

Learning Jazz Language by Aural Imitation: A Usage-Based Communicative Jazz Theory (Part 2)

MATTIAS SOLLI, ERLING AKSDAL, JOHN PÅL INDERBERG

Abstract. *How can imitation lead to free musical expression? This article explores the role of auditory imitation in jazz. Even though many renowned jazz musicians have assessed the method of imitating recorded music, no systematic study has hitherto explored how the method prepares for aural jazz improvisation. The article uses Berliner's assumption that learning jazz by aural imitation is "just like" learning a mother tongue. The article studies three potential stages in the method, comparing them to the imitative, rhythmic, multimodal, and protosymbolic behavior of infant perception (building on the works of Stern, Trevarthen, and Merleau-Ponty). The demonstrations of the aural imitation method draw on pedagogic experiences accumulated since 1979 at the Jazz Program at the Norwegian University of Science and Technology. By analyzing structures of behavior suggested by the method, the article indicates key traits that render aural jazz improvisation possible, such as a fundamental sense of rhythm, formation of*

Mattias Solli, PhD, is a Norwegian philosopher. He is a former jazz student and jazz musician, currently working as a researcher in the Department of Philosophy and Religious Studies, Norwegian University of Science and Technology (NTNU). In addition to jazz music, his research interests are phenomenology, hermeneutics, aesthetics, and ethics. Solli is a co-editor (with Ståle Finke) of *Oppløsningen av det estetiske: Kunstteori og estetisk praksis* (2021).

Erling Aksdal is a Norwegian jazz pianist and composer. As a pianist, he has performed and recorded with Chet Baker, Elvin Jones, Warne Marsh, and Lee Konitz, as well as many notable Norwegian musicians. He has served as council member of Association Européenne des Conservatoires (AEC) and was a founding member of AEC's Pop Jazz Platform. Aksdal currently heads the Jazz Program at the Department of Music, NTNU.

John Pål Inderberg is a Norwegian jazz musician and bandleader and a Professor Emeritus of the Department of Music, NTNU. As a saxophonist, he has performed and recorded with Warne Marsh, Lee Konitz, Chet Baker, Bob Brookmeyer, Gil Evans, Chick Corea, and Pat Metheny, as well as with many notable Norwegian jazz musicians. Inderberg taught the saxophone to several generations of Norwegian jazz musicians.

symbolic behavior, joint musical attention, and the facility to “hear via the other.” In conclusion, we critically address a frequent theoretical model describing musical improvisation as a synthesis of discrete elements or building blocks.

This is Part 2 of the article (for Part 1, see JAE 55.4). It continues with the section enumeration.

4. Enhancing the Process

The previous sections in Part 1 (JAE 55.4) began fleshing out Berliner’s assumption, that is, the assumption *that learning jazz by aural imitation is “just like” learning a mother tongue*. The all-important role of rhythm in the aural imitation method suggested the relevance of rhythmic protolinguistic behavior. We indicated how understanding rhythm by aural imitation includes a bodily transformation. The aural imitation method is about to form a general functional ability embedded in an auditory awareness toward general potential in the music and spontaneous behavioral impulses.

Let us introduce the next stage in the aural imitation method to see how the practice suggests more theoretical perspectives (Table 2; table included also in Part 1). Our focus will be on the learning of the Western tonal language. This does not infer that rhythmic learning is considered finished or surpassed, only that we shift attention toward another aspect of the musical language as a whole.

4.1 *Embodying Tonal and Harmonic Relations*

4.1.A. ENHANCED PROCESS

Everything happening in Table 2 is a further development of what we discussed in Part 1. The exercises bring in no new building blocks, but they intensify the mimetic enactment along axes established in the first phase.

4.1.B. ENHANCED MULTIMODALITY

The exercises initiate an enhanced differentiation between the mental and the bodily production of the music, which further reinforces the multimodal coordination tasks introduced in Table 1 (see Part 1). While rhythmic features are played on the instrument, the melodic material is heard in mind only (Instr. 2), and while the instrument is used to break chords, the keynote is continuously heard in the inner ear (Instr. 5).

In addition, the student is encouraged to record the playing (Instr. 4). Recording oneself functions as a reality check; rather than relying on mere intrasubjective evaluations, the student now brings in external documentation of the process.

Table 2. Phase 2: Tonal and harmonic orders and the use of the instrument

Step	Description	Tool
Instr. 1	<p>Repeat the exercises in Act 1a and b (Acts 1–7 are presented in Table 1). Play the individual rhythmic elements and optionally the rhythmic whole that can be performed naturally on your instrument. Use feet and, if possible, voice to complement your playing.</p> <p>Let the basic rhythm of the recording that you imitated in Act 1b “sound” in your ear. After a few minutes of practicing this, emphasize parts of this rhythm on the instrument.</p>	Instrument, singing, clapping, foot-tapping
Instr. 2	<p>Listen to the core melodic material of the composition/recording. Play the melody rhythm while “listening” to your perceived version of the melody line. Select one or more random tones.</p> <p>Repeat Act 4. Listen to your inner hearing while playing the same melodic material on your instrument in multiple registers. Play with different levels of dynamic, timbre, and tempo.</p>	Instrument, singing, clapping, foot-tapping
Instr. 3 and Instr. 4	<p>Listen to your inner hearing’s version of the rhythmic basis of the recording. Play the melody lines from Act 4 simultaneously. Play as if you were singing in the instrument. Play the same melodic material while allowing individual rhythmic elements to be expressed using feet (or voice).</p> <p>Vary dynamic, timbre, and tempo. Remember to make a recording of your playing. Is there a match between your hearing and playing on the recording?</p> <p>Percussionists who do not play a melodic percussion instrument sing the exercises in Instr. 3, Instr. 4, and Instr. 5.</p>	Instrument, singing, clapping, foot-tapping
Instr. 5	<p>The tone that represents the center of the chord (root) is played (see Act 7). One note in each chord throughout the recording should match the root tone in your hearing.</p> <p>Play the four lower notes in all the chords from the root to the seventh. Then arpeggiate the chords up and down in a sequence that you spontaneously decide. Use the full range of the instrument (see Act 7).</p> <p>Make music from this exercise. No exercise should be just technical.</p>	Instrument, singing, clapping, foot-tapping
Instr. 6	<p>Practice your perception of period by, e.g., playing arpeggiated chords (see Act 7b), main melodic lines (see Act 4) or melodic bass lines (see Act 6). Keep the recording’s main rhythm going in feet or voice.</p> <p>Make recordings so you can check if you are keeping the periods. Play diverse sub-elements from this table with varying intensity, timbre, register, and dynamics.</p>	Instrument, singing, clapping, foot-tapping

However, on a more profound phenomenological note, listening to self-recording means a further cross-linking in the student's behavior, namely, between a mode of listening *that is and can only be subjective* and a way of listening to music that is *principally intersubjective*. Although only the student can hear her own inner, mental presentation of the music, the recording documents a musical expression available to others. Any discrepancy between the subjective and the intersubjective both broadens the understanding of one's own game and prepares for actual interaction with other people.

4.1.C. USE OF INSTRUMENT IS SECONDARY TO THE HOLISTIC AUDITORY AND MUSICAL ACTIVITY

The student involves the musical instrument gradually, comprehensively, and in ways fully determined by the music. The instrument is brought in via participation in the rhythmic (qualitative and vital) form, before gradually enacting melodic and harmonic structures.

Most importantly, however, imitational work is still holistic. It involves the whole body. The task of coordinating breath, fingers, and limbs to produce sounds on the specific instrument is not a limited task of its own but integrated into the overall embodied enactment. In this sense, producing sounds on the musical instrument is secondary to the general embodied, expressive engagement in the music. The musical instrument is accidental compared to the musical essentials enacted by the whole body.

Let us say, as a thought experiment, that the instrument had been brought in from the start. In that case, the student would be tempted to figure out the music not from the vantage point of the ear but from physical movement. Mechanical movements would easily be primary, whereas the ear would be secondary. Consequently, the modes of perceiving and replicating music would be limited by contingent embodied factors, that is, by not only individual technical skills but also the instrument-specific ways of being in the music. While chord-instrument players like pianists and guitarists would approach the music through the possibilities of putting down chords vertically, players of melodic instruments like saxophone or trumpet would approach the music horizontally. Habits would stay in the way of the general musical sense unfolding in Adderley's music.

Whereas much pedagogic literature on playing improvisational jazz focuses on the instrument and only indirectly on the ability to hear the musical sense,¹ outspoken aural musicians underline the need to think *ear first*. As Galper states, the idea that the musical instrument is the instrument is an illusion: "*You are the instrument. Everything you are working on is internal to you.*" The body proper is the instrument. "*Your tools are your mind, your body, and your emotions. They are trainable.*"²

Now, in a limited sense, the current usage of the musical instrument can be compared with the famous phenomenological example of a blind person's use of a stick.³ For the blind man, the world becomes available through

the gradual tip-tapping of the stick. He explores and maps out the environment through physical movements, interaction, and sensor-motor feedback. Analogously, the student now uses her musical instrument to explore the musical world. Moving the fingers over the keys of the saxophone or the keyboard of the piano is skillful probing stretching toward possibilities *out there*. Both stick and musical instruments are prolongations of the sensor-motor capacities of the phenomenal body.

The limitation of the analogy marks the difference to the sonic-environmental approach. For, as we pointed out previously, the environment about to be explored by the student is not “just” a sonic environment but a symbolic, expressive language. The skillful probing is a way to explore a communicative system of equivalences—the symbolic system of the major-minor tonality—through the musical instrument. In this respect, the use of the musical instrument is comparable to the moving of lips, vocal cords, and lungs in speaking with others. The musical instrument is part of the communicative, expressive powers of the body proper in the sense that the moving stick is not.

In practice, the implementation of the musical instrument will require much exploratory effort from the student. There might be new sounds or difficult technical passages that need to be figured out.

While the subsequent sections will expand on the music as a symbolic communicative system, we need to set aside the questions about the use of the instrument, in particular, the tricky issues of the relationship between aural and motor intentionality. Suffice to say that, when a child starts imitating the linguistic sounds of people around, motor intentionality (say, of lip movements or guttural sound production) is secondary compared to the expressive drive to *say something* in relation with others. Similarly, the student uses the instrument just to sound “like that” as much as possible, as Galper puts it,⁴ adjusting breath, fingers, lips, and so forth, to the unfolding musical sense.

4.1.D. ENHANCED FOCUS ON TONAL-HARMONIC RELATIONS

The student trains more explicitly to recognize the root of the chord, while at the same time practicing spontaneous voicings of the chords as a whole (Instr. 5). Where the first step (Table 1) established primary auditory contact with keynotes and key relations, the following steps (Table 2) encourage a somewhat freer emulation of the harmonics, exploring the tonal relations latent in the harmonies. In effect, the student begins to juxtapose a vertical and horizontal mode of listening within the harmonics. She strives to hear both how the tones of chords are “piled” vertically and how they can be varied and flexed in voicing while still being the same chord.

As it has been all the way, the point of the exercises is not to plan tonal pathways but to try them out *directly*, with the instant feedback of the increased sensitivity of the aural attention. The student trains to discriminate more

and more precisely within the musical wholes of the major-minor tonality. And in this specific sense, too, the usage of the instrument is subordinate to the aural investigation of the tonality. The student does not use the instrument to hear tonal relations but, contrariwise, to enact the musical relations that she hears.

In this enhanced focus on tonal-harmonic relations, we find the theme that we now need to elaborate on: Where the previous section pursued specifics belonging to rhythm and rhythmic learning, we turn now to the questions about what it implies to learn the Western tonal system, with a particular interest for the generative orders of the tonal system.

4.2 Hearing Generative Tonal Sense

What does it mean to learn the generative syntax of the tonal system in the ways suggested by the exercises described above? To elaborate on an answer, we first need to see what it implies to target the tonal system as a thick audio-perceptual sense (that is, not an intellectual or computational system and not as a system associated with visual notation).⁵ Scruton illuminates how the Western tonal system represents a stringent syntax that is both limited and unlimited at the same time. On the side of the limitation, the system consists of *only* twelve tones. Besides, granted that the music is tonal (that is, not nontonal or atonal, as the musical languages fostered in some of the free-jazz traditions), it is organized by inevitable tensions and releases. In Scruton's definition, tonal music requires that the following four conditions are met:⁶

1. The melodic line feels fully "closed" only when it comes to rest on a certain privileged tone (the tonic).
2. The final move on to the tonic has (in standard cases) the character of a "cadence"—a loosening of tension.
3. Octaves are heard as equivalent—so that the effect of closure is duplicated at the octave.
4. Other tones are heard in relation to the tonic—as more or less distant from it, as tending toward or away from it.

A tonal *key*, thus, represents a perceptual norm or a center of aural gravity. All other notes have a force relative to the center. This brings melody and harmony into close and constant relations. All melodic horizontal lines get harmonic implications relative to the key, and all harmonies link to how the melody and the key are heard.⁷

On the other side, the potential *use* of the perceptual system is unlimited, especially when the equally tempered version of the tonal system is used.⁸ On standardized contemporary instruments, the Western tonal system is equally transposable into all keys, starting on all twelve tones. This possibility gives the keys a certain perceptual indeterminacy.⁹ Each key is defined by a field of possible tones of diatonic and chromatic organization. It unfolds in

relation to the other keys, according to the almost perfect circle of fifths. Each key relates by a fifth to its “neighbor” key, except the key starting on the seventh tone, which relates to its “neighbor” with an imperfect fifth. The melodic order of scales that compose a key generates a synthesis of melodic and harmonic perception. The diatonic scales can be heard as a system whereby the two harmonic affinities—the octave and the fifth—are worked into the very substance of the tonal music, and their intrinsic relation is resolved.¹⁰

The generative potential of this system cannot be “used up.” It can be unfolded in endless musical varieties, a fact that Adderley, Coltrane, Shorter, Bley, Schneider, Bach, Beethoven, Shostakovich, and others demonstrate in their unlimited creative potential of the system. While their works sound distinct and personal, all these artists follow the same perceivable laws for polyphonic musical processes, harmonic relationships, cues, scale relationships, and melodic and chromatic formations.

The potentialities of *this* system are what our student is about to explore. By trying out the various modes of listening within the harmonies, she explores the aural syntax developed through centuries of aural artistic investigations. However, as was the case with rhythm, if she grew up with songs in major-minor tonality, the student does not start from scratch when she starts imitating the record. She already knows the fundamentals. In the womb, the fetus is already surrounded by natural ratios of the harmonic series produced by its mother’s voice, literarily affecting the formation of the auditory capacity.¹¹ Besides, the postnatal rhythmic interactions described above typically have harmonic implications in overtones and melodies.¹² Thus, in a certain sense, the tonal syntax is already part of the ontogenetic and phylogenetic institution of the student’s body schema. She spontaneously hears within the aural gravity of the tonic. Even before she begins the methodological ear training, she perceives something *in* the musical organization, as Scruton puts it above. She hears the moving force, how specific forces in the music propel it in certain directions, as when a tune *moves* toward the tonic. In other words, she already has an aural concept of the generative syntax of the Western tonal system. She hears generative potentials of harmony, the intangible yet normative *something* organizing the music from within.

However, it is one thing to possess a more or less vague preunderstanding of the tonal syntax but another to have a clear understanding of the harmonic system as such, that is, *an understanding of the system qua aural system*. The latter form of knowledge is both a goal and a crucial necessity among professional musicians playing tonal jazz.¹³ To compose and generate music in real time, as Bill Evans puts it,¹⁴ the musician needs a familiarity with the harmonic syntax as profound and nuanced as the mother tongue. The musician needs to *breathe* harmony.¹⁵

Hence, as a future aural musician, the student will need an advanced understanding of how to catch and enact tonality in the spur of the moment,

filling in the adequate melodic and harmonic solutions called for. She will need the facility to perceive—with high precision—the melodic and harmonic implications acted out by herself and others (often many others) while simultaneously responding adequately in the same musical language. In other words, she will need a discriminating sensitivity for what goes on in the music and creative capacity for tonal production. In short, she will need to be fluent in the Western tonal language. She will need to have the whole language at her disposal—and the readiness to use it.

4.3 Symbolic Behavior and Joint Musical Attention

If aural fluency within the Western tonal language is the desired outcome of imitational work described above, what, then, are the critical structures of behavior implied in fluency? If we know the contours of an answer to this question, we can better see how the method works toward that goal and how ontogenetic development prepares for the leap into fluency.

In our context, it makes sense to suggest that a full-fledged aural understanding of the tonal generative syntax must have something to do with the awareness and multimodal fluency associated with forms of vitality. Skilled aural musicians reporting of their tonal fluency seem to have incorporated and internalized the order of the tonal system to the degree that the energy and multimodal fluency of the forms of vitality are transposed into the keys of the twelve tones. That is, they perceive perceptual groupings organized not only due to shared temporal forms but also due to their participation in the tonal organization. The body schemas of the musicians transpose not only between sense modalities but also along with the system of equivalences offered by the tonal system. The human life is potentially transposed into the all the keys of the twelve tones (so to say).

What complicates the analysis, however, is the fact that the harmonic system is not “just” forms of vitality but a generative *system* in the sense that forms of vitality, for themselves (whatever this means), are not. The tonal system is a full *language*, a complex *whole*, whose intrasystemic pathways are strictly organized by the twelve tones and the nonarbitrary rules of the major-minor tonality.

Compared to the forms of vitality described above, the tonal system constrains behavior in new ways. Just like verbal language begins to narrow down the expressive potential of a child around nine months of age,¹⁶ the twelve tones and the tonal syntax of major-minor tonality narrow behavior. Not just anything goes: unless the unfolding musical sense calls for it, a major third hammered out on a minor chord just sounds *wrong*. However, just like verbal language opens new possibilities of expression for the child to express itself vis-à-vis the other, the restrictions of tonality *also* enable full new possibilities of expression and new modes of “being with the other,” as Stern would say.¹⁷ The aural musician has acquired a complete and

intersubjective medium that makes her capable of expressing *herself* in live communication *through* the aural generative syntax shared with others.

To elaborate on the tonal generative facilities in agreement with Trevarthen's and Stern's theories, we conceive the facility as a specialized form of *symbolic behavior*. Now, in other contexts, the words *symbol* and *symbolic* often become associated with intellectual cognition or verbal signs. For instance, Reybrouck associates symbolic with mental representation, thinking, computation, and conceptualization.¹⁸ By contrast, in our usage-based approach, symbolic behavior equals neither intellectual computation nor linguistic expression as such. Instead, it designates the domain-general, perceptual ability to let *something symbolize something else*—in radically open, transformative, and intersubjective relations.¹⁹

Symbols imply the mental ability to grasp something as an invariant under a diversity of aspects and perspectives. Thus symbols imply the ability to grasp something as an *object*, in the phenomenological sense of something that remains invariant through perspectival variation and is graspable for the subject and also available for other subjects.²⁰

Assuming we do not overintellectualize the word *mental* (which is something Thompson thoroughly criticizes) and do not protest against the use of the word *object* in musical contexts (music is not an object, but in the phenomenological sense, it is an *intentional object for the subject*), it seems reasonable to say that aural musicians have cultivated symbolic behavior in the form of spontaneous, real-time music-making. They can grasp and vary harmonic invariants embedded *in* the tonal and rhythmic language. Along the complex, intrasystemic pathways of the rhythmic and tonal syntax, they can catch and enact generative potentialities embedded in the musical flow. They can hear how there is something *in* the sound, something that moves with a force of its own, immediately responding within the lawfulness of the tonic atmosphere.²¹

Thus, aural facilities qua symbolic behavior imply the ability to identify and produce indeterminate yet normative perceptual potentialities emerging within the hearable constraints of the twelve tones and the laws of tonality. The musicians can differentiate precisely how the music *leads* according to the suggestions of the octave, the circle of fifths, the diatonic and chromatic scales, and the rest of aural lawfulness that regulates the polyphonic musical processes. They can catch and enact lateral relations opened by the possibility of varied expressions of the same intrasystematic sense.²²

However, as also indicated by Thompson, cultivating symbolic behavior cannot be reduced to the facility to perceive and manipulate symbols for themselves. It implies an all-important intersubjective dimension. Symbol manipulation is an activity executed *with*, *against*, and *in accordance with* how other human beings perceive the same symbolic order. As Trevarthen points

out above, musicality is a *communicated talent*—a talent for communicating in live, direct, or virtual company with others.

Recall that jazz musicians generally recognize aural facilities as fundamental. Though there will be empirical variation as to how much the idea of auditory interaction is accomplished, there should be no doubt that the musicians who set the standards of the business fulfill the ideal. The recordings of Adderley, Coltrane, Armstrong, Parker, Mingus, Ellington, and Evans are exemplarily prototyped documents of *music heard*.²³ They are aural experts and teachers in the highest artistic and pedagogic sense, to borrow Archie Shepp's phrasing.²⁴ Moreover, it should not be controversial to say that these musicians are masters of the collective and polyphone musicianship first emerging in the African context.²⁵ Their recordings document the spontaneous, auditory, and musical behavior unfolding within the constraints and possibilities of shared musical languages. They document the genuine accomplishment of the human potential for real-time communicating in live and instantaneous communication. (For proof, listen to their recordings.)

Now following Tomasello,²⁶ it seems reasonable to say that the aural facilities of high-skilled aural musicians imply a form of *joint attention*—or joint *musical* attention, as we will call it. Joint musical attention implies the abilities to

- direct attention toward the same music as heard by others. (This is a banal yet crucial condition for collective music-making);
- hear not only how things are played but also how they could be played, that is, the facility to perceive the rhythmic and tonal generative potentials latent in the music;
- follow the musical attention of the other. This implies the facility to perceive the rhythmic and tonal generative potentials about to be acted out by fellow musicians;
- lead the attention of the other toward self-perceived musical potentialities. This brings in a mutual dialectic in the live company. Both parties lead the attention of the other;
- learn through aural imitation. Imitation is not only a propaedeutic concern but conditions also the activity of pursuing the same musical sense as unfolded by peer musicians.

These criteria are modifications of Tomasello's list defining joint linguistic attention, a fact reflecting the shared background of music and language in human musicality. Moreover, while the criteria indicate a self/other relationship, their implications can be rephrased with a focus on the musical sense unfolding between the listening subjects. From the vantage point of the music, joint musical attention implies the abilities to

- hear how perceptual and musical categories of similar and distinct musical gestalts are formed and dissolved;

- form perceptual and musical categories of how similar and distinct musical gestalts are formed and dissolved;
- hear musical transpositions based on similar functional roles of the musical gestalts;
- form musical transpositions based on similar functional roles of the musical gestalts.

To appreciate these points, we need to see the phenomenological correlation between the aural abilities on the side of the listening *subjects* and the generative potentials in their intentional objects, namely, the music. The music is the intermediary reality, the communicative sense that unfolds, reflects, and embeds the aural horizons of the players. Joint musical attention, then, is the ability to hear the rhythmic and harmonic generative potentials in substantial co-perception with others, as the meaningful possibilities unfold in real-time interaction. It is the ability to catch multiple aural horizons latent in the music, emerging because other people hear the same music differently. For, as we recall that humans *always* perceive more or less differently,²⁷ the generative potentials can ever be pulled in different directions in joint musical attention. There is no one way to hear the unfolding music, but many—in fact, as many as there are possible ways of listening.

To illustrate, let's say a pianist suddenly hears a latent sharp nine in an unfolding dominant seventh chord, suggested by the flat intonation of the saxophonist. By subtle manipulation of the music, the pianist can now lead the attention of the others toward this potentiality, perhaps by indicating a substitute chord. The others might respond to this initiative directly or by "transposing" what they hear into a rhythmic structure that, for them, serves a similar functional role in the spur of the moment. The crux of joint musical attention evolves in this facility to hear how rhythmic and harmonic gestalts form and dissolve themselves within the context of living, instantaneous, and polyphone communication.²⁸

Crucially, this description of joint musical attention will prove essential to see where the student's process potentially is headed. Learning the tonal language and cultivating joint musical attention are but aspects of the same process. It is worth noting how our definition of joint musical attention differs from contemporary concepts of joint musicianship. Seddon, Seddon and Biasutti, and Phillips-Silver and Keller seek to clarify how skilled musicians are capable of rapid mind-reading and understanding the plans of the other.²⁹ In criticizing their accounts, Schiavio and Høffding suggest an account of joint musical awareness, focusing on how a group of interviewed musicians is thoroughly absorbed in their doing, with neither time nor need to read the mind or emotional states of the peer musicians.³⁰ However, despite substantial differences in other regards, none of these theorists takes substantially into account how collective musicianship is a *mediated activity*. That is, they do not consider the fact that skilled musicians listen

to each other in and through a thick symbolic communicative medium. The music is just there, so to say, as the neutral sonorous medium between the players.

In our context, by contrast, questions about degrees of conscious awareness of self and others are irrelevant, compared to the fact that skilled musicians listen to each other precisely in and through the shared artistic and communicative medium. The *jointness* in the joint musical attention expresses no direct relation between subjects but is a reflected relation. The musicians listen to the same musical sense, and they unfold it together. They share attention by unfolding the same generative potential from moment to moment.

5. Hearing Symbolic Indications

The previous subsections suggest a series of perspectives on what it implies to learn the Western tonal system by ear and to be able to use the language in live communication with others. Along with these perspectives, the exercises described above are a way to train symbolic behavior. Importantly, rhythmic learning has embedded a symbolic dimension. In cultivating her overall capacity to identify and vary coherent rhythmic unity across the multimodal variation, the student has enacted subtle modes of selfsameness under a diversity of aural perspectives. In Thompson's definition, she trained to grasp something as an invariant under a variety of aspects and perspectives.

However, with the exact replication of the tones and relations between tones (Table 1) and the slightly more spontaneous variation encouraged in Table 2, symbolic learning takes a more specified form, as the student is about to explore the aural selfsameness and equivalences suggested with the tonal atmospheres. She tries out the various modes of listening latent within the keys. She is about to hollow out the multiple, systematic pathway variances implied in the tonal invariances. In other words, the student trains her capacity to catch and enact the generative harmonic potentials. She tries to grasp the invariances unfolding *in* the tonal language, something that moves with a force of its own, propelling the music in various directions.

In Scruton's parlance, the student tries out how the tonal atmosphere of the chords and the sequence harbors latent possibilities of variations within the musical atmosphere of their functioning.³¹ She tries out how the octaves can be heard as the "same again" across the tonal differences, how various voicings of the II-V-I cadences create slightly different tensions and releases in the music while still fulfilling the same functional roles, and how the harmonic and melodic potential of the tonic separates into other diatonic and chromatic relations. In other words, she tries to hear how tones and relations between tones belong to the same tonal atmosphere and how the tonal

atmosphere can be strengthened or weakened by tone sequences moving more or less distant to the key, tending toward the center or away from it. She tries to hear unused potentialities within the tonal atmosphere—foreign tones excluded from the principal regions of the key, also creating transitions and potential tensions within the atmosphere. In short, the student explores the tonal language as a perceptual system of equivalences sketching out the generative potentials embedded in the Western tonal language.

The all-important crux, then, is to see how this mode of symbolic learning has an intrinsic dimension of intersubjective and communicative relatedness. In line with the general argument of this article, the student is not replicating and using an abstract, computational generative system. She is about to learn a genuine symbolic language used in real-time communication between real human beings, namely, the Adderley Quintet. By replicating, enacting, and incorporating *their* music, she is about to imitate and incorporate their means of communication in joint musical attention, that is, their aural indications embedded in the harmonic (and rhythmic) generative potentials.

Herein evolves new profundities implied in Berliner's assumption. As we will come to see, this mode of symbolic learning involves structural similarities to a new phase in ontogenetic language development related to the entry into verbal, symbolic language by the age of seven to nine months. By exposing these factors (everything we invoke will eventually turn out relevant), we will finally be ready to conceive the jucier aspects of how the aural imitation method works and how imitating recordings prepares for collective music-making.

5.1 Generalized Interactions and the Indicative Role of Joint Attention

To see what resources Table 1 and Table 2 cultivate into joint musical attention and fluent use of the tonal language, we need to invoke perspectives explaining how the ability of symbolic behavior starts showing itself and how it forms into linguistic capacities. Considering the syntactical dimension of the tonal system, we also need aspects belonging to the learning of linguistic grammar.

In the developmental narrative introduced by Trevarthen³² and Stern,³³ symbolic behavior begins to show itself soon after birth. That unity in rhythm and forms of vitality emerge across the many sense modalities indicates a certain selfsameness or generality in the constant variance of life is indeed a protoversion of symbolic behavior. Besides, rhythm and forms for vitality, too, have directions. They *lead* somewhere; they indicate.

However, encircling the mode of protosymbolic generality further (in a way that will prove fruitful to capture the generality of aural learning), Stern suggests the term RIG, an acronym for "Representations of Interactions that have been Generalized": "[I]nfants have some abilities to abstract,

average, and represent information preverbally."³⁴ From two to seven months, together with the formation of the core self, they begin to show capacities to aggregate experience and distill, or abstract out, an average prototype out of perceptual variety. "RIGs are flexible structures that average several actual instances and form a prototype to represent them all."³⁵

RIGs have a somewhat ambiguous status when it comes to representing brute reality. On the one hand, the distilled prototype represents *all* the events that made it come into being. On the other hand, the prototype does not represent *any* of the events as such. That is, the RIG does not necessarily correspond to anything in a one-to-one sense: "A RIG is something that has never happened before in exactly that way, yet it takes into account nothing that did not actually happen once."³⁶ The ambiguity regarding "realness" makes up a flexible and generative moment. Things and events are indexed and reindexed in a fluid and dynamic fashion. Attributes of many different kinds gradually form meaningful networks. Invariants emerge in the constant variance of perceptual life.

Now, put simply, RIGs are protostructures for *words*. When the child is eighteen to twenty-four months and begins to use and understand linguistic symbols, this process is a continuation of the generalization that has been going on for a while. Language fills in the need for more advanced representations and communication of experience across the constant change of perceptual singularity.

Crucially, the ability for joint attention is critical for the process from RIGs to linguistic symbol manipulation. In Trevarthen's framework, joint attention represents a shift from *primary* to *secondary intersubjectivity*, emerging around nine months.³⁷ At this age, the infant moves from coordination of self and others based mainly on timing, form, and intensity, to the inclusion of objects and more explicit engagement in cooperative exchange of referential gestures. The infant shows increased initiative-taking to the systematic combining of purposes to partner and object. The infant begins to generate meaningful acts in a new sense, such as rudimentary demands, refusals, and inquiries, or more awareness of objects. "We are born to generate shifting states of self-awareness, to show them to other persons, and to provoke interest and affectionate responses from them."³⁸

In Stern's framework, the process of sharing attention starts around seven to nine months, when the infant develops a new sense of self, which Stern calls the *subjective self*.³⁹ The infant now begins to show a new awareness of self vis-à-vis others. It "discovers" that there are other minds out there as well as its own.⁴⁰ Self and other are no longer only core entities of physical presence, action, affect, and continuity; they also include mental states, such as feelings, motives, and intentions, things that lie beyond the physical happenings in the domain of core-relatedness. The infant shows a new organization of the subjective perspective, defined by a qualitatively new sense of

self vis-à-vis the other. Mental states can be “read,” matched, aligned with, or attuned to, in a more articulate sense. The infant shows capacities for sharing a focus of attention, for attributing intentions and motives to others and apprehending them correctly, and for attributing the existence of states of feeling in others and sensing whether they are congruent with one’s states or feelings.⁴¹

The gesture of pointing and the act of following another’s line of vision are among the first overt acts that permit inferences about the sharing of attention or the establishment of joint attention.⁴² While the infant shows a preliminary form of the ability to follow the gaze directions of others *before* nine months, the ability to share attention suggests a new ability to perceive *pointing* matures around that age. The child still maps out affectively and *very closely* the behavior of the mother but also begins imitating the *goal* of the mother’s actions in another sense. “To imitate is not to do what the other does, but to arrive at the same result,” as Merleau-Ponty would say.⁴³ When the mother looks in one direction and the child does the same, the child does *not copy* the movement as such but imitates the aim of the looking gesture, which is to attend to the same as the mother.

Interestingly enough, recent studies accentuate precisely the auditory aspects of this early ability to share attention. Launching the term *telecommunicationality*, Schiavio, van der Schyff, Kruse-Weber, and Timmers demonstrate how infants between six and ten months begin to *aim* toward something in the sounds of the surrounding.⁴⁴ The goal seems to be both to create a non-boring and meaningful environment for its own sake *and* to share this auditory sense with the mother. By trying to imitate the sounds of the mother, infants develop their repertoires of goal-directed actions that will allow them to explore the environment in a meaningful (for example, musical) way. At the same time, infants respond creatively to the sonic situation. By mastering certain actions, they also develop adequate perceptual abilities that seem to motivate further ways of interaction with the world. That is, by understanding the goal of a given sound-related activity performed by another individual (for example, the caregiver or peer), infants could typically begin to play with their own sounds as invitations to more mutual understanding embedded in the auditory phenomena.⁴⁵

Joint attention implies both ability to transcend egocentrism by decentering attention into the interest and purposes of the mother and, the other way around, the ability to steer the attention of the other, thus expanding the “control zone” of the ego.⁴⁶ Differently put, together with the ability to *follow* another person’s attention to distal objects and events outside immediate interaction comes the reciprocal ability to *direct* the attention of others to distal objects by pointing, showing, and using other nonlinguistic gestures.⁴⁷

The child engages in the process: it validates whether joint attention has been achieved, and if it has failed, the child will initiate more interaction to

gain the joint perspective. Thus, joint attention is not passive reception but active participation *with, against, and according to* the other's way of perceiving things. The child begins a dialectic mediation of perspective. It becomes a subject in an interpersonal exchange of perspectives. In other words, a new dimension emerges in what we call the intermediary third. The thirdness takes up the distinct perspectives of partners.

In Stern's framework, the leap into the mutuality of *my* and *your* attention indicates the rudimentary formation of a new self, *the verbal self*.⁴⁸ The indicative dimension of joint attention is the key to the development of what later will turn into an explicit understanding of verbal symbols. It is also at the core of understanding generative rules and procedures for interactions.⁴⁹ Within the framework of affective attunement and mutually created meaning, the child will gradually recognize stable traits in "how we do things." It will begin to identify complex invariants in behavior and to perceive how these invariants contain latent potentialities of novel behavior. It will become aware of how overt behavior is one of several possible manifestations of the same, as Stern puts it.⁵⁰ In other words, the child will become aware of how something can symbolize something else beyond the concrete reality of RIGs. It will perceive how each manifestation of something has some degree of substitutability and potential variability latent in how others perceive the same object.

Naturally, it falls outside the scope of this article to pursue further how the child acquires verbal language of words and grammar. Suffice to say that what Tomasello calls early grammaticalization is rendered possible by joint attention.⁵¹ Gradually and then from inside the rhythmic communication full of RIGs and exchange of gaze and expressive gestures, grammatical and generative structures of the mother tongue will emerge for the child as organizing forces embedded in communication. The child will need no meta-awareness of the syntax qua syntax but will eventually pick up general ways to generate utterances. He or she will follow the direction of the gesture—the sense of the movement, as Merleau-Ponty would say: "[B]eginning with the first phonetic oppositions, the child *speaks*, and only afterward will he learn to apply the principle of speech in diverse ways."⁵² From the inside of the phonetic dimension of communicative sense, the child will begin to apply general principles embedded in the communication.

5.2 Guided by Music

As we return to the aural imitation method, we recall first how the student and teacher collaborate in the process of exploring the music. By listening together, they genuinely share attention. They mutually guide the other's attention toward nuances in the music, either by saying things like, "Listen to how the saxophone phrases here, compared to here," or raising an eyebrow after a hefty passage or just by singing or playing what they have

in mind. They can indicate or point toward specific nuances in the music, accumulating a growing auditory awareness within the shared musical language.

On a more intriguing and potent level, however, it now makes sense to say that the student also shares attention with the musicians she listens to on the record—not directly, of course, as if they were standing next to her, *but in a mediated sense*. That is, she does not necessarily engage in empathetic attunement with the musicians (as Seddon would suggest⁵³), nor does she use “mental imaginary” to plan the productions of her own sound or to predict the upcoming sounds of the players heard on the record.⁵⁴ She “just” listens *within* the same musical language as the Adderley combo once used in live communication. She lets herself be guided by musical indications embedded *in* the music as an intermediary third. Gradually, she can become aware of indications unfolding in subtler details of the music, and through the enactive efforts, she can become capable of using similar musical indications. Finally, we see also why the student could have picked a solo performance as a model for imitation. Solo performances unfold a language of indications.

Let’s say our student has imitated and incorporated (very precisely) how the bass plays the root in the falling fifths relations of “Autumn Leaves” and how the piano, trumpet, and saxophone voices out the other tones of the chords vertically and horizontally (Acts 6–7, Table 1, Phase 1), before she now begins to vary the voicings according to Instr. 5 (Table 2, Phase 2).

The current point is this: The student does not perform these forms of enaction in a void but *relative* to the music once carried out by Adderley and his combo. She directs her attention toward the same music as they did at the time of recording. But she also trains her ability to hear and accomplish other latent possibilities in the same music. She tries to hear not only how the generative potentials of “Autumn Leaves” once were heard and fulfilled in a recording studio in New York but also how the potentials *could* have been heard and carried out. In other words, she does not *invent* new options for the musical language as much as she tries to fulfill possibilities latent in the musical language. These “dormant” or “quiescent” pathways are already part of the musical organization. All she has to do is bring them forth by hearing them and acting them out.

In our framework, it makes sense to say that the ability to hear distal musical potentialities is rendered possible by the protosymbolic imitation of joint attention. Without the ontogenetic formation of joint attention, it is improbable that the student would be able to follow indications within a complex tonal system. Without the ability to understand the directing of pointing, she would probably not be able to hear unaccomplished, intrasystematic leads unfolding in the musical language. If she had not experienced

the mutual dialectics creating an intermediary third, she would not have had the ability to hear and generate musical potentials in real-time polyphonic synchronization with other subjects.

At the same time, it makes sense to say that the aural imitation method implies a further development of the student's general capacity to follow the attention of others. She is "forced" to expand on her abilities to follow subtle indications embedded in the intersubjective symbolic language. The music pushes her into a general regrouping in her ways of perceiving general sense latent in the concrete auditory unfoldment.

In prolonging our previous observations regarding rhythmic perception and understanding, we note how the auditory attention toward the harmonic sense also harbors a potential proprioceptive dimension. Discovering new pathways in the music is coextensive with the student's discovery of new channels in her behavior. The energy, spontaneity, and self-relation associated with forms of vitality are about to be channeled into the symbolic sense of the tonal language. Multimodal behavior is about to acquire a new medium of expression within the stringent pathways of the tonal language. She has to discover and establish new latent behavior within herself, letting adjustments of motor excitation participate in the harmonic generative potentials of the music.

5.3 *Learning the System as a Whole*

As we know from the previous sections, it is one thing to hear and enact local or partial latencies belonging to the tonal language and another to have at one's independent disposal the whole tonal language in the ways associated with tonal fluency. How, then, do the exercises of Tables 1 and 2 help the student crack the system as a whole? Alternatively, put in a quasi-Saussurean way, how can she construct a tonal langue ready to be transformed into musical parole at any time, without the support of an external audible source like the record or the instrument?

One key evolves in Stern's RIG concept. Recall how RIG conceptualizes how learning, already from the early and protolinguistic phase of ontogeny, involves a certain generality embedded in the concrete. That is, long before she learned to handle general categorizations of words and linguistic symbols, the student had begun to distill, or abstract out, average prototypes out of perceptual variety. She had started to form flexible structures that averaged several actual instances and formed a prototype that represented all the pertinent situations.

In the aural imitation context, something similar must somehow take place. The student must hear that the partial tonal sequences and generative potentials that she imitates embed a generality—namely, the generality of the *whole* tonal system. Although it is utterly impossible to hear through *all*

possible pathways in the tonal language, it *is* possible to grasp the general functionality of the whole system.

The harmonic falling fifth relation characterizing “Autumn Leaves” gives the student (and us) a hint. By what we said above, this harmonic progression makes use of one of the characteristic features of the Western tonal system: the circle of fifths. Simply put, “Autumn Leaves” makes abbreviated, musical use of precisely this trait of the system.

Herein unfolds the generality that the student somehow must grasp in the audible phenomenon: she needs to hear how the concrete use of generative potentials manifests a systematic *part* of a systematic *whole*. That is, she needs to hear how the tune exemplifies or “symbolizes” the generative syntax of the major-minor tonality, the almost complete circle of fifths, the twelve tones, and the diatonic and chromatic transitions. In other words, she needs to hear how there are practically unlimited aural possibilities embedded in the limited sequences. The individual sequences and the generative potentials embed and exemplify the generative syntax of a system that ultimately transcends the concrete realizations of *that* tune.

While it is easy to overintellectualize the learning process involved in the potential cracking of the whole system, Stern’s RIG helps us keep track of the perceptual and embodied mode of learning. The ways that intangible wholeness of the system is present in “Autumn Leaves” is structurally similar to how each manifestation of a RIG represents the average or distillation of several of several experiences. For the student, the system as a perceptual whole is something that never actually can be heard (how could it be?), yet the concrete tune that she is about to imitate takes into account nothing that does not belong to the whole system. In other words, “Autumn Leaves” is what Stern would call a perceptual prototype. It is a distillation of the transgressing, imperceptible whole.

Exposing the RIG point further, we consider briefly a study discussed by Merleau-Ponty as our point of departure. Against the general background that children often learn to discriminate colors relatively late, a group of young children was asked to distinguish objects with a small set of colors.⁵⁵ The moment the children first learned to recognize and name two or three colors, the Sternian generality of RIGs kicked in. Suddenly, at some moment, the children were able to identify and discriminate *colors*:

[W]hat is acquired is not properly speaking the discrimination of these two qualities as such; it is a general power of comparing and distinguishing *colors*: all pairs of colors benefit from the distinction of red and green and differential behavior progress not from one to the other, but by a finer discrimination with regard to all of them.⁵⁶

In the same strike of understanding, the children had learned to *see* more distinctions and to name these distinctions. The color RIG is generalized,

nuanced, and “symbolized,” so to say. The combination of grown-up guidance (“look here and try to distinguish this from that”) and a limited number of examples helped the children broaden their sensual, qualitative, and symbolic form of perception.

Now, just like the children learning to sort colors, the jazz student has narrowed down the options of variations. The children in Merleau-Ponty’s case handle *three* colors, not endlessly many colors, and *this* limitation helps them understand the perceptual generality *exemplified* by the colors. Similarly, the student handles *one* recorded tune organized by the Western tonal system (not many recordings at once), and *this* specific limitation is the vehicle that potentially will make her able to perceive the system as a whole. Moreover, in analogy to Merleau-Ponty’s case, this general learning will go hand in hand with an increased awareness for perceptual details. Just like the child learning to *differentiate* between colors, the student will begin to distinguish within the tonic atmosphere in more precise manners. Diatonic organizations harbor chromatic transitions, and V–I sequences harbor a potential II–V7, or whatever.

In contrast to Merleau-Ponty’s case, however, the generality potentially grasped by the student is a symbolic network of *sense generating sense*, that is, of a complete, generative syntax. By the stringent constraints discussed above, each new discrimination within the system will potentially lead to more nuances regulated by the same order. Moreover, in contrast to Merleau-Ponty’s case, the ultimate guide to this latent generality unfolds in the perceptual phenomenon itself, that is, *in the music*. Adderley’s recording harbors its own explanation: it contains indicative forces powerful enough for the student to crack more or less the whole system.

The indicative role of the music was implied when we said that the student joins attention with the recorded musicians *in a mediated sense*. Their use of the tonal language is indicative: *how* the musicians unfold the concrete generative potentials can lead the student’s ear in the direction of the general invariant syntax of the tonal language. Analogous to how the student as a child once picked up the invariants of linguistic grammar without necessarily being *told* that “this is grammar,” she can pick up the qualitative indications toward generality unfolding in the flow of the music.

Let’s say the student has yet to make the final discovery. She manages to swing and hear and generate relevant tonal sequences but has not cracked the system as a whole. This will come in the next section.

6. Subjective (Personal) Variation

Before we say more, we turn to the final stage of the aural imitation method (Table 3, identical to Table 3 presented in Part 1).

Table 3. Phase 3: Subjective (personal) variation

Step	Description	Tool
Pers. 1	Let the basic rhythmic foundation (Act 1b) sound in your ears and sing, play, and beat new rhythmic ideas/patterns simultaneously. This must occur spontaneously. Vary the length of ideas from short motifs to longer themes. In this exercise, play harmonically and melodically freely. Make recordings of your own playing. What do you like and dislike?	Instrument, singing, clapping, foot-tapping
Pers. 2	Pick some melodic motifs from the recording. Sing this melody line along with the recording so many times that you know it by heart. Play it together with the recording. Play the same thing alone while the rest of the recording sounds in your ear. Sing and then play spontaneous melodic lines. Be sure to convey ideas from your own musical imagination. Record your playing along with the recording as well as your soloing. What do you think of your own playing? What do you want to change? Percussionists: Sing exercises in Pers. 2.	Instrument, singing, clapping, foot-tapping
Pers. 3	Concentrate on the harmonic progression of the recording. Sing the chord sequence you hear on the recording. Break the same chords in arpeggio exercises on your instrument. Let the rhythmic foundation of the recording go into your ear while you spontaneously play new chords. Preferably, piano is used, but singers and melodic instrumentalists sing/play broken chords in inversions of their own choice. Set your harmonic imagination free.	Instrument, singing, clapping, foot-tapping

6.1 *Enhancing the Process Even Further*

Again, everything encouraged in this third phase of the aural imitation method is a further development of the process started with the earlier stages presented in Tables 1 and 2 (partly discussed in Part 1, and partly in the Sections 4–5 above.). We pay extra attention to the following features.

6.1.A. *STRENGTHENED MENTAL HEARING*

The student enhances the facility to let the music resound in her mind. She pays attention to and explores how the mental or “intrabodily” music sounds relatively independent of the real voice and the instrument. She tries it out in cross-modal behavior, acted out relative to stamping, clapping, singing, and the use of the musical instrument.

6.1.B. *STRENGTHENED PERSONAL VARIATION*

More explicit than before, the student is now encouraged to act out personal variations within the musical language spontaneously. By doing so,

she stretches the limitations and possibilities of rhythm, melody, and harmony—not by thinking and cognitive evaluation but by impromptu enactment. Rhythmic and melodic motifs from the record function as springboards for embellishments of various lengths.

6.1.C. ENHANCED DYADIC CHARACTER OF ATTENTION

Viewed together, everything we have said here implies a strengthening of the dyadic structure of the auditory attention established above: The mode of attention going “inwards” (toward the self-produced mental music) and “outwards” (toward the self-generated musical variations hearable to others) are ways of releasing and hearing one’s spontaneity *relative to* the music that streams out of the loudspeakers. The strengthened contact with the ultimately subjective intramental way of listening goes hand in hand with the intensified contact with the expressive behavior that blends into the intersubjective resounding music.

In other words, for the mode of aural attention encouraged in these exercises, the *personal* ways of enacting the music imply no conflict between the “inner” and the “outer”—just as there is no conflict between the self-generated music and the music streaming from the loudspeakers. The “inner” and the “outer” are not poles of expression but rather dimensions of the same musical sense being explored and expressed. Everything accomplishes the same musical language: Every mode of behavior is played out with, against, and in accordance with each other—and relative to the music that resounds from the record.

7. Leaping into Fluency

To analyze the implications of Table 3, we need not import many new theoretical perspectives but rather unpack how the potential outcome of the exercises accomplishes aspects belonging to an intersubjective relatedness that has been there all the way. In the introduction, we noted how musicality, according to Trevarthen,⁵⁷ is a communicated talent—a talent for communicating in live company. Through the critical investigations of Berliner’s assumption, we have tried to indicate how intersubjective relatedness is part of every reasonably normal childhood in the forms of protoconversational, rhythmic interaction and the gradual formation of joint attention.

Based on what we have seen, it makes sense to suggest that imitating and incorporating the rhythmic and tonal forms of the music by ear works because the method stimulates, cultivates, and canalizes energies associated with the imitative relatedness of musicality. The imitative method utilizes, animates, and transforms general musical potential into a specialized, aural skill. The fundamental ability to move rhythmically and in self-generated behavior and the ability to hear and generate symbolic orders within the

tonal system were already there, formed and carried out elsewhere in the general ability of joint attention and linguistic facilities.

However, as we might remember, we left our student a little hanging in the previous section. She had begun to hear latent possibilities within the tonal language but had not crossed the symbolic threshold for real. She could hear latent pathways embedded in the tonal syntax but had yet to crack the tonal grammar of the whole. How, then, can the exercises described in Table 3 help the student take the leap into fluency? What, more specifically, would the leap have to do with the intersubjective relatedness which has been there all the way?

One crux of an answer evolves in the intimate relationship between the dyadic character of the aural attention and the indicative aspect of joint attention exposed in the previous sections. Recall how the formation of joint attention implied both a new ability to transcend egocentrism by the decentering of attention into the interest and purposes of the other and a more active ability to steer the other's attention and in that sense expand the "control zone" of the ego. In other words, the child learned to *be manipulated* by the behavior of the other and to *manipulate* the behavior of the other. One general goal of the manipulative activity was to achieve and sustain joint attention, which seems valuable for its own sake for the infant (and the mother).

In effect, striving with the exercises in Table 3, the student now seeks a similar mutuality. For as we just saw, when she varies the music according to her spontaneity, she *relates* to the music through her independent manipulation of it. At the same time, having her ears and body glued to minute details in the music streaming out of the loudspeakers, she picks up how the musical voices relate both to each other *within* the band and to herself as a potential listener. She picks up how subtle manipulations and negotiations go on in the group and how each initiative both singularly and collectively manipulates her spontaneous behavior. In other words, she becomes aware of the aural, dialectical encounter between self and other and how the intermediary music dynamically propels behavior in all parties.

This attended encounter of aural indications and reciprocal manipulations evolved what we could call the key to the tonal fluency and the general understanding of the tonal system.

7.1 Hearing via the Other

Let us say this alteration happens: Suddenly, the student *hears* the general, harmonic glue that holds the tonal organization together. She hears how the same tonal lawfulness organizes every contingent tonal initiative coming from herself and the others and how the generative lawfulness ultimately transgresses the contingent manifestations. She hears how *these* precise phrases and chord progression exemplify the general order of the tonal system *as such*.

Simultaneously, she realizes that her earlier attempts to enact the tonal orders came a little too much *from herself*. She was ego-centered, as Stern would say.⁵⁸ She listened to the tonal generative potentialities mainly from her subjective aural perspective. By contrast, after the transformation, she is more *alter-centered*. She relates differently to the whole music. She “takes in” how Adderley and his band relate to each other differently. This is the moment when she “falls into” the tonal fluency. The musical sense “comes to” her—just like verbal language comes to her in a spontaneous conversation with a listening and responding friend. What has happened?

To unpack the structure of the potential transformation, we turn briefly to the philosophical literature on understanding, which reflects the transformative encounter between self and others in ways that now seem pertinent. “It’s enough to say that we understand in a *different way if we understand at all*,” writes Gadamer.⁵⁹ Understanding *differently* means, for Gadamer, a genuinely nondirected openness for how other humans understand: it “involves recognizing that I myself must accept some things that are against me, even though no one forces me to do so.”⁶⁰

Merleau-Ponty makes a similar point. To understand another human in a conversation is to *speak via the other*:

When I speak [and] understand, I experience the presence of others in myself or of myself in others. . . . To the extent that what I say has meaning, I am a different “other” for myself when I am speaking; and to the extent that I understand, I no longer know who is speaking and who is listening.⁶¹

The *other* discussed by Gadamer and Merleau-Ponty *can* be a real person, but it can also be a fictive, generalized other, located in oneself, so to speak.⁶² The point is that the person who *understands* must decenter from a mere private way of conceiving things. Understanding implies a genuine openness to different ways of perceiving.

Gadamer and Merleau-Ponty stress how openness to the other implies the emancipation of what we, with Benjamin,⁶³ call the intermediary third. Where neither of the parties control the human encounter of perspectives but contribute and attune to the other’s ways of understanding, the intermediary is allowed to play itself out, according to its intrinsic norm. The intermediary is accomplished in full-fledged symbolic form. Something is allowed to emerge, something that remains an invariant through an unlimited amount of perspectival variation.

Our previous sections have prepared for Gadamer’s and Merleau-Ponty’s points. The rhythmic interaction and forms of vitality implied implicit relational knowledge—a sense of direction in synchronized behavior. Joint attention conceptualized the sharing of perspectives, and the concept of joint musical attention exposed the structure of the mutual aural and indicative dialectics among fluent aural musicians. What Gadamer and

Merleau-Ponty currently help us indicate represents the accomplishment of the relational competence. The more mature ability to speak “via the other” is rendered possible by the early development. The formation of the verbal self and the capacity for intersubjective relatedness enables the more mature distribution of aural and musical sense.

At the same time, Gadamer and Merleau-Ponty will now help us encircle how understanding the general syntax of the tonal system and the ability to *hear* and generate tonal sense “via the ear other” are *genuinely aspects of the same transformation*.

In our context, it makes sense to say that, by attending to her spontaneous variations of the music heard in mind, the student is already relating to a fictional or depersonalized other—a way *one* could listen to the music. Simultaneously, she relates to the music produced by real others, which is to say, Adderley’s combo. She actively relates to the auditory product emerging when these exact individuals once practiced joint musical attention mediated by the tonal language. So construed, the student is already engaged in a pluralism of distinct aural horizons. She actively relates to multiple highly specific ways that a human ear can enact the generative potentials of the Western tonal system.

Moreover, in the process of trying out personal variation, the allocentric, open, and nondirective mode of aural attention allows everything to blend. It becomes unclear *who* is listening and who is playing, as Merleau-Ponty would say. The student becomes a “different other” for herself, as everything becomes mediated through the tonal language.

Finally, the student hears something general *in* the constant variations. She understands how every concrete tonal organization pivots around or exemplifies something general—something that ultimately remains the same across every contingent tonal variation. This selfsameness is the general syntax of the tonal system. However, the general order is inseparable from the possible aural perspective—the *different* ways of hearing the generative potentials. And this is what the student now hears: how there are *infinite ways of listening embedded in the same tonal organization*. She perceives how the generative potentials are universal pools of possible auditory enaction—pools that cannot be “used up” because there will always be yet another way to listen within the syntax.

In other words, the student has understood the tonal syntax as a symbolic and intersubjective order that enables the current form of tonal communication.

7.2 No Building Blocks but Transformation and Differentiation

In this article, we have tried to describe the formation of an aural, embodied, and communicative knowledge, which is nonscriptural by nature. Nothing in the student’s learning process can be fixed conceptually and passed over

in written or semi-written forms. Ultimately, it can only be communicated in music.

To conclude, we can note how our usage-based approach to improvisation differs from a theoretical model that dominates much contemporary research on jazz improvisation and jazz education. Many theorists believe it is a good idea to approach jazz improvisation as if the musical behavior was an activity of synthesizing discrete units. Call this the *building block approach*.

Berliner exemplifies the approach. When he reflects (briefly) on how imitation can prepare for improvisation, he chooses to pick up Nettl's famous metaphor,⁶⁴ describing improvisation as a combination and recombination of building blocks: "Many students begin acquiring an expansive collection of improvisational building blocks by extracting those shapes they perceive as discrete components from the larger solos they have already mastered and practicing them as independent figures."⁶⁵ Wilf follows Nettl's and Berliner's path: "[I]mprovisation involves imitation insofar as it is a recombination of previously available building blocks created by other improvisators."⁶⁶ Philosopher Benson sees no trouble in doing more or less the same: "For improvisation is a sense of 'putting together.' One takes the basic rhythmic and chord structures of the genre in which one works, and puts them together in different ways."⁶⁷ In fact, across the broadest spectrum of theories with interest in improvisational and musical behavior, the same approach is used in various ways.⁶⁸

The building-block approach is perhaps a residue of Hume's empiristic description of human perception,⁶⁹ or it may be a conception taken from the Western literary tradition, where the building blocks are like letters in a sentence. Either way, the leap into tonal fluency that we just tried to illuminate would equal the ability to execute rapid combinations of individual tones or sequences of tones. Our student would now be capable of producing a specific set of notes within the acquired framework of implicit and explicit rules. She would also be able to combine various rhythmic structures in different ways, perhaps even into novel combinations.

From the perspective of a perceiving subject, this way of describing the learning process might seem all right. Working with the exercises described above, the student can very well isolate distinct phrases and rhythmic patterns and "put them together" with other isolated elements in focus, thus having the sense of performing a synthesizing activity. However, from the reflected point of the usage-based approach elaborated in this article, the building-block approach is downright wrong. For, as we have seen, not one of the exercises described encourages a synthesizing activity. Not one of the tasks inspires the construction of a musical language by adding rhythm *plus* harmony *plus* tonal sequences or whatever. Instead, constructing the musical language was *always* about hearing musical *wholes* (or matrices) of rhythm, tonality, harmony, and individual style. From the vantage point of

the ear, the generative potentials make sense only as musical wholes or aural matrices.

Besides, in a fundamental philosophical sense supported by this article, *there is no such thing as discrete parts of music* or building blocks. The least fragment of a rhythmic pattern or tonal sequence presupposes and resounds the entire lifeform in which it originated. We have seen how rhythm organizes the first relationship, how the general ability to follow and accomplish tonal sense originates in the ontogeny of joint attention, and how rhythm and tonality have the longer histories in phylogenetic evolution. Each musical whole belongs to a larger whole, implying culture, family, the local language, and everything else that characterizes life with others. Thus, it is more apt to say that constructing a musical language by aural imitation is learning to discriminate within larger wholes, rather than approaching the process as a synthesis of parts into a whole. The differentiation also has an artistic and expressive dimension. To borrow phrasing from Welsh, “The very idea of ‘being an individual’ can only take place against the relief of a shared social world from which I seek to individuate myself.”⁷⁰ Just as imitation prepares the child for the emancipative transformation within verbal language, the aural imitation method prepares for emancipative transformations from inside the musical sense.

Our imagined student now packs her horn and goes to meet, play, and improvise with her band. The precise musical language is hers—and it unfolds in the collective.

Notes

Thanks to Lars Sigfred Evensen, Ståle Finke, Bengt Molander, Hans Magnus Solli, Lasse Thoresen, Ingebjørg Seip, Roger Jeffs, Michael Duch, Pradeep A. Dhillon, and members of the Research Group for Aesthetics and Phenomenology at NTNU for highly valuable readings and commentaries on various drafts of the manuscript. Thanks to Eldbjørg Raknes and Vigleik Storaas for important background information on aural jazz musicianship. Double thanks to Njål Ølnes for commenting on an early draft of the manuscript and being a well-informed and inspiring discussion partner in the writing process. Thanks to the Faculty of Humanities at NTNU for financing the project.

1. Bob Mintzer, *12 Contemporary Jazz Etudes: B-Flat Tenor Saxophone* (Van Nuys, CA: Alfred Publishing, 2004); Bob Mintzer, *Playing Jazz Piano* (Van Nuys, CA: Alfred Publishing, 2004).
2. [Hal Galper,] “Hal Galper’s Master Class—*The Illusion of an Instrument*,” May 2, 2010, https://www.youtube.com/watch?v=y_7DgCrziI82010, accessed Aug. 9, 2021.
3. Maurice Merleau-Ponty, *Phenomenology of Perception*, ed. Donald Landes (London: Routledge, 2012), 144; Alva Noë, *Action in Perception* (Cambridge, MA: The MIT Press, 2006), 1.
4. Hal Galper, “Able Bodied—An Interview with Hal Galper,” *JazzImprov Magazine* 3, no. 3 (2001): 3, Hal Galper, <https://www.halgasper.com/interviews-2/jazzimprov-magazine-interview/>, accessed Aug. 9, 2021.

5. The history of the Western tonal system is neatly interwoven with the Western literary tradition, as demonstrated by Richard Taruskin, *The Oxford History of Western Music*, vol. 1 (Oxford: Oxford University Press, 2005). However, in accordance with what was said in Section 1.4, Part 1, we set aside this perspective, approaching the musical system merely as an audible phenomenon.
6. Roger Scruton, *The Aesthetics of Music* (Oxford: Clarendon Press, 1997) 240.
7. Scruton, *The Aesthetics of Music*, 248.
8. James Barbour, *Tuning and Temperament* (New York: Da Capo Press, 1972); Ross W. Duffin, *How Equal Temperament Ruined Harmony (and Why You Should Care)* (New York: W.W. Norton, 2007).
9. Scruton, *The Aesthetics of Music*, 248ff.
10. Scruton, *The Aesthetics of Music*, 250.
11. Martine Van Puyvelde et al., "Tonal Synchrony in Mother–Infant Interaction Based on Harmonic and Pentatonic Series," *Infant Behavior and Development* 33 (2010): 387–400, <https://doi.org/10.1016/j.infbeh.2010.04.003>, accessed Aug. 9, 2021; Beatriz Ilari, "Music and Babies: A Review of Research with Implications for Music Educators," *Update: Applications of Research in Music Education* 21 (2002): 17–26; Stanley N. Graven and Joy V. Browne, "Auditory Development in the Fetus and Infant," *Newborn and Infant Nursing Reviews* 8 (2008): 187–93, <https://doi.org/10.1053/j.nainr.2008.10.010>, accessed Aug. 9, 2021; F. Hicks, "Theatre Nursing: The Power of Music," *Nursing Times* 88 (1992): 72–74.
12. Jon Roar Bjorkvold, *The Muse Within: Creativity and Communication, Song and Play from Childhood through Maturity* (New York: HarperCollins, 1992).
13. Paul Franklin Berliner, *Thinking in Jazz* (Chicago: University of Chicago Press, 1994); David Schroeder, *From the Minds of Jazz Musicians: Conversations with the Creative and Inspired* (New York: Routledge, 2018); Eitan Y. Wilf, *School for Cool: The Academic Jazz Program and the Paradox of Institutionalized Creativity* (Chicago: University of Chicago Press, 2014).
14. Louis Cavrell, "The Universal Mind of Bill Evans," TV documentary, 1966, <https://www.youtube.com/watch?v=QwXAqlaUahI1966>, accessed Aug. 9, 2021.
15. Berliner, *Thinking in Jazz*, 94.
16. Daniel N. Stern, *The Interpersonal World of the Infant: A View from Psychoanalysis and Developmental Psychology* (London: Karnac, 1998), 162ff.
17. Stern, *The Interpersonal World*, 162.
18. Mark Reybrouck, "Musical Creativity between Symbolic Modelling and Perceptual Constraints: The Role of Adaptive Behaviour and Epistemic Autonomy," in *Musical Creativity: Multidisciplinary Research in Theory and Practice*, ed. Irène Deliège and Geraint Wiggins (Hove, Sussex, UK: Psychology Press, 2006), 58–76; Reybrouck, "Music as Environment: An Ecological and Biosemiotic Approach," *Behavioral Sciences* 5, no. 1 (2015): 1–26.
19. Terrence William Deacon, *The Symbolic Species: The Co-Evolution of Language and Brain* (London: Norton, 1997); Hans-Georg Gadamer, *Truth and Method*, trans. Joel Weinsheimer and D. G. Marshall (London: Continuum, 2004); Maurice Merleau-Ponty, *The Structure of Behavior*, trans. A. L. Fisher (Pittsburgh, PA: Duquesne University Press, 2011); Daniel N. Stern, *Forms of Vitality* (Oxford: Oxford University Press, 2010); Michael Tomasello, *Constructing a Language: A Usage-Based Theory of Language* (Cambridge, MA: Harvard University Press, 2003); Tomasello, *Origins of Human Communication* (Cambridge, MA: MIT Press, 2010).
20. Evan Thompson, *Mind in Life* (Cambridge, MA: The Belknap Press of Harvard University, 2007), 76; Italics original.
21. While Reybrouck's computational framework is largely distinct from ours, he too describes music in terms of indication; see Mark Reybrouck, "Music as Environment," and Reybrouck, "Music Cognition and Real-Time Listening: Denotation, Cue Abstraction, Route Description and Cognitive Maps," *Musicae Scientiae* 14 (2010): 187–202.
22. Thompson, *Mind in Life*, 76.

23. Hal Galper, "The Oral Tradition," Hal Galper, <http://www.halgasper.com/articles/the-oral-tradition/2012/>, accessed Aug. 9, 2021.
24. Ben Sidran, *Black Talk: How the Music of Black America Created a Radical Alternative to the Values of Western Literary Tradition* (New York: Payback Press, 1981), xiv.
25. Galper, *The Oral Tradition*; Mellonee V. Burnim and Portia K. Maultsby, eds., *African American Music: An Introduction* (New York: Routledge, 2014); Simha Arom, *African Polyphony and Polyrhythm: Musical Structure and Methodology* (Cambridge: Cambridge University Press, 1991).
26. Tomasello, *Constructing a Language*, 3–4.
27. Merleau-Ponty, *Phenomenology of Perception*; Charles Keil, "Participatory Discrepancies and the Power of Music," *Cultural Anthropology* 2, no. 3 (1987): 275–83; Tiger Roholt, *Groove: A Phenomenology of Rhythmic Nuance* (New York: Bloomsbury Publishing, 2014).
28. For compatible descriptions of polyphonic openness in real-time aural music-making, see Berliner, *Thinking in Jazz*; Ingrid Monson, *Saying Something: Jazz Improvisation and Interaction* (Chicago: University of Chicago Press, 1996); LeRoi Jones, *Blues People: Negro Music in White America* (New York: Harper Perennial, 2002); George Lipsitz, "Improvised Listening: Opening Statements Listening to the Lambs," in *The Improvisation Studies Reader: Spontaneous Acts*, ed. Ajay Heble and Rebecca Caines (New York: Routledge, 2014), 27–34.
29. Frederick Seddon, "Modes of Communication During Jazz Improvisation," *British Journal of Music Education* 22 (2005): 47–61, <https://doi.org/10.1017/S0265051704005984>, accessed Aug. 9, 2021. Frederick Seddon and Michele Biasutti, "A Comparison of Modes of Communication between Members of a String Quartet and a Jazz Sextet," *Psychology of Music* 37, no. 4 (2009): 395–415, <https://doi.org/10.1177/0305735608100375>, accessed Aug. 9, 2021. Jessica Phillips-Silver and Peter Keller, "Searching for Roots of Entrainment and Joint Action in Early Musical Interactions," *Frontiers in Human Neuroscience* 2, no. 26 (2012), <https://doi.org/10.3389/fnhum.2012.00026>, accessed Aug. 9, 2021.
30. Andrea Schiavio and Simon Höfding, "Playing Together without Communicating? A Pre-Reflective and Enactive Account of Joint Musical Performance," *Musicae Scientiae* 19, no. 4 (2015): 366–88, <https://doi.org/10.1177/1029864915593333>, accessed Aug. 9, 2021.
31. Scruton, *The Aesthetics of Music*.
32. Colwyn Trevarthen, "Musicality and the Intrinsic Motive Pulse: Evidence from Human Psychobiology and Infant Communication," *Musicae Scientiae* 3, no. 1 suppl. (1999): 155–215; Trevarthen, "First Things First: Infants Make Good Use of the Sympathetic Rhythm of Imitation, without Reason or Language," *Journal of Child Psychotherapy* 31, no. 1 (2005): 91–113, <https://doi.org/10.1080/00754170500079651>, accessed Aug. 9, 2021.
33. Daniel N. Stern, "Face-to-Face Play: Its Temporal Structure as Predictor of Socio-affective Development," in *Rhythms of Dialogue in Infancy: Coordinated Timing in Development*, ed. Joseph Jaffe et al. (Boston: Wiley/Society for Research in Child Development, 2001), 144–49, 147; Stern, *The First Relationship* (Cambridge, MA: Harvard University Press, 2002); Stern, *The Interpersonal World of the Infant*.
34. Stern, *The Interpersonal World of the Infant*, 97.
35. Stern, *The Interpersonal World of the Infant*, 110.
36. Stern, *The Interpersonal World of the Infant*, 110.
37. Colwyn Trevarthen and Penelope Hubley, "Secondary Intersubjectivity: Confidence, Confiding and Acts of Meaning in the First Year," in *Action, Gesture and Symbol: The Emergence of Language*, ed. A. Lock (London: Academic Press, 1978), 183–229; Penelope Hubley and Colwyn Trevarthen, "Sharing a Task in Infancy," *New Directions for Child and Adolescent Development* 1979 (1979): 57–80; Beebe et al., "A Comparison of Meltzoff, Trevarthen, and Stern."
38. Colwyn Trevarthen, "What Is It Like to Be a Person Who Knows Nothing? Defining the Active Intersubjective Mind of a Newborn Human Being," *Infant & Child*

- Development* 20 (2010): 119–35, 119, <https://doi.org/10.1002/icd.689>, accessed Aug. 11, 2021.
39. Stern, *The Interpersonal World of the Infant*, 26–34 and 124–61.
 40. Stern, *The Interpersonal World of the Infant*, 27.
 41. Stern, *The Interpersonal World of the Infant*, 27
 42. Stern, 129.
 43. Maurice Merleau-Ponty, *Child Psychology and Pedagogy: The Sorbonne Lectures, 1949–1952*, trans. Talia Welsh (Evanston, IL: Northwestern University Press, 2010), 22.
 44. Andrea Schiavio, Dylan van der Schyff, Silke Kruse-Weber, and Renee Timmers, “When the Sound Becomes the Goal: 4e Cognition and Teleomusicality in Early Infancy,” *Frontiers in Psychology* 8, no. 1585 (Sept. 25, 2017), <https://doi.org/10.3389/fpsyg.2017.01585>, accessed Aug. 9, 2021.
 45. Schiavio et al., 6.
 46. Stern, *The Interpersonal World of the Infant*, 129.
 47. Phrase borrowed from Tomasello, *Constructing a Language*, 3.
 48. Stern, *The Interpersonal World of the Infant*, 162ff.
 49. Stern, *The Interpersonal World of the Infant*, 133.
 50. Stern, 161.
 51. Tomasello, *Constructing a Language*, 8–17.
 52. Maurice Merleau-Ponty, *Signs*, trans. Richard McCleary (Evanston, IL: Northwestern University Press, 1987), 40.
 53. Seddon, “Modes of Communication during Jazz Improvisation.”
 54. Phillips-Silver and Keller, “Searching for Roots.”
 55. Merleau-Ponty, *The Structure of Behavior*, 96–97. Merleau-Ponty cites Kurt Koffka, *The Growth of the Mind* (New York: Brace, 1925).
 56. Merleau-Ponty, *The Structure of Behavior*, 96–97.
 57. Trevarthen, “Musicality and the Intrinsic Motive Pulse,” 158.
 58. Stern, *The Interpersonal World of the Infant*, 124–37.
 59. Gadamer, *Truth and Method*, 296. Italics original.
 60. Gadamer, *Truth and Method*, 355.
 61. Merleau-Ponty, *Signs*, 97.
 62. Stein Bråten, “Participant Perception of Others’ Acts: Virtual Otherness in Infants and Adults,” *Culture and Psychology* 9 (2003): 261–76.
 63. Jessica Benjamin, “Beyond Doer and Done: An Intersubjective View of Thirdness,” *The Psychoanalytic Quarterly* 73, no. 1 (2004): 5–46. See also Part 1, Section 3, of the current article.
 64. Bruno Netti, “Thoughts on Improvisation: A Comparative Approach,” *The Musical Quarterly* 60, no. 1 (1974), 1–19, 13.
 65. Berliner, *Thinking in Jazz*, 101.
 66. Wilf, *School for Cool*, 134.
 67. Bruce Ellis Benson, *The Improvisation of Musical Dialogue: A Phenomenology of Music* (Cambridge: Cambridge University Press, 2003), 136.
 68. Aaron L. Berkowitz, *The Improvising Mind: Cognition and Creativity in the Musical Moment* (Oxford: Oxford University Press, 2010); Robert Gjerdingen, *Music in the Galant Style* (Oxford: Oxford University Press, 2007); Alfred Pike, “A Phenomenology of Jazz,” *Journal of Jazz Studies* 50 (1974): 88–94; Jeff Pressing, “Improvisation: Methods and Models,” in *Generative Processes in Music: The Psychology of Performance, Improvisation, and Composition*, ed. John Sloboda (Oxford: Oxford University Press, 1988), 129–78; Reybrouck, “Music as Environment.”
 69. David Hume, *Enquiries concerning Human Understanding and concerning the Principles of Morals* (Oxford: Clarendon Press, 1975).
 70. Talia Welsh, *The Child as Natural Phenomenologist: Primal and Primary Experience in Merleau-Ponty’s Psychology* (Evanston, IL: Northwestern University Press, 2013), 49.