milner, T. E., and Kawato, M. (2001). The Central Nervous System Stabilizes Unstable Dynamics by Learning Optimal Impedance. *Nature*, 414, 446–449.

Buzsáki, G. (2006). *Rhythms of the Brain*. Oxford University Press.

Chartrand, T. L. and Bargh, J. A. (1999). The Chameleon Effect: The Perception-Behavior Link and Social Interaction. *Journal of Personality and Social Psychology*, 76(6), 893–910.

Chion, M. (1994). *Audio-Vision: Sound on Screen* (C. Gorbman, Trans.). Columbia University Press.

Cox, A. (2001). The Mimetic Hypothesis and Embodying Musical Meaning. *Musicae Scientiae*, 5(2), 195–212.

Cox, A. (2016). *Music and Embodied Cognition: Listening, Moving, Feeling, and Thinking*. Indiana University Press.

Craig, A. D. (2002). How Do You Feel? Interoception: The Sense of the Physiological Condition of the Body. *Nature Reviews Neuroscience*, 3(8), 655–666.

Davis, M. (1984). The Mammalian Startle Response. In R. C. Eaton (Ed.), *Neural Mechanisms of Startle Behavior* (pp. 287–351). Plenum Press.

Dawkins, R. (1976). *The Selfish Gene*. Oxford University Press.

Dimberg, U., Thunberg, M., and Elmehed, K. (2000). Unconscious Facial Reactions to Emotional Facial Expressions. *Psychological Science*, 11(1), 86–89.

Donald, M. (1991). *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Harvard University Press.

Egermann, H., Fernando, N., Chuen, L., and McAdams, S. (2015). Music Induces Universal Emotion-Related Psychophysiological Responses: Comparing Canadian Listeners to Congolese Pygmies. *Frontiers in Psychology*, 5, Article 1341.

Fadiga, L., Fogassi, L., Pavesi, G., and Rizzolatti, G. (1995). Motor Facilitation During Action Observation: A Magnetic Stimulation Study. *Journal of Neurophysiology*, 73(6), 2608–2611.

Fadiga, L., Craighero, L., Buccino, G., and Rizzolatti, G. (2002). Speech Listening Specifically Modulates the Excitability of Tongue Muscles: A TMS Study. *European Journal of Neuroscience*, 15(2), 399–402.

Flyxion (2026). Hyperbionic Reading: A Visual Projection of Prosodic Parameter Space. Unpublished manuscript, released in the standardgalactic/hyperbionic-reading repository on 22 September 2026 (commit 5a4b191). `https://standardgalactic.github.io/hyperbionic-reading/hyperbionic_reading.pdf`

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A. D., and Koelsch, S. (2009). Universal Recognition of Three Basic Emotions in Music. *Current Biology*, 19(7), 573–576.

Gazzola, V., Aziz-Zadeh, L., and Keysers, C. (2006). Empathy and the Somatotopic Auditory Mirror System in Humans. *Current Biology*, 16(18), 1824–1829.

Godøy, R. I. (2003). Motor-Mimetic Music Cognition. *Leonardo*, 36(4), 317–319.

Godøy, R. I. and Leman, M. (Eds.). (2010). *Musical Gestures: Sound, Movement, and Meaning*. Routledge.

Graham, F. K. (1975). The More or Less Startling Effects of Weak Prestimulation. *Psychophysiology*, 12(3), 238–248.

Griffin, M. J. (1990). *Handbook of Human Vibration*. Academic Press.

Grillner, S. (1985). Neurobiological Bases of Rhythmic Motor Acts in Vertebrates. *Science*, 228(4696), 143–149.

Grossman, P. and Taylor, E. W. (2007). Toward Understanding Respiratory Sinus Arrhythmia: Relations to Cardiac Vagal Tone, Evolution and Biobehavioral Functions. *Biological Psychology*, 74(2), 263–285.

Groves, P. M. and Thompson, R. F. (1970). Habituation: A Dual-Process Theory. *Psychological Review*, 77(5), 419–450.

Haken, H., Kelso, J. A. S., and Bunz, H. (1985). A Theoretical Model of Phase Transitions in Human Hand Movements. *Biological Cybernetics*, 51(5), 347–356.

Heyes, C. (2010). Where Do Mirror Neurons Come From? *Neuroscience and Biobehavioral Reviews*, 34(4), 575–583.

Heyes, C. (2011). Automatic Imitation. *Psychological Bulletin*, 137(3), 463–483.

Heyes, C. and Catmur, C. (2022). What Happened to Mirror Neurons? *Perspectives on Psychological Science*, 17(1), 153–168.

Hickok, G. (2009). Eight Problems for the Mirror Neuron Theory of Action Understanding in

Monkeys and Humans. *Journal of Cognitive Neuroscience*, 21(7), 1229–1243.

Hoffman, H. S. and Ison, J. R. (1980). Reflex Modification in the Domain of Startle: I. Some Empirical Findings and Their Implications for How the Nervous System Processes Sensory Input. *Psychological Review*, 87(2), 175–189.

Hogan, N. (1984). Adaptive Control of Mechanical Impedance by Coactivation of Antagonist Muscles. *IEEE Transactions on Automatic Control*, 29(8), 681–690.

Hurley, S. L. (1998). *Consciousness in Action*. Harvard University Press.

Hurley, S. (2008). The Shared Circuits Model (SCM): How Control, Mirroring, and Simulation Can Enable Imitation, Deliberation, and Mindreading. *Behavioral and Brain Sciences*, 31(1), 1–22.

Hurley, S. and Chater, N. (Eds.). (2005). *Perspectives on Imitation: From Neuroscience to Social Science* (2 vols.). MIT Press.

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. MIT Press.

Janata, P. and Grafton, S. T. (2003). Swinging in the Brain: Shared Neural Substrates for Behaviors Related to Sequencing and Music. *Nature Neuroscience*, 6(7), 682–687.

Jeannerod, M. (2001). Neural Simulation of Action: A Unifying Mechanism for Motor Cognition. *NeuroImage*, 14(1), S103–S109.

Juslin, P. N. and Västfjäll, D. (2008). Emotional Responses to Music: The Need to Consider Underlying Mechanisms. *Behavioral and Brain Sciences*, 31(5), 559–575.

Kelso, J. A. S. (1995). *Dynamic Patterns: The Self-Organization of Brain and Behavior*. MIT Press.

Kirby, S., Cornish, H., and Smith, K. (2008). Cumulative Cultural Evolution in the Laboratory: An Experimental Approach to the Origins of Structure in Human Language. *Proceedings of the National Academy of Sciences*, 105(31), 10681–10686.

Koch, M. (1999). The Neurobiology of Startle. *Progress in Neurobiology*, 59(2), 107–128.

Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., and Rizzolatti, G. (2002). Hearing Sounds, Understanding Actions: Action Representation in Mirror Neurons. *Science*, 297(5582), 846–848.

Köhler, W. (1947). *Gestalt Psychology* (2nd ed.). Liveright.

Lakoff, G. and Johnson, M. (1980). *Metaphors We Live By*. University of Chicago Press.

Landis, C. and Hunt, W. A. (1939). *The Startle Pattern*. Farrar and Rinehart.

Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (1990). Emotion, Attention, and the Startle Reflex. *Psychological Review*, 97(3), 377–395.

Large, E. W. and Jones, M. R. (1999). The Dynamics of Attending: How People Track Time-Varying Events. *Psychological Review*, 106(1), 119–159.

Leman, M. (2007). *Embodied Music Cognition and Mediation Technology*. MIT Press.

Marder, E. and Bucher, D. (2001). Central Pattern Generators and the Control of Rhythmic

Movements. *Current Biology*, 11(23), R986–R996.

Matsuoka, K. (1985). Sustained Oscillations Generated by Mutually Inhibiting Neurons with Adaptation. *Biological Cybernetics*, 52(6), 367–376.

McGurk, H. and MacDonald, J. (1976). Hearing Lips and Seeing Voices. *Nature*, 264, 746–748.

Meltzoff, A. N. and Moore, M. K. (1977). Imitation of Facial and Manual Gestures by Human Neonates. *Science*, 198(4312), 75–78.

Morton, E. S. (1977). On the Occurrence and Significance of Motivation-Structural Rules in Some Bird and Mammal Sounds. *American Naturalist*, 111(981), 855–869.

National Institute for Occupational Safety and Health (1998). *Criteria for a Recommended Standard: Occupational Noise Exposure* (DHHS (NIOSH) Publication No. 98-126). U.S. Department of Health and Human Services.

Nozaradan, S., Peretz, I., Missal, M., and Mouraux, A. (2011). Tagging the Neuronal Entrainment to Beat and Meter. *Journal of Neuroscience*, 31(28), 10234–10240.

Ohala, J. J. (1984). An Ethological Perspective on Common Cross-Language Utilization of F0 of Voice. *Phonetica*, 41(1), 1–16.

Phillips-Silver, J. and Trainor, L. J. (2005). Feeling the Beat: Movement Influences Infant Rhythm Perception. *Science*, 308(5727), 1430.

Pikovsky, A., Rosenblum, M., and Kurths, J. (2001). *Synchronization: A Universal Concept in Nonlinear Sciences*. Cambridge University Press.

Ramachandran, V. S. and Hubbard, E. M. (2001). Synaesthesia: A Window into Perception, Thought and Language. *Journal of Consciousness Studies*, 8(12), 3–34.

Repp, B. H. and Su, Y.-H. (2013). Sensorimotor Synchronization: A Review of Recent Research (2006–2012). *Psychonomic Bulletin and Review*, 20(3), 403–452.

Robles, L. and Ruggero, M. A. (2001). Mechanics of the Mammalian Cochlea. *Physiological Reviews*, 81(3), 1305–1352.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching*, 10(3), 209–231.

Sherrington, C. S. (1906). *The Integrative Action of the Nervous System*. Yale University Press.

Solomon, R. L. and Corbit, J. D. (1974). An Opponent-Process Theory of Motivation: I. Temporal Dynamics of Affect. *Psychological Review*, 81(2), 119–145.

Spence, C. (2011). Crossmodal Correspondences: A Tutorial Review. *Attention, Perception, and Psychophysics*, 73(4), 971–995.

Strack, F., Martin, L. L., and Stepper, S. (1988). Inhibiting and Facilitating Conditions of the Human Smile: A Nonobtrusive Test of the Facial Feedback Hypothesis. *Journal of Personality and Social Psychology*, 54(5), 768–777.

Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (1996). Heart Rate Variability: Standards of Measurement, Physiological Interpretation, and Clinical Use. *European Heart Journal*, 17(3), 354–381.

Todd, N. P. M. and Cody, F. W. (2000). Vestibular Responses to Loud Dance Music: A Physiological Basis of the "Rock and Roll Threshold"? *Journal of the Acoustical Society of America*, 107(1), 496–500.

Tversky, B. (2019). *Mind in Motion: How Action Shapes Thought*. Basic Books.

von Békésy, G. (1960). *Experiments in Hearing*. McGraw-Hill.

Wagenmakers, E.-J., Beek, T., Dijkhoff, L., Gronau, Q. F., Acosta, A., Adams, R. B., et al. (2016). Registered Replication Report: Strack, Martin, and Stepper (1988). *Perspectives on Psychological Science*, 11(6), 917–928.

Wittgenstein, L. (1953). *Philosophical Investigations*. Blackwell.

Zatorre, R. J., Chen, J. L., and Penhune, V. B. (2007). When the Brain Plays Music: Auditory-Motor Interactions in Music Perception and Production. *Nature Reviews Neuroscience*, 8(7), 547–558.