



SARAH BRO TRASMUNDI, MATTHEW ISAAC HARVEY
University of Southern Denmark, Odense, Denmark

A BLENDED QUANTITATIVE-ETHNOGRAPHIC METHOD FOR DESCRIBING VOCAL SONIFICATION IN DANCE COACHING

In this paper we present a micro-analytic description of the role vocalizing plays in a single case of professional dance instruction. We use a novel mix of qualitative and quantitative tools in order to investigate, and more thoroughly characterize, various forms of vocal co-organization. These forms involve a choreographer using vocalization to couple acoustic dynamics to the dynamics of their bodily movements, while demonstrating a dance routine, in order to enable watching dancers to coordinate the intrabodily dynamics of their own simultaneous performances. In addition to this descriptive project, the paper also suggests *how* such forms of coordination might emerge, by identifying those forms of voice-body coupling as potential instances of “instructional vocal sonification”. We offer a tentative theoretical model of how vocal sonification might operate when it is used in the teaching of movement skills, and in the choreographic teaching of dance in particular. While non-vocal sonification (both physical and computer-generated) is increasingly well-studied as a means of regulating coordinated inter-bodily movement, we know of no previous work that has systematically approached vocal sonification. We attempt to lay groundwork for future research by showing how our model of instructional vocal sonification might plausibly account for some of the effects of vocalization that we observe here. By doing so, the paper both provides a solid basis for hypothesis generation about a novel class of phenomena (i.e., vocal sonification), and contributes to bridging the methodological gap between isolated descriptions and statistical occurrences of a given type of event.

Key words: vocal sonification, dance choreography, cognitive ethnography, interaction dynamics, interpersonal coordination

1. Introduction: Setting the Scene

1.1. Conceptual Background: Vocal Sonification in the Context of Movement Instruction

This paper makes a methodological contribution to multimodal interaction studies (Goodwin, 1994; 2002; 2003; 2007, Streeck, 2009; Streeck et al., 2011) by adding a straightforward quantitative component to an existing ethnographic analysis (Trasmundi, forthcoming). Specifically, this section defines a narrow but frequently-recurring class of phenomena: the use of vocal sonification in the teaching of complex movement skills (henceforth “instructional vocal sonification”). In broad strokes, such an activity consists of a coach, teacher, or choreographer using rhythmic vocalizing to guide their students during practice or performance of high-skill movement activities such as dance and martial arts. The current section gives a theoretical and conceptual overview of vocal sonification as part of movement instruction, and concludes that this class of phenomena becomes much more intelligible – more amenable to clear description and potentially to hypothesis-generation as well – when approached using a combination of quantitative and qualitative tools.

Section 2 illustrates this point by exploring an example of instructional vocal sonification in some detail, providing an in-principle demonstration that thick qualitative description (Geertz, 1973) is well-suited to describing teaching strategies and identifying pivotal teaching events, but that complementary quantitative tools are required to adequately describe the role of sonification involved. While Trasmundi (forthcoming) offers a “thick” qualitative description (Geertz, 1973) of the wide variety of teaching techniques that the choreographer employs, analyses of that type are unable to draw generalizations about the specific contributions made by each technique. Specifically, in that article Trasmundi identifies three interrelated instructional activities: *pinpointing*, *perfection*, and *demonstration*. These are, respectively, the main instructional foci for cycles 1, 2, and 3 of the dance routine. One aim of this paper is to provide detailed descriptions of the temporal relations within and between these instructional cycles – as these relations might support those prior theoretical divisions – in order to show how instructional sonification in whole-bodied interaction can be used to emphasize small-scale attentional cues for skilled dancers.

By contrast to qualitative descriptions, a rich quantitative literature documents the positive effects on performance of using live auditory feedback during the practice of movement skills in a variety of athletic activities. However, this literature is restricted to non-vocal sonification, in which sounds have been synthesized using purpose-built software (see e.g., Effenberg, Fehse, Schmitz, Krueger, & Mechling 2016; Schaffert & Mettes 2015). This paper is thus a very first attempt at describing vocal sonification

as an instructional technique in the teaching and practicing of complex movement skills, achieved by bringing qualitative interaction analysis together with a minimal quantitative analysis. We argue that much can be gained from linking qualitative case-studies with quantitative analysis, and this paper proposes some of the groundwork required to systematically compare examples of vocal sonification in different contexts. In other words, this paper should be treated as a qualitative case-study that links thick descriptions with quantitative measurements of end- and start-times of movements in relation to vocalizations. By doing so, it contributes to (i) bridging the gap between isolated descriptions of an event and statistical occurrences of an event, and (ii) providing a solid basis for hypotheses generation about a particular phenomenon, in this case the role of sonification in instructional practices. As such this paper provides insights into *what* would be important variables to measure to explain the role of sonification in instructional practices. Thus, it can also be used as a stepping stone for research that aims at providing general explanations of how sonification works and how it adds value to instructional practices.

Our conceptual explanation begins with “sonification”. The most general and conservative definition is provided by the National Science Foundation’s ‘Sonification Report’: “sonification is the use of nonspeech audio to convey information, [or] more specifically, sonification is the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation” (Kramer, et al., 2010, p. 4). Several parts of this definition – the phrases “convey information” and “for the purposes of facilitating communication or interpretation” – are at odds with an ecological-distributed model of interaction, as they imply a model on which interaction involves the transfer of information between two discrete and isolated information-processors. The core of the definition does not depend on these phrases, however, as we will explain. That core is the idea that in its most minimal sense, sonification involves a systematic mapping from some target domain into acoustic structure. Crucially, that systematic mapping must be produced by synthesizing, articulating, or emitting sound. Thus, sonification is not *only* a structural relationship. It is also an active process, both in the sense that sonification must be achieved and in the sense that its achievement enables a perceiving person to carry out previously impossible actions in the target domain.

In some cases, sonification is just a matter of physical laws, as when kinetic events sonify themselves by producing perceptible sound (think of the hum of a car passing by or the rush of a glass filling under the tap). But most of the time, and in all cases where sonification is employed as part of human interaction, sonification is a *deliberate* process of data collection and sound synthesis. Human methods for sonifying data vary widely, from using datasets as input to synthesizers with the minimum-possible amount

of processing to elaborate parameter mappings between domains or the construction of complex acoustic models (see Dubus & Bresin, 2013, for overview). In all such cases, though, sonification itself is embedded within various forms of interaction in order to modify them by enabling new forms of sensorimotor coordination. The definition given by Kramer and colleagues (2010) is ambiguous on this score. We can be more precise by noting that sonification itself is simply the mapping of data relations in a non-acoustic domain into acoustic structure. Various methods or types of sonification – whether natural or digital – can be embedded in human interaction in order to modify human activity, and it is these second and third, logically distinct steps that the authors refer to with the phrases “convey information” and “for the purposes of facilitating communication or interpretation”. Thus, sonification itself is logically a three-place relation between (i) one or more measurable parameters or identifiable patterns in a non-acoustic domain, (ii) a sound-generating or -synthesizing process, and (iii) one or more measurable parameters in the acoustic domain. In ecological-distributed terms, we can then add that Kramer and colleagues (2010) are specifically interested in instances of sonification, that, *unlike* the example of a glass filling under a tap, are undertaken deliberately by one person in order to affect the activity of another. These instances thus also always involve (iv) a human audience who can perceive an isomorphism between the acoustic parameter(s) produced and the non-acoustic parameter or pattern being sonified, and (v) some change in how the perceiver can or does interact with the non-acoustic parameter or process.

Our interest here is in the *vocal* sonification of human movement for instructional purposes. That is to say, we are concerned with a variety of deliberate sonification where the sound is modulated by a human voice in order to enable their listener(s) to modify the dynamics of their own ongoing or anticipated movements in specific ways. For instance, change in the pitch of a vocalization might sonify spatial displacement or acceleration along a particular vector, thus enabling the listener to match the magnitude or acceleration of their own movement to that of the movement being sonified. Cases of sonification that fulfill these criteria are common in daily life but almost entirely absent from the scholarly literature. In fact, we are aware of only one piece of formal research on instructional vocal sonification: Fagergren’s (2012) analysis of the teaching practices of professional choreographer Wayne McGregor (henceforth “WM”).

Fagergren’s work makes use of video data collected by David Kirsh and his colleagues at the University of California San Diego (see Kirsh, 2012), and because the rest of this paper will be spent discussing a clip from one of those videos, it is necessary now to clarify our views of the data, and the types of claims it licenses. These are that neither we nor Fagergren (2012) have any quantitative basis at present for asserting (or denying) that WM *in fact*

sonifies anything at all with his nonverbal vocalizations. This is both because the available data – the video recordings – are simply not of sufficient quality to afford quantification of the magnitudes of the dancers' movements, and because there has been no systematic effort to identify the kinetic parameters that define any of the movements involved in any of the distinct and coherent forms of sensorimotor coordination the company's dancers perform. In other words, it is possible neither to determine what movement parameters WM is attempting to sonify in any given instance, nor to measure those parameters if they could be determined.

Our proposals are made within an ecological framework whose primary assumption is that action and interaction are constituted by sensorimotor coordination, on multiple spatial and temporal scales, between organisms and their environments (Hodges 2007; Järvillehto 2009; Steffensen & Pedersen 2014). This means that we understand distinct forms of action as distinct forms of sensorimotor coordination, uniquely characterized by certain features of their coordination dynamics (Bingham, 1988; Buhrmann, Di Paolo, & Barandiaran, 2013; Turvey, 2007; Warren, 2006; Wilson & Golonka, 2013). By focusing on a single, perceptually-clear phenomenon, the present paper ignores the vast majority of the interpersonal coordination that takes place in our chosen example of interaction, including most of the ways in which vocalizing might plausibly be involved. Our focus throughout will primarily be on non-verbal vocal activity on the part of the instructor, and the ways in which it could *plausibly* function to specify, regulate, or provide feedback on the relative timing of the movements they are practicing. Thus, even though the final section of the paper situates our work with respect to the language sciences more generally, almost the entirety of what follows concerns the elucidation of one example of vocal sonification in the context of teaching movement skills. Our hope is that this elucidation, as well as the methods employed, will pave the way for more rigorous, comparative work on vocal sonification.

Thus, this paper is not a scientific report of a traditional empirical study. Rather, it concerns the issue of how one can effectively observe a new and complicated phenomenon, and how observation and reports of experience can be used to decide what exactly needs to be measured and formally documented. The article aims to shed light on a poorly-understood class of phenomena. In addition, section 2.2 argues that WM very likely *does* sometimes succeed in vocally sonifying movement parameters, and in so doing provides a conceptual frame for the empirical work presented in section 2.

1.2. Choreographic Practice and Sonifying “Feelings of Moving”

Despite lacking the tools or information necessary to determine whether WM's vocalizations are genuine sonification, Fagergren (2012) succeeds in establishing several convincing claims about how those vocalizations fit in to

his choreographic instruction. The first is that regardless of whether he succeeds, WM certainly *intends* to sonify dance movements as part of his instructional practice, and very often attempts to do so. Fagergren reports finding 90 unambiguous episodes of WM's nonverbal vocalizing in approximately 10 hours of video (Fagergren, 2012, p. 7). The second is that these vocalizations are intended to evoke, for a listener, what it feels like to experience some particular bodily movement. WM describes this variously as a "feeling of moving" or a "feeling of sound" (Fagergren, 2012, p. 34). The third claim, also derived from WM's self-reports, is that the sensations thus evoked are always intended to guide or constrain the listener's ongoing movements.

These claims amount to saying that WM's vocalizing is an effort to actively regulate certain parameters of his dancers' ongoing movements. This seems likely to be a core property of instructional vocal sonification, just because it is definitionally a teaching strategy and therefore is intended by the teacher to be a means of changing behavior. The immediate question, then, is whether this is possible. Can an acoustically-evoked "feeling of moving" constrain the dynamics of actual, ongoing movements? In interviews (2012, pp. 34–35) WM states that his vocalizing is intended to elicit particular kinesthetic sensations, so that these sensations will then change how the dancers perform a new movement. In one instance, he wants an arm movement whose dynamics and kinesthetics produce feelings that he describes as "elasticity", and as being "high" rather than "low", and each of these descriptions accompanies a nonverbal sound. When asked whether he will remember the movement in terms of the sound he produced to elicit those feelings of elasticity and height, he replies, "well, only if I see that they are not doing that [i.e., not doing the movement correctly]. I can see the difference between this [the correct version] and this [the incorrect version]" (Fagergren, 2012, p. 35). While the language here is vague, and self-report might be a dubious source of insight into coordination dynamics, WM suggests that his nonverbal utterances are direct regulatory responses to perceiving incorrectly-performed movements.

Crucially, there are independent considerations that suggest those responses may be both well-motivated and effective. Consider the complexity of the movements involved in WM's choreography: he is working with professionals who have the highest levels of expertise, fluent in a vast array of standard gestures and subtle modifications to them. When he teaches his dancers a new phrase, he will quickly reach a stage where every dancer in the group can perform the movements, but do so with a myriad of tiny deviations from his exact desires. Imagine, for instance, that the angle of one dancer's forearm relative to the angle of her neck is just slightly off during one quarter of a turn, and at the same time her extended foot is a few degrees too tightly arched, and her gradual rise up from crouching to standing takes place over 4.5 seconds rather than 4.2. In other words, it is at least plausible that

the instructional corrections and interventions WM must make are often too complex – too numerous, too subtle, and too interrelated – to be directly pointed out with verbal description. If this is the case, then WM's strategy of evoking particular "feelings of moving" – elasticity, stickiness, and so on – may be the most direct, most explicit means available to him to specify his desired changes to the dancers' current movements. This view gains support from the writings of Ivar Hagendoorn, another professional choreographer.

[W]hat is composed in a choreography is at once the material – movement – and the sensation it entails, first and foremost a sensation of movement, but extending to other feelings, events and contingencies. [...] The art of dancing is as much 'moving' as it is knowing the effect a movement sequence has on the observer. It is knowing where to put an accent, which phrase to emphasize, when to accelerate or when to release. Most of this knowledge is implicit [...] It encompasses the principles of perception and motor control and what is summed up by experience, the product of years of training... (Hagendoorn, 2004, pp. 104–105.)

The interesting claim here is not that Hagendoorn feels "sensations of movement" are intrinsic to the artistic practice of dance and choreography, but rather that he understands expertise in these domains as the possession of implicit knowledge of how those feelings are realized by nitty-gritty details of the dynamics of sensorimotor coordination. Compare Hagendoorn's very general statement to a more specific one made by WM.

I wanted this energy of high and low but [...] it wasn't just about a spatial organization for me it was about a feeling of moving and it was from the inside out rather than the outside in. [...] I felt that sound wise I needed [...] a layer of information [...] that wasn't about words [and] wasn't about interpretation of words, I could sing them a rhythm that would help them express the physicality so I could express what the transition sounded like to give them a kind of an embodied image of it. (Wayne McGregor, as reported in Fagergren (2012, p. 34), lightly edited for repetition.)

In summary, these statements imply that dance instruction has to do with the ability to express "feelings [or "sensations"] of moving" as small changes in the dynamics of one's own ongoing sensorimotor activity, and reciprocally to perceive subtle, complex patterns in the sensorimotor activity of others *in terms of* those "feelings of moving".

That model of expertise in choreography and dance is one of the reasons to think that WM's instructional strategy – using nonverbal vocalizations to evoke or elicit very particular "feelings of moving" – may be effective. The other reason is the growing body of evidence suggesting that

technologically-achieved, synthetic sonification can both increase the efficacy of training for movement skills and improve expert performance in real time (Dubus & Bresin, 2013; Dyer, Stapleton, & Rodger, 2017; Pizzera & Hohnmann, 2015; Schaffert, Mattes, Barrass, & Effenberg, 2009; Sigrist, Rauter, Reiner, & Wolf, 2013; Sors, Murgia, Santoro, & Agostini, 2015). Moreover, sonification is effective in improving performance because it “educates the attention” in at least three ways. It can (i) map a very high-dimensional variable onto a low one in real time, allowing athletes or performers to directly regulate their activity with respect to previously-unavailable information. Examples include sonifying the pressure a speed-skater exerts on the ice with multiple different parts of his skate blade (Stienstra, Overbeeke, & Wensveen, 2011), thereby enabling him to adjust the angle and consistency of his strokes. Further, the sonification of a continual, accurate measure of boat velocity lets rowers adjust the relative lengths of the various parts of the rowing cycle, increasing the time spent on propulsion and decreasing the time spent on recovery (Schaffert & Mattes, 2015; see Dyer, 2017 for discussion). Sonification can also (ii) provide “models” or “targets” for movement dynamics in real time. For example, in a case study where increased acoustic intensity seemed to elicit increased application of force by weight lifters (Murgia, et al., 2012), or in experiments where expertise in hurdling and skateboarding was found to be “guided by auditory events that trigger proper anticipations of the corresponding patterns of movements” (Cesari, Camponogara, Papetti, Rocchesso, & Fontana, 2014 p. 1; Kennel, et al., 2015). Finally, sonification can (iii) be used to direct a watcher’s attention to complex parameters of a movement, thereby increasing both the accuracy of their subsequent perceptual judgments and their own performance on the same task (e.g., Effenberg, 2004; Ramezanzade, Abdoli, Farsi, & Sanjari, 2014).

Taken together, these claims are coherent and complementary. On the one hand, WM and Hagendoorn’s account of choreographic expertise needs a story about how nonverbal vocalization can elicit particular “feelings of moving”. The literature on artificial sonification provides such story, provided the vocalizations are at least semi-systematically isomorphic with the relevant kinematic parameters. On the other, the literature on artificial sonification needs a model of choreographic instruction as a well-defined mode of sensorimotor coordination, with characteristic parameter ranges and behavior.

One of this article’s goals is to offer hypotheses for future exploration based on the foregoing argument and the careful description of the phenomenon of instructional vocal sonification that we undertake in section 2. Thus far, we are in a position to speculate that Wayne McGregor is engaged in genuine sonification, as that term was defined at the beginning of this section, and that his sonifications will necessarily be less precise than artificial sonification is, and probably involve a mix of the three varieties just

enumerated. Also, rather than a straightforward mapping from one information structure (e.g., kinetic events) to another (e.g., acoustic dynamics), his instructional vocal sonification most likely involves complex mappings at least from articulatory dynamics to acoustic dynamics (that is, from vocalizing to sound), and from acoustic dynamics to kinesthetic and other physiological parameters that together comprise “feelings of moving”.

In order to develop these conjectures further, and to give a clearer picture of the phenomenon of instructional vocal sonification, section 2 provides a qualitative description of one instance of WM’s use of nonverbal vocalizations as a choreographic tool. It focuses most closely on ways in which his vocalizations might be sonifications, and now these sonifications might plausibly be enabling his dancers to regulate their own movements. It does this by describing both his teaching strategies and apparent goals, and documenting co-occurrences of his vocalizations and his bodily movements. Our hope is that the kinds of ethnographic and multimodal based descriptions (Goodwin, 2013; Kirsh, 2010; Pedersen, 2015) we are able to provide here will both give a clear sense of what instructional vocal sonification is, in practical terms, and by so doing open the way to systematic study of the phenomenon in the future.

1.3. A Blended Quantitative-Ethnographic Method: Presenting the Case-Study of Instructional Vocal Sonification

The dataset used in this article derives from a large cognitive, ethnographic study conducted by David Kirsh and his team at the University of California, San Diego. It was carried out in collaboration with the choreographer Wayne McGregor and his *Random Dance Company* (Kirsh, 2012). The study was conducted in London, and the data consists of a large set of video-recordings of dance rehearsals and interviews with company members. However, while the entire database contains dozens of hours of video, the analysis here restricts itself to a 25-second video clip that captures a short teaching sequence. The example is of interest for several reasons. Its basic structure is cyclical, with WM working his way through the same dance phrase three times in a row. As he does, he gives more careful treatment to one step in particular (the movement of the left leg from bent to straight). For instance, he fixes joint attention on it in various ways, and otherwise emphasizes its performance throughout the teaching event. Additionally, the clip involves a striking pattern of highly coordinated vocal activity by WM, which is highly rhythmic, clearly central to the organization of his teaching activities, and clearly articulated (these vocalizations are described and visualized below). For these reasons, the pattern persuasively invites exploration of its potential status as a sonification, potentially of one or more parameters of WM’s own concurrent dance performance.

In order to explore these features of the video data, we describe WM's instructional uses of vocalization in terms of coordinated movements, in order to investigate whether that vocalization might plausibly be an instance of sonification – and if so, what it sonifies, and how it does so. Our initial expectations were that WM's vocalizations would be highly coordinated with gestures of other modalities, and in particular that they would display a relatively high degree of synchrony, as indicated by the degree to which vocalizations and identifiable gestures co-occurred or “overlapped” with one another. From the ethnographic notes and a large archive of video data and interviews, it is evident that the very skilled random dance company rely on a high degree of micro-scale, multimodal adaptive behavior in order to achieve precision and synchrony in their dance expressions. It is important to recognize the exploratory nature of this description: co-occurrence alone is not evidence of synchrony, that is, of coordination by means of a common constraint. We know WM's understanding of the dance routine includes “feelings of moving”, and we are speculating that these feelings constrain both his performance of the dance and his accompanying vocalizations. Here, we do no more than argue that this assumption is plausible, and that the co-occurrence we observe is the result of true coordination; later testing will establish the truth of the matter.

The description below proceeds in two parts. The first is a cognitive-ethnographic examination of WM's evident teaching goals and the strategies he uses to pursue them, as well as of the overall sequence of events (the cyclical structure, identifying the components of the dance phrase being rehearsed, etc.). This process was based on the analysis already completed for Trasmundi (forthcoming), and was primarily concerned with documenting WM's vocalizations in greater detail. The second component of the description is quantitative, and consists of annotating the start- and end-times of each identifiable movement and vocalization performed by WM. That process is described further here.

This exclusive focus on WM, as opposed to any of the dancers, was motivated by several considerations. First, he is the only person in the clip who fully performs the routine three times. Second, he is both the instructor and the only person who engages in any significant vocalization, and as such has a monopoly on actions of special interest to us, and finally, WM's behavior never appears to be especially responsive to that of any of the dancers, with one notable exception described in section 2.1. The process was simplified by limiting the annotation to four somewhat artificial layers or “modalities”: vocal utterances, arm/hand gestures, leg/foot gestures, and head orientation (intended to be a proxy for gaze, and omitted here as lacking sufficient detail to add any value). The bodily gestures were identified in ELAN, and differentiated by sudden changes of direction or velocity, and by periods of stillness. ELAN analysis proceeded frame-by-frame, noting start- and

end-times as well as overlaps and un-categorizable transitions. Additionally, vocalizations were identified in Praat, using the sound file, while gestures were identified in ELAN without sound. Segments were then identified using the compiled list of gestures. Vocalization start- and end-times were identified in Praat using surface-level features such as acoustic intensity and formant values.¹ The resulting lists of vocalizations and bodily movements were used to document the relative timings of voice, hand, and foot movements.

That process produced a list of some 147 observed “gestures” and vocalizations whose start-times and end-times are accurate to within either 15 ms (for bodily movements) or 2 ms (for vocalizations). Individual gestures included, for example, a step with one foot (from initiation to termination of movement), raising or lowering a hand, bending a knee, and so on. Gestures were differentiated in the grounds of either (a) stillness between gestures or (b) sharp changes of direction of the limb in question.

This list made it possible to analyze the temporal structure of WM’s activity in relation to the segmental and cyclical nature of the rehearsal, noteworthy contingencies, and other patterns. Intuitively, WM appears perform to a single dance phrase, or dance “routine”, three times over the course of the video clip. These three repetitions are identified below as “cycles” 1, 2, and 3. Our first task was to identify the routine’s component movements, or “segments”. These segments were general movement features that WM repeated from one cycle to the next, and to which he gave special attention. These were also the movements that his dancers repeated in their own movements. Throughout, the terms “segment” and “routine” will be used to indicate abstract units of choreography, while “gesture”, “vocalization”, and “performance” will indicate specific observed movements. The Macarena, for instance, is a dance “routine”, involving “segments” that include putting your left hand in, putting your right hand in, and so on, while the movement by means of which some particular person performs one particular step of the dance on some particular occasion would be a “gesture”.

In this instance WM is teaching a six-segment routine, which he performs three times; as mentioned, this repetition (and corresponding repetition by his dancers) is the means by which the segments were identified. During these performances, he makes many gestures with his hands, feet, and torso, as well as vocalizations. Some of those gestures are directly involved in performing segments of the routine – such as the lowering of his left leg that lasts from 13.18 and 14.28 seconds in the video timeline – while others are not. For example, between 12.28 and 12.36 seconds, WM is holding his arms in a circle in front of him, with his hands meeting over his lower stomach. During that precise interval, his left hand rises slightly above his

¹ There are only two instances of speech by people other than WM, both of them brief and quiet. Those vocalizations were excluded from the analysis, along with very brief “transition” gestures, such as a pause between raising an arm and lowering it again, and overlaps between gestures of the same modality.

right, and then stabilizes. This is a gesture, a specific observed movement, but not one that is crucial to the performance of the routine. A visualization of the main gestures making up cycle 3 is presented in Figure 1; this basic sequence is repeated three times (cycles 1, 2, and 3) with slight variation, and that variation is the main focus of this explorative description.

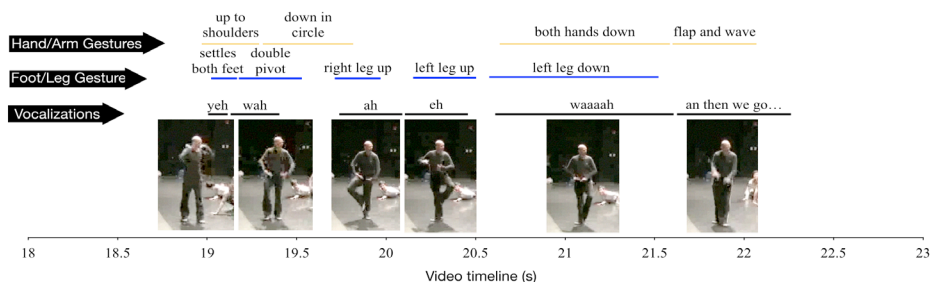


Figure 1. A visualization of cycle 3, WM's third performance of the entire dance routine that he iterates three times during the video clip. Screenshots show him performing each of the six segments. The colored lines show the approximate temporal duration of each gesture associated with a segment of the routine; gestures not associated with any segment are not shown. Text above the colored lines provides a verbal label for each gesture.

Once the list of segments composing the routine was established (presented in Figure 2), we determined which gestures from the full list – if any – seemed to correspond to each segment. This process was complicated by the facts that WM does not perform every segment of the routine all three times, and many of his gestures (steps, shuffles, waves, etc.) are not part of any segment. Each segment is defined by one key gesture – hand movements for the first two segments, foot movements for the third, fourth, and fifth, and a general whole-body movement for the sixth. These are referred to below as “segment-defining” gestures, and were easily identified. For instance, the first segment (‘hands up to shoulders’) is defined by the simultaneous upwards movement of both hands from the waist to the shoulders. In cycle 1, WM performs this segment between 1.79 and 2.26 seconds in the video timeline, raising his hands smoothly up behind his head. His legs and also shift slightly during this time-frame, but these movements are of no direct relevance to the dance routine or its component segments.

In addition to these segment-defining gestures, almost every segment has a clearly associated vocalization, and many also have a linked gesture in the other modality. For instance, in cycle 3, while the ‘hands down in circle’ segment is defined by an appropriate hand gesture, it is also linked to the ‘double pivot’ foot gesture with which it overlaps (see Figure 1). This determination was made on the grounds that the gestures appear to be part of the same whole-body movement: WM pivots his feet during ‘hands down in circle’ in order to have his torso and feet facing the same direction prior to performing the next segment, ‘right leg up’. We identified segment-

defining gestures and segment-linked gestures for each of the six segments of the dance, in each of the three cycles.

Once this process was complete, we were able to integrate the two strands of our analysis, producing the mixed qualitative-quantitative description found in section 3. It proceeds chronologically through the video, describing the teaching strategies and event-sequences that occur in each cycle, and noting how these may have depended on observable patterns of cross-modal coordination in WM's behavior.

Segment Number	Segment Name	Segment Description
1	Up to shoulders	Hands start at waist-height, in line with the front of the body, palms out. They are raised and rotated so that they end directly above the shoulders, palms facing straight up. (Feet in neutral standing position.)
2	Down into circle	Elbows move out to the sides and hands turn to have palms facing forward, fingertips pointing at each other behind the neck. Hands/arms are raised up to form a circle, and maintain that configuration as they are lowered down to hip-height. (Feet in neutral standing position.)
3	Right leg up	Right foot is raised straight up and right knee moves out to the side; bottom of foot is placed flat against inside of left leg, at knee-height. Leg is lowered after a pause. (Hands remain in circle.)
4	Left leg up	Beginning as the right foot touches the ground, the left foot is lifted and placed flat against the inside of the right knee, mirroring segment 3. (Hands remain in circle.)
5	Left leg down	Left foot is slowly lowered to the floor, keeping the toes pointed straight down. (Hands remain in circle.)
6	Coda	Drops down onto knees, which are spread out to the sides. Arms are swept back in winglike fanning motion, then brought back in, with one hand on the floor and one hand behind the back. Ends crouched, with head bowed.

Figure 2. List of segments in the dance routine, with a rough description of the bodily movements involved in each.

2. Segmentation and Vocalizing in Dance Instruction

2.1 Overview of the Video Clip

The goal of the following description, is to (a) identify and describe the phenomenon of instructional vocal sonification and (b) suggest several means by which that phenomenon can be further explored. More specifically, this and the next four sections present a fine-grained description of the vocal and gestural components of WM's instruction activity. The clip records a 25-second period of rehearsal during which the members of the company are spaced out across the practice stage, while WM occupies a position at the front (see Figure 3, which gives an overview of the setting). He stands side-by-side with another dancer, and demonstrates three repetitions of the routine over the course of the clip – this even though the dancers already know all the steps (see Figure 2 for depictions and descriptions of the six segments that compose the routine). As he demonstrates, the other dancers all practice the routine while two short bars of music loop continuously in the background.

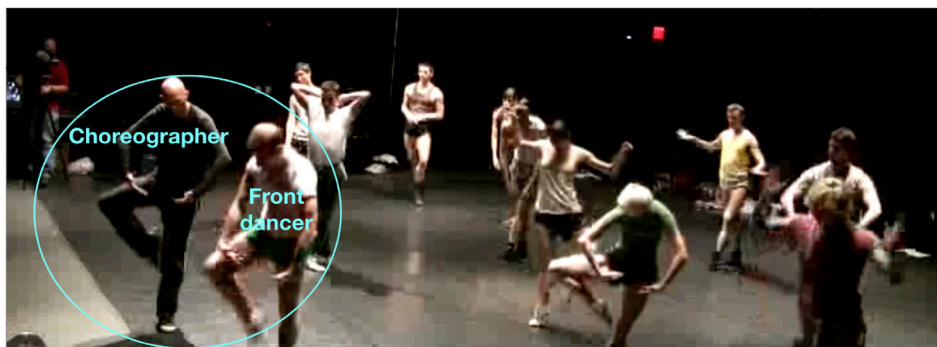


Figure 3. Screenshot giving an overview of the Random Dance Company rehearsal. The circle indicates a dialogical system comprising WM and the dancer standing at the front of the stage, with whom he works closely for two out of three cycles in the video clip.

Even though the video is long enough for each dancer to have repeated the routine three times, none of them do so (other than WM himself). Instead, they rehearse in their own ways; some strike and hold a single position from the routine, others stand and watch WM, some mimic one or more of his movements, fully performing some and minimizing others, and so on. However, many dancers' transitions from one position to another take place at the appropriate time. For instance, a dancer might hold the initial position for one dance move for several seconds, and complete it at the same time as WM, having waited for him to cycle through the preceding moves. Thus the rehearsal seems to be organized around continuous repetition of the full routine; dancers' movements deviate from it in specific, deliberate ways rather than either following along or acting independently.

The next three sections each present one of the cycles of the routine in greater detail. The descriptions they provide include many interpretive statements concerning the instructional and interpersonal consequences of what we have observed, as is *de rigueur* in rich qualitative description (Goodwin, 2000; Streeck, 2009). Our intention in giving these interpretations is both to give our best understanding of the example under discussion, and to demonstrate the utility of the style of description used here. By adding quantitative elements to qualitative description, we are able to identify potentially interesting aspects of the situation in greater detail than would otherwise be possible, which both aids in our understanding of the phenomenon of instructional vocal sonification and lets us ask more specific questions for future research.

As the subsequent sections give these descriptions, they will therefore return to the key points of interest in the example. These include the overall organization of WM's vocalizations, and whether they might plausibly exert control over the dancers' movements or sonify real or potential movement patterns, as well as changes in the highly-structured task environment the rehearsal space constitutes. We also explore WM's potential sonifying

by exploring his careful treatment of the ‘left leg down’ segment in each of the three cycles, examining how he emphasizes its performance in his demonstrations as well as in his feedback and instructions. Finally, we identify three nested instructional strategies in this dance event, each of which appears to be enabled by a particular form of multimodal embodied coordination that involves both vocalization and gesture. As mentioned earlier, that is, each cycle has a main instructional focus, respectively, these are *pinpointing*, *perfection* and *demonstration*.

2.2 Instruction and Pinpointing: Cycle 1

In this section we argue that WM’s actions during the first cycle provide specific instructions to the dancers and help to pinpoint movement features to be targets of joint attention within the overall dance routine. The cycle takes place over 7.1 seconds, and involves the segments listed above. During this time, WM employs a whole-bodied strategy in order to (1) coordinate shared attention among members of the dance company, (2) correct the front dancer who is mistiming his performance of the routine, and (3) pinpoint the crucial segment that needs specific improvement. We go through this whole-bodied strategy by investigating the function and coordination of pauses, vocalizations and gestures. In Figure 4, we present WM’s vocalizations as he draws attention to the routine and runs the segments through his body.

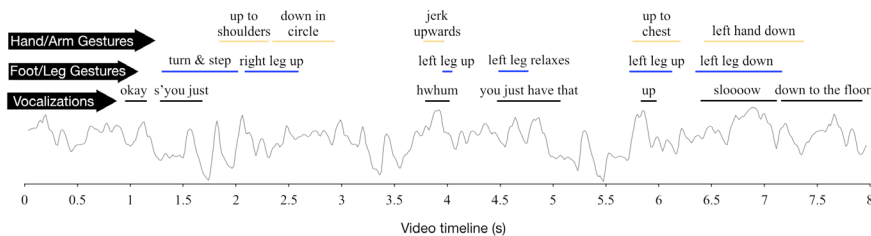


Figure 4. A visual presentation of the timing of WM’s vocalizations in the first cycle of the routine, along with other gestures. Labels for vocalizations are simple orthographic transcriptions, and the plotted line is the net acoustic intensity of the audio file accompanying the video clip.

While vocalizations are extraneous to the routine as it will later be performed for an audience, here during the rehearsal session WM produces a distinct vocalization for each segment of the routine. Throughout the clip, his vocalizations all either align closely with specific gestures, or, in the case of “okay”, “s’you just”, and “you just have that”, are verbal transitions between periods of dancing-vocalizing.² That is, *almost all* of WM’s vocal activity in the clip is the utterance of nonverbal, rhythmic sounds that are

² In most cases, only WM’s non-verbal vocalizations are potential sonifications, but there is one exception: when he says “sloooooow” during ‘left leg up’ and ‘left leg down’ in cycle 1. Here, the central vowel seems to sonify the descending leg movement (as discussed below), suggesting that even verbal vocalizations might be sonifying under the right circumstances.

equal in number and similar in duration to the segments of the dance routine he is performing with his body; in cycles 2 and 3 the individual syllables of those utterances each co-occur with particular segments of the routine.

Additionally, it appears that WM's vocalizing is part of a deliberate instructional strategy he is carrying out, one that centrally involves the coordination of multimodal dance segments with the corresponding utterances. The vocalizations in this first run-through of the dance are mostly verbal, which together with his gestures help to clarify what WM wants the dancers to attend to in the movements he's making: "okay s'you just", "you just have that", "up", "sloooooow down...". As the segments are coordinated together with the vocal utterances, this functions as a show-and-tell strategy (Trasmundi, forthcoming.), which is presented in greater step-by-step detail in Figure 5.

WM's first action is to establish joint attention among the dancers, focusing them on his own upcoming run-throughs of the routine. Before the cycle begins, the dancers are practicing individually; WM walks to the front of the stage and gathers the dancers' attention by turning his torso to face them and pausing as he waits for their gazes to find him, and then uttering "okay s'you just" (Figure 5.1). As he turns his torso back to face "forward" (i.e., towards the audience, away from the other dancers who are arrayed across the stage behind him), he focuses his gaze on the ground between himself and the front-most dancer, who is standing next to him. This visual focus on their shared space enables the other dancers to attend to the two of them as a unit, and so also to anticipate the emerging teaching event. In Figure 5 we observe how the dancers have adapted to this instruction phase: they have moved closer towards WM and they observe and mirror his behavior carefully. The front dancer in particular is mimicking his movements, watched closely by WM.

The teaching event then encounters a snag when the front dancer mis-aligns his performance of one segment with WM's performance. That WM perceives this immediately is demonstrated by the speed with which he interrupts the flow of the instruction in order to correct the misalignment. Immediately following the mis-timed segment, his gaze fixates on the front dancer's feet and he halts his performance of the routine, holding his position instead. He also gestures 'stop' with his hands (Figure 5.3). The front dancer adapts to this change in instruction mode and gazes at the WM, who then utters "you just have that" as he holds the position mid-way through the routine, before relaxing and then resuming it from where he left off, uttering "up" as he lifts his left leg (see Figure 5.4–6).

Having successfully drawn joint attention to his performance of the routine, and so also the front dancer's attention to him as a cognitive resource in instruction, WM then focuses that attention on the 'left leg down' segment (Figure 5.6). His narrowing focus is evidenced by two key moments in

the sequence of cycle 1 events as it is presented above. The first is WM's pause in Figure 5.4, which appears to be a flat refusal to begin 'left leg down' without the front dancer's attention. The second is his slow, careful preparation for performing left leg down – gradually positioning his hand and his foot, and moving his gaze to his own performance space – once that attention is gained. Further, the work of his hands gestures reinforce attention on the dynamics of 'left leg down'.







		
<p>1. WM says “okay, s’you just”; as he finishes he moves his gaze to focus on the ground between himself and the group and then finally between him and the front dancer, on whom he is focused for the remainder of the cycle.</p>	<p>2. He says “hwhum”, and as he does performs ‘left leg up A’, lifting his left foot to the starting position for ‘left leg down’; the end-time of this movement (4.09) is only 73 ms later than the end time of “hwhum” (4.017).</p>	<p>3. He then flares out his hands into a “stop” gesture, breaking the proper “circular” position.</p>
		
<p>4. There is a pause and WM says “so you just have that”, and lets his left foot inattentively drift back to the floor (4.99-5.79). He shifts his gaze to the front dancer (5.10-5.93), and at about 5.7 seconds the front dancer looks over and they make eye contact.</p>	<p>5. WM then shifts his gaze to the floor in front of himself (5.93), says “up” (5.831-5.972), lifts his hands up to shoulder height 5.66-6.17, and lifts his left foot back to the starting position for ‘left leg down’ 5.71-6.17).</p>	<p>6. He lowers his left hand (6.40-7.38) as he performs ‘left leg down’ (6.345-7.15) and says “sloooow” (6.429-7.15). The start times are within 0.084 seconds of one another, the end times are identical for the voice and foot, and the hand lags them by only 230 ms.</p>

Figure 5. Screenshots and verbal descriptions of the gesture-sequence in which WM draws the lead dancer's attention to “left leg down” during Cycle 1.

But while the qualitative analysis gives us confidence that WM's focus is on the 'left leg down' segment, it does relatively little to illuminate which aspect(s) of the segment's performance he is focused on. For this, we turn to a rudimentary quantitative description made possible by the cross-modal gesture-overlap data in Table 1. 'Left leg down' vocalizations (transcribed as "slooow", "eeei", and "waaah", for cycles 1, 2, and 3, respectively) co-occur with their segment-defining foot gestures for 87%, 90%, and 98% of their durations, respectively, in each of the three cycles. By contrast, across all other segments in all three cycles, the average degree of overlap between vocalizations and their corresponding segment is just 57%. Moreover, the 'left leg down' segments display very similar degrees of overlap between vocalizations and hand gestures as well as between hand gestures and foot gestures, while most other segments either don't have associated gestures in all three modalities, or, where they do, overlap only very little.

Other observations also support the possibility that WM's focus is on the duration of 'left leg down'. In particular, Figure 11 and Tables 2-4 each independently document that 'left leg down' and 'left leg up' are the only segments that are linked to three gestures in every cycle. Of the two, 'left leg down' seems to be the more favored, for three reasons. First, 'left leg up' is inherently preparatory for 'left leg down'; this might not be the case if 'left leg down' performances were rushed or careless, but they are not. Second, 'left leg up' is far less synchronous across its several gestures, with average vocalization-foot gesture overlaps of only 65% in cycle 1 and 54% in cycle 3. Third, 'left leg up' displays far greater variability in the durations of its realizations than does 'left leg down', with a third of its nine realizing gestures shorter than 0.28 s, and another third longer than 0.5 s. By contrast, over half of 'left leg down's nine vocal and gestural realizations are between 0.9 s and 1.1 s. In sum, then, while 'left leg up' has just as many gestural realizations as 'left leg down', they are far less consistent in duration and overlap both within and between cycles.

These data suggest that WM's focus on 'left leg down' is, more specifically, a focus on the durations of the vocal, manual, and pedal versions of 'left leg down'. Thus, at least one aspect of his instructional use of vocalizing may be the precise sonification of gesture duration as syllable duration. This is the utility gained by adding even simple quantification to qualitative description, and this is a clear example of their complementarity.

Additionally, there is at least one aspect of the performance of 'left leg up' that weighs in favor of the assumption that WM is paying special attention, not to it, but to 'left leg down'. This consideration involves the gestures depicted in Figure 6, namely, the apex of the hand gestures associated with 'left leg up' in each cycle (for cycle 1, the image is taken from the second instance of that segment). WM executes these gestures with care and active self-monitoring (e.g., visual self-monitoring), but the care is not taken during

the main part of the movement but specifically as it nears its apex as depicted in Figure 6. Given that ‘left leg up’ is unavoidably preparatory this suggests that WM’s carefulness in fact has little to do with that segment itself, and more to do with positing his hand so as to execute a downward movement during the subsequent ‘left leg down’. Moreover, these details of positioning can only make a small difference to WM’s execution of the movement; the most plausible explanation for this pattern is thus that WM is extremely concerned to produce hand movements that are precisely isomorphic with the leg and/or vocal tract movements ‘left leg down’.

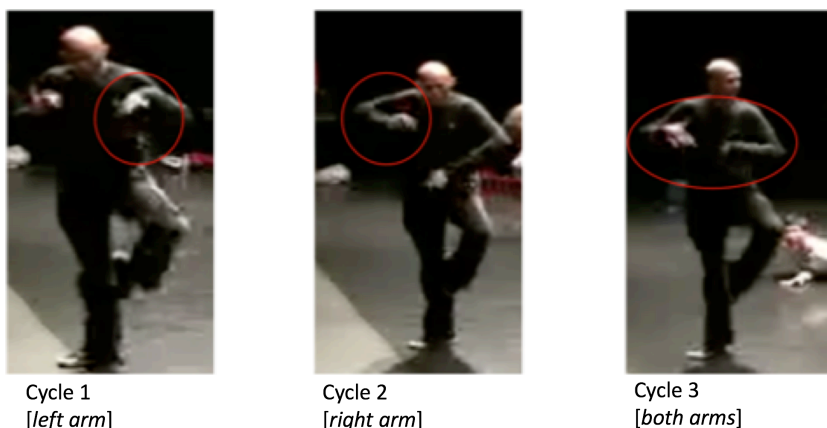


Figure 6. The moment in each cycle just before WM performs the ‘left leg down’ segment. In each case his left leg is raised, and one or both of his hands are also raised to his chest, poised to descend in tandem with his left foot. Note that these arm movements are not part of the dance routine.

Table 1. This table gives the degree to which WM’s vocalizations overlap with his accompanying gestures. (Segments marked “unimodal” had no associated vocalization, so there could be no overlap.) The values given are the percentage of the duration of each vocalization that occurred at the same time as WM was making the accompanying gesture. For ‘left leg down’, we have additionally given the co-occurrence percentages for all of WM’s gestures and vocalizations, in order to illustrate that that segment’s unusually high degree of cross-modal overlap.

Segment	Cycle 1	Cycle 2	Cycle 3	Average
1 Hands up to shoulders	Unimodal segment	0%	67%	34%
2 Hands down in circle	Unimodal segment	42%	24%	33%
3 Right leg up	Unimodal segment	41%	79%	60%
4 Left leg up	65%	84%	54%	68%
5 Left leg down (voice-hand overlap)	95%	93%	87%	92%
5 Left leg down (hand-foot overlap)	87%	90%	98%	92%
6 Coda	85%	97%	89%	90%

In order to investigate the use of rhythmic vocalizing as an instructional tool for coordinative behavior in general, we measured the cross-modal synchrony in this instruction event in depth. The measurements are visible in Table 2 below (for cycle 1; see subsequent sections for similar discussion of cycles 2 and 3) and emphasize the precision of WM's coordination of specific segments with certain verbal articulations in order to demarcate start and end times of segments.

Table 2. This table gives the degree of cross-modal overlap between vocalizations and gestures in cycle 1.

Segment	Number of modalities	Average overlap between...			
		vocalization and segment-defining gesture	vocalization and non-segment-defining gesture	hand/arm gesture and foot/leg gesture	average overlap across all modalities present
1 Hands up to shoulders	2	No vocalization	No vocalization	45%	45%
2 Hands down in circle	1	No vocalization	No vocalization	No foot/leg gesture	N/A
3 Right leg up	1	No vocalization	No vocalization	No hand/arm gesture	N/A
4 Left leg up	3	65%	64%	95%	75%
5 Left leg down	3	95%	87%	85%	89%
6 Coda	0	Not performed	Not performed	Not performed	N/A

The average percentage overlap across all modalities in relation to 'left leg down', the crucial segment, is 89% here. This information indicates that the overall overlap is high, even though it can be much higher as we observe in the second cycle. By relating those observations to the interviews (2012, pp. 34–35) – where WM states that his vocalizing is intended to elicit particular kinesthetic sensations, so that these sensations will then change how the dancers perform a new movement – we argue that cross-modal coordination is an important instructional quality, as dance is performed as a flow of nested movements with no clear starts and ends. From the data it appears that WM's instructional actions here do not serve as a representation of an ideal performance of the routine by the dancers. Rather, his actions function as an extended joint-attentional focus on timing elements of a flow of movements. Hence, vocalizations are important in pinpointing those elements of the dance, as these are difficult to discern visually in the flow of movements. This is both because they are very subtle, and because the dancers are perceiving the timing of WM's gesturing not only in relation to the timing of

the other segments of the routine, but in relation to their own movements as they follow along. WM's vocalizations plausibly function to specify these relative timings in manner

As emphasized in previous qualitative analysis of this video clip (Trasmundi, forthcoming) WM uses multiple embodied techniques (spanning gaze direction, pauses and verbal instruction, which are also elaborated in Figure 5 above) to direct the dancers' attention to aspects of the routine – usually aspects of his current performance of the routine – that he wants them to focus on and improve. The idea in this paper is that in addition to deliberate gaze, specifically timed pauses, verbal instruction and other tools, he also uses sonification. When it is relevant, WM scaffolds the front dancer's attention by coordinating segments and vocalizations as shown in Figures 1, 3, and 5. We can therefore assume that the vocalizations are at least potentially effective as an instructional device or as part of an instructional strategy. Further, we see in Table 1 how the verbal instruction “sloooooow down to the floor” overlaps with his left-leg-down movement by 87% and as such WM uses sonification to coordinate sensory-motor action.

Based on this example, dance instruction begins with an initial phase characterized by the vocal pinpointing of the start and end times of segments. This seems to be done in order to break down the continuous flow of the routine into small, distinct pieces that can then each be practiced with a degree of independence. Vocalizations coordinated with specific gestures are being used here to bracket out much of the complexity of the dance, as this discrete and step-wise structural decomposition of the routine is different from the dancers' bodily perception of dance flow. This demarcating activity appears to be one important function in this case of dance instruction.

Another function involves emphasizing specific dynamics of individual segments. Given that the dancers are aware of which segments they need to improve, WM is able to use specific aspects of his performance to help them make those improvements, in order to achieve The main goal: a precise performance of the routine exactly as he intends it. This interpretation of the evidence is supported by another potential sonification in WM's vocalizing. This time, the vocal parameter involved is pitch, and the parameter being sonified appears to be either foot height or foot acceleration during WM's performances of 'left leg down'. Figure 7 provides the average pitch of the vocalizations associated with this segment in each cycle, showing gradually decreasing pitch in cycle 1 (the “-ooow” from “sloooow”) and almost no change in cycles 2 and 3 (“eeei” and “waaah”, respectively). These vocalizations are notable for their sustained central vowels, and this, together with the highly perceptually salient pitch drop of the “-ooow” in “sloooow” and the equally salient *lack* of pitch change in “eeei” and “waaah”, suggest that this correspondence is highly worthy of exploration as potential sonification, and as evidence of WM's instructional goals.

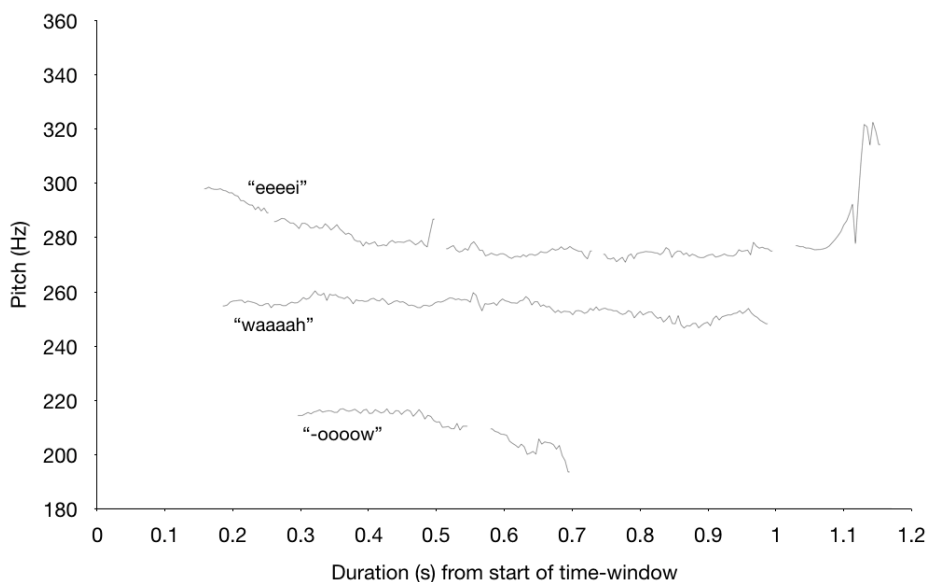


Figure 7. Overlay of three pitch charts extracted from Praat, with identical pitch ranges (y-axis scale) and time-window durations (x-axis scale). Praat pitch settings had default values with the following exceptions: pitch range 180Hz-350Hz, voicing threshold 0.22, octave cost 0.06, octave-jump cost 0.6, voiced/unvoiced cost 0.3. Note: “ooooow” comes from cycle 1’s “sloooooow”, “eeeeei” is from cycle 2, and “waaaah” is from cycle 3.

In the following paragraphs we describe how WM builds on the company’s shared experience of cycle 1 as he moves into the second cycle, in order to further refine the dancers’ performances.

2.3. Instruction and Perfection: Cycle 2

Having established the situation as an instruction task, WM moves on from pinpointing the task and its elements to emphasizing the perfect performance, which involves precise duration, force, coordination, and so on. Again as in the first cycle, vocalization is an important resource for guiding attention and marking the start and end times of different segments within the dance routine. However, in this cycle WM relies on vocalizations that sonify duration and force of movements, rather than symbolic content after he has pointed out that the cycle is going to be repeated “so I’ve gone”. Cycle 2 thus involves precision-sonification of acceleration as pitch change (i.e., both are constant/non-changing) in addition to the duration-as-voicing, which is much more precise than in cycle 1 and as we will see later in cycle 3 as well. First, we visualize how the segments are coordinated with the vocalizations in this cycle in Figure 8 below.

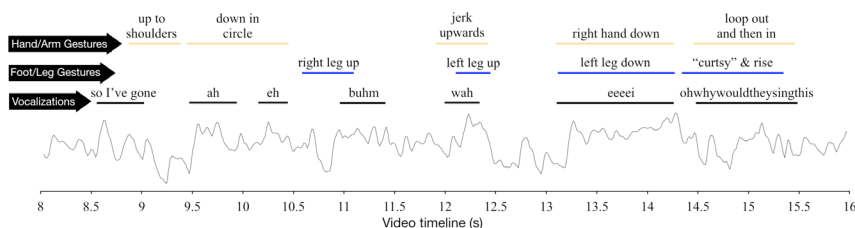


Figure 8. A visual presentation of the timing of WM's vocalizations in the second cycle of the routine, along with other gestures. Labels for vocalizations are simple orthographic transcriptions, and the plotted line is the net acoustic intensity of the audio file accompanying the video clip.

In cycle 2 we observe that WM switches from a particular pattern of word-based instruction to one of rhythmic sonification. A change in vocalization pattern seems to enhance performance in terms of efficiency and precision when analyzed as cross-modal synchrony (Table 3). We have measured the general overlap within this cycle (see Table 3), and found a high degree of multimodal co-occurrence. Further, when taken in isolation, the gestures associated with “left leg down” are far more precisely coordinated in the second cycle than they are in the first cycle.

Table 3. This table gives the degree of cross-modal overlap between vocalizations and gestures in cycle 2.

Segment	Number of modalities	Average overlap between...			
		vocalization and segment-defining gesture	vocalization and non-segment-defining gesture	hand/arm gesture and foot/leg gesture	average overlap across all modalities present
1 Hands up to shoulders	2	0%	No foot/leg gesture	No foot/leg gesture	0%
2 Hands down in circle	2	42%	No foot/leg gesture	No foot/leg gesture	42%
3 Right leg up	2	41%	No hand/arm gesture	No hand/arm gesture	41%
4 Left leg up	3	84%	62%	66%	71%
5 Left leg down	3	93%	90%	97%	93%
6 Coda	3	N/A	N/A	N/A	86%

“Okay s’you just” and “so I’ve gone” are preparatory to the choreographer’s beginning cycles 1 and 2, respectively. Where WM relies on a word-based vocalization pattern in cycle 1, he changes this strategy in cycle 2, appearing to move beyond traditional verbal instruction by vocalizing rhythm, force and duration and/or acceleration. Even though cycle 2 is almost just as long as cycle one (cycle 1 is 7.01 seconds and cycle 2 is 7.04), the cross-modal coordination is tighter both in general and when focusing on the crucial segment in the dance phrase. While the segments are identical and performed in

similar ways in all the cycles, it is primarily the vocalization pattern and its timing that reveal significant changes in the instruction strategy. Thus from Table 3 we identify an average combined overlap of 93% for ‘leg left down’ within this cycle, which appears related to the changes in WM’s vocalizations. In all cycles, WM uses audible voicing to sonify the active performance of the segments. However, one reason why the overlap is higher in cycle 2 could be that he does not need to coordinate the length of words with gestures. For instance, the “buhm” “wah” “eiuiii” in cycle 2 are equivalents to “hwhum” “up” “sloooooow down to the floor” in cycle 1. Finally, as he has the opportunity to repeat the dance phrase over and over, he can emphasize different aspects of the task over time. In this cycle, precision appears to be an important aspect compared to the first cycle’s emphasis on segmenting the main challenges.

2.4. Instruction and Demonstration: Cycle 3

Between the second and third cycle, WM turns towards the dance company and observes for a couple of seconds before he initiates the third and final cycle. This cycle 3 deviates from the previous cycles in multiple ways. First, the length is approximately half the length of cycle 1 and 2 (cycle 3 is 3.375 seconds). Further it is looser and more demonstrative as we will describe in detail below.

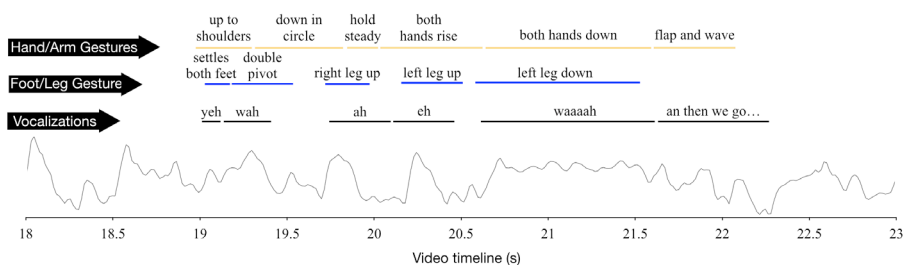


Figure 9. A visual presentation of the timing of WM’s vocalizations in the third cycle of the routine, along with other gestures. Labels for vocalizations are simple orthographic transcriptions, and the plotted line is the net acoustic intensity of the audio file accompanying the video clip.

First, WM sonifies each segment in real-time, but this time he does not provide the dancers with any preparatory signals as he does in cycles 1 and 2. And second, he moves to a new part of the stages and performs while facing the whole group of dancers, rather than remaining paired-up with the dancer in front (see Figure 10).

WM’s reorientation marks a different instructional focus from the previous cycles. Having moved away from the front dancer and engaged with the group at large, he now runs the dance through his body in real-time. In the two previous cycles, he adapted to the front dancer’s movements and coordinated his actions and perceptions in relation to the performance of the front dancer. While WM always sonifies his own movements, the sonification is – in each

cycle – specific to the way those movements are intended to affect the dancers (i.e., specific to the instructional effect he is trying to have). Thus, cycle 2 was a methodical and focused demonstration, and accordingly in that cycle WM carefully sonified his own performance. Cycle 3, however, is a faster and less methodical run-through of the routine. It is thus possible that WM may be more focused on the surrounding dancers during this cycle, and accordingly, it may be that his vocalizations entirely or in part sonify parameters of the dancers' performances rather than his own. The voicing functions here to sonify the duration of a segment-defining gesture so that the dancers can learn the segment's timing without relying on visual perception. This is a useful teaching tool, as the dancers often watch their own bodies as they rehearse. As WM has pinpointed how the segments are performed and timed with the music in a multimodal embodied way to all dancers, he has established a shared memory, where the dancers need fewer cues to adapt their movements to the intended outcome. WM has created the basis for practicing and he is now in a position where he perceives the dancers' progress, that is, he is observing how they are adjusting to his instructions in cycle 1 and 2. However, rather than just observing the dancers, he engages in observation *and* instruction which provide the dancers with rich resources for continually adjusting their behavior in relation to WM's vocalization.



Figure 10. The layout of the rehearsal stage after WM shifts position and orientation during the interim between cycles 2 and 3. Here, he is performing the 'left leg down' segment of cycle 3.

WM's focus during this cycle appears to be on a combination of instruction, demonstration and observation: he turns towards the group, he speeds up to real-time performance, and he does not give the dancers preparatory cues to coordinate their practice in synchrony with him. Segments are less cross-modally synchronous here than they were in the perfection-focused cycle 2, as seen in Table 4.

Table 4. This table gives the degree of cross-modal overlap between vocalizations and gestures in cycle 3.

Segment	Number of modalities	Average overlap between...			
		vocalization and segment-defining gesture	vocalization and non-segment-defining gesture	hand/arm gesture and foot/leg gesture	average overlap across all modalities present
1 Hands up to shoulders	3	67%	71%	72%	70%
2 Hands down in circle	3	24%	72%	51%	49%
3 Right leg up	3	79%	65%	55%	72%
4 Left leg up	3	54%	79%	70%	68%
5 Left leg down	3	87%	98%	89%	91%
6 Coda	3	N/A	N/A	N/A	81%

2.5. Comparing the Three Cycles: Nested Functions in the Dance Instruction Task

The routine's three cycles share several key similarities: they each contain the same segments, they each involve rich multimodal forms of instruction, and they each involve three versions of the 'left leg down' segment with durations close to one second (see Figure 11 for a full overview of all three cycles). On the other hand, the cycles also differ in significant ways. In cycle 1 WM is more verbal and concerned with fixing joint attention; in cycle 2 he is focused on the nuances of his performance and sonifies his own movements more precisely, and in cycle 3 he performs at full speed, and allows his vocalizations and gestures to become less synchronous. Key observations in support of these suggestions are that cycle 3 is half the length of cycles 1 and 2, that cycle 1's vocalizations are far more verbal than are those of cycles 2 and 3, and that cycle 2 has the greatest degree of cross-modal synchrony in all segments, including the crucial 'left leg down'. If the aim of dance instruction were a precise, fixed outcome, one would not expect such a high degree of variation across multiple repetitions of a routine. Moreover, we have noted small-scale changes in WM's performances in relation to (i) tempo, (ii) cross-modal gestural overlap, and (iii) vocalization duration and pitch. We have suggested that these changes are adaptive responses to his shifting instructional focus.

For instance, we have suggested that WM's vocalizations in each cycle may have been part of an attempt to elicit sensations of movement that served the instructional purposes of pinpointing, perfecting, and demonstrating, respectively. In all three cycles, WM uses audible voicing (vibration of the vocal folds) to sonify the active performance of dance-step-defining

movements. That is, he sonifies gesture duration as vocalization duration. His vocalization durations are correlated with the durations of the gestures that define each dance step. We have described this by showing that they overlap in time and by comparing the degree of overlap within and between cycles. This sonification, together with the proposals developed in section 1.2 and our preceding descriptions of each cycle's instructional context, provides tentative support for our analysis in terms of pinpointing, perfecting, and demonstrating.

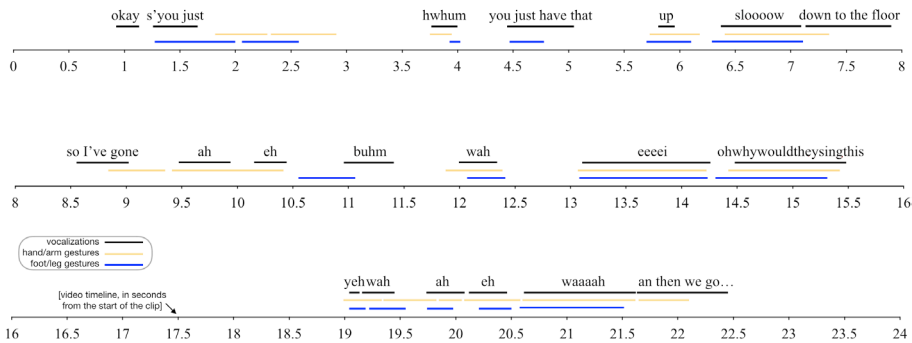


Figure 11. A visualization of WM's full, three-cycle performance of the routine, and the approximate overlaps between his vocalizations and gestures as he does so.

Perhaps the clearest evidence in favor of this cross-cycle functional difference in WM's vocalizing relates to the way in which he sonifies 'left leg down'. He is more precise and careful in his performances of 'left leg down' than he is when performing other segments. His sonifications of step-defining gesture duration as voicing is (1) more precise for gestures linked to this segment than for any other gesture within each cycle, and (2) more consistent across cycles for this segment than for any other (see Figure 11 and Tables 1 and 5). Because cycle 3 is twice-as-fast as the other cycles, it is notable that 'left leg down' has the same duration in cycle 3 as it does in cycle 2, and a *longer* duration than it does in cycle 1 (see Table 5 for gesture durations in all three modalities). This is not true for any other gesture. (For instance, both 'right leg up' and 'left leg up' are about one half as long in cycle 3 as they are in cycle 1.) This supports our claim that this segment in particular is being emphasized and so must be of special relevance in the instruction task. From the descriptions we see how 'left leg down' has high degrees of overlap between vocalizations and *both* other modalities in *all* cycles, where other segments have, at most, high degrees of overlap between vocalization and a dance-step-defining gesture in *one* modality in *some* cycles. 'Left leg down' displays not only longer duration, but also more consistent duration as well as greater cross-modal overlap. Our point here is just that this one segment displays all three of those properties in all three

cycles. Thus we're assuming that increased duration is the result of deliberate attentional emphasis because it co-occurs with these other properties of consistent duration and greater cross-modal overlap.

Table 5. A table giving the durations of each segment-linked gesture in all three cycles, in seconds.

Segment name	Gesture duration (s)								
	Cycle 1			Cycle 2			Cycle 3		
	Vocaliza- tions	Hand/Arm gestures	Foot/Leg gestures	Vocaliza- tions	Hand/Arm gestures	Foot/Leg gestures	Vocaliza- tions	Hand/Arm gestures	Foot/Leg gestures
Hands up to shoulders	–	0.470	0.760	0.418	0.610	–	0.131	0.380	0.170
Hands down in circle	–	0.580	–	0.333	0.960	–	0.303	0.480	0.360
Right leg up	–	–	0.490	0.542	–	0.540	0.382	0.250	0.273
Left leg up A	0.322	0.210	0.110						
Left leg up B	0.141	0.510	0.460	0.407	0.520	0.280	0.322	0.550	0.220
Left leg down	0.721	0.980	0.805	1.204	1.140	1.100	1.017	1.050	0.928
Coda	–	–	–	0.956	0.960	0.880	0.705	0.440	–

Altogether, we have shown how cycle 1 serves as a contextualizing phase where the dancers' attention is guided by pointing out relevant segments in general terms ("up", "sloooow down to the floor", "just like that" etc.). In the second cycle WM is more concerned with precision and dynamics of the movements as the dancers are now aligned and tuned in on what is crucial based on the work in cycle 1. Cycle 3 differs as WM faces the group of dancers and runs the dance sequence through his body in real-time, providing the dancers with sonification so that there is no need for them to pay visual attention to his movements in order to match their own timing to his. However, as they have performed cycles 1 and 2 while also paying visual attention to the multimodal instruction, the dancers are now relying primarily on the sonifications as instructional anchors for practicing.

This conclusion fits with the observed degrees of cross-modal overlap for different segments that are reported above. From those measurements we can conclude that the three cycles vary in their degree of precision, in their instructional focus and intent, and in their vocal and kinesthetic phenomenology. We argue that this is not due to chance, but is rather a result of professionalism: teaching is a complex activity that relies on flexible adaptation on the part of the teacher, who needs to coordinate with the needs of and feedback from other dancers.

3. Conclusion

3.1. Overview of the Video Clip

In this paper we have attempted to outline some of the general features of "instructional vocal sonification", a hitherto-unstudied class of phenomena in the domain of interpersonal interaction. We have described in detail the various intra- and inter-bodily dynamics that are involved in instructional sonification. As we discuss in the final section below, this study paves the way for future work by identifying important variables in this complex phenomenon.

In this paper, we chose three starting points. The first of these was single clear example of the phenomenon as practiced by a professional choreographer working with his dance company, captured on video and discussed in two previous publications (Fagergren, 2012; Trasmundi, forthcoming). The second was the tradition of cognitive-ethnographic description of multimodal features of interaction, and the third was the rapidly-expanding literature on the use of artificial sonification for training and performance-enhancement in skilled movement.

First, we presented a general definition of "sonification" as the perceptible projection of data relations from a source domain into the domain of acoustic structure. Second, we discussed this phenomenon in relation to the source of the video data used here – Wayne McGregor ("WM") and the *Random Dance Company*. This discussion concluded that instructional vocal sonification

likely involves mappings between multiple high-dimensional domains of data relations, such that the kinetic patterns being perceived, sonified, and consequently changed are too complex for humans to apprehend directly. We propose that expert teachers may engage with them, instead, by attending to “feelings of moving”. These are explicitly described by choreographers as characterizing different ways of performing a single gesture, and uses them to describe the desired outcomes of his instructional interactions with dancers. That is, we hypothesized that rather than explicitly tracking and vocally sonifying high-dimensional kinetic variables, choreographers instead attend to narrow kinesthetic aspects of the movement in question, and use these both as the basis for their sonifying vocalizations and also as targets for their instructional interventions.

Altogether, we have identified a local systematicity in the way the choreographer uses vocal sonification and gestural dynamics strategically to attract attention to and mark specific aspects of the dance performance. Our interpretation is that differences in the precision of his sonification of gesture duration appear to parallel changes in his teaching strategy from one repetition of the routine to the next.

These observations, we believe, warrant future applications of the mixed methodological tools used here to other cases of instructional vocal sonification, as well as to instances of interpersonal interaction that involve other forms of vocal co-organization. These applications, if they include sufficiently robust data-gathering and quantitative components, will develop the aims of the current paper in three ways. First, by testing the hypothesis mentioned above, that is, that choreographers use “feelings of moving” both as bases for attempted sonification and as instructional targets. Second, they may also expand it into a wide variety of additional hypotheses motivated by a general form of the idea developed in section 1, along the lines that vocal sonification is likely to involve domain-specific means of those kinds to talk about, compare, recall, and notice the extremely complex, many-dimensional environmental parameters being sonified. As such, these additional hypotheses would concern the conceptual, verbal, and attentional means that expert practitioners in different domains employ to perceive and act with respect to crucial but subtle aspects of the phenomena with which they are engaged. Third and finally, as elaborated below, we hope that by means of this or some other process, future applications of the ideas proposed here will result in the development of systematic tools for the description and comparison of instances of instructional vocal sonification.

3.2. Building on the present study

The foregoing description is intended to lay groundwork for future studies of instructional vocal sonification. Having outlined and described the phenomenon here, in an admittedly tentative fashion, the way is clear

both to describe other examples of the phenomena in greater detail – using improved versions of the tools chosen in the present article – and to compare those examples one to another. Not only might instances of choreographic instructional vocal sonification be compared, but dance might be compared to other movement-based skill acquisition, and furthermore, there may be opportunities to compare instructional vocal sonification to other related phenomena, such as the use of rhythmic nursery rhymes as part of games like “patty cake”.

There are several specific questions raised here, which will demand other methods as well as other types of data. First, there is the question of what natural vocal sonification means. In order to address this, explicit attempts at sonification will need to be recorded under natural conditions, ideally using motion tracking and multiple cameras to capture limb movement and high-fidelity microphones to capture sound. By working closely with choreographers and dancers and collecting interview data as well as natural behavioral data researchers can obtain a far more in-depth description of the intentions of choreographers in using nonverbal vocalizations as instructional tools, and how dancers experience the function of such methods used. Thus, targets for sonification would need to be provided – for instance by providing rich phenomenological descriptions of how choreographers plan and experience their use of vocal instructions and long-term investigations of how sonification help narrow down the possible information available to dancers from instructional vocal sonification, by identifying particular movement parameters – onset timing, rhythm, force-related aspects, and so on – that they modify in response to changes in selected acoustic parameters. This research might be inspired by the work of Sheets-Johnstone (2011), Kimmel and Preuschl (2016), and others. These methods might even be augmented by others, such as assisted recall – wherein the dance practitioners themselves would watch videos of their activity and verbally recall intentions, feelings, and other aspects of their experiences. Such work would allow quantitative comparison of the putative sonification to candidate sonified movements, in order to determine whether attempted sonification is successful.

Resolving this question will involve simultaneously answering several others, including: what are the non-acoustic parameters being sonified, and what acoustic parameters are they mapped onto? Do these non-acoustic parameters correspond to “feelings of moving” as described by choreographers? To what extent do these parameters vary from one instance of instructional vocal sonification to another? These questions are suited to expert quantitative analysis of acoustic and motion-tracking data, perhaps inspired by work done by Effenberg (2004), Cesari and colleagues (2014), and Schaffert and Mattes (2015).

Finally, it will be necessary to build up a corpus of this work over time in order to make any generalizations about instructional vocal sonification. Such a corpus will enable the development of explanatory models for instructional vocal sonification, through the gradual exploration of its functions and enabling conditions, both in dance and in other varieties of interpersonal interaction.

Acknowledgements

Many thanks are due to each of our three extraordinarily dedicated reviewers for working through many different aspects of this paper in incredible depth. Their efforts were both helpful and instructive. We would like to thank Stephen Cowley, as well, for his own detailed and extremely helpful feedback on an earlier draft of this paper. Rasmus Nielsen provided invaluable help with several figures, as well as analytical expertise and moral support. All together, we couldn't possibly have wished for more capable advisors. We are grateful, as well, to David Kirsh for the kind use of his video data, and to Joanna Rączaszek-Leonardi for her inexhaustible patience.

References

- Arvaniti, A. (2009). Rhythm, timing, and the timing of rhythm. *Phonetica*, 66, 46–63.
- Auvray, M., Lenay, C., & Stewart, J. (2009). Perceptual interactions in a minimalist virtual environment. *New Ideas in Psychology*, 27, 32–47.
- Baber, C., Chemero, A., & Hall, J. (2017). What the jeweller's hand tells the jeweller's brain: Tool use, creativity, and embodied cognition. *Philosophy & Technology*.
- Beckner, C., Ellis, N. C., Blythe, R., Holland, J., Bybee, J., Ke, J., . . . Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, 59(suppl. 1), 1–26.
- Bingham, G. P. (1988). Task-specific devices and the perceptual bottleneck. *Human Movement Science*, 7, 225–264.
- Brinck, I. (2016, 8-10 Dec). Joint improvisation in the arts practices: Entrainment, engagement, and expert skill. Proceedings from “*A Body of Knowledge – Embodied Cognition and the Arts*”, Irvine, CA.
- Buhrmann, T., Di Paolo, E., & Barandiaran, X. (2013). A dynamical systems account of sensorimotor contingencies. *Frontiers in Psychology*, 4, 285.
- Burrell, S. (1995). Tune into “show and tell”. *First Teacher*, 16(1), 12–13.
- Butler, E. A. (2011). Temporal interpersonal emotion systems: The “TIES” that form relationships. *Personality and Social Psychology Review*, 15(4), 367–393.

- Cesari, P., Camponogara, I., Papetti, S., Rocchesso, D., & Fontana, F. (2014). Might as well jump: Sound affects muscle activation in skateboarders. *PloS One*, 9(3), e90156.
- Chollet, D., Madani, M., & Micallef, J. P. (1992). Effects of two types of biomechanical bio-feedback on crawl performance. In D. MacLaren, T. Reilly, & A. Lees (Eds.), *Biomechanics and medicine in swimming (swimming science)* (pp. 48–53). London: E & F Spon.
- Cowley S. J. (1998). Of timing, turn-taking, and conversations. *Journal of Psycholinguistic Research*, 27(5), 541–71.
- Cowley, S. J. (2007). Cognitive dynamics and distributed language. *Language Sciences*, 29(5), 575–583.
- Cowley, S. J. (2009). Distributed language and dynamics. *Pragmatics & Cognition*, 17(3), 495–507.
- Cowley, S. J. (2011). *Distributed language*. Amsterdam: John Benjamins.
- Cowley, S. J., & Harvey, M. (2016). The illusion of common ground. *New Ideas in Psychology*, 42, 56–63.
- Cowley, S. J., & Madsen, J. K. (2014). Time and temporality: Linguistic distribution in human life-games. *Cybernetics and Human Knowing*, 20(3), 172–185.
- Cuffari, E. C. (2014). Keep meaning in conversational coordination. *Frontiers in Psychology*, 5(1397).
- Cummins, F. (2009). Rhythm as an affordance for the entrainment of movement. *Phonetics*, 66, 15–28.
- Cummins, F. (2011). Periodic and aperiodic synchronization in skilled action. *Frontiers in Human Neuroscience*, 5(170), 1–9.
- Cummins, F. (2012). Looking for rhythm in speech. *Empirical Musicology Review*, 7(1–2), 28–35.
- Cummins, F., & Simko, J. (2009). Notes on phase coordination, and their application to rhythm and timing in speech. *Journal of the Phonetic Society of Japan*, 13(3), 1–12.
- Dale, R., Fusaroli, R., Duran, N. D., & Richardson, D. C. (2014). The self-organization of human interaction. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (1st ed., pp. 43–96). New York, NY: Elsevier Inc., Academic Press.
- De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, 6, 485–507.
- Di Paolo, E., Rohde, M., & Iizuka, H. (2008). Sensitivity to social contingency or stability of interaction? Modelling the dynamics of perceptual crossing. *New Ideas in Psychology*, 26, 278–294.
- Dubus, G., & Bresin, R. (2013). A systematic review of mapping strategies for the sonification of physical quantities. *PLoS One*, 8(12), e82491.

- Duran, N. D., & Dale, R. (2014). Perspective-taking in dialogue as self-organization under social constraints. *New Ideas in Psychology*, 32, 131–146.
- Dyer, J. F. (2017). *Human movement sonification for motor skill learning*. PhD thesis. Belfast: Queen's University.
- Effenberg, A. O. (2004). Using sonification to enhance perception and reproduction accuracy of human movement patterns. Proceedings from *The International Workshop on Interactive Sonification*, Bielefeld.
- Effenberg, A. O. (2005). Movement sonification: Effects on perception and action. *IEEE Multimedia*, 53–59.
- Effenberg, A. O. (2007). Movement sonification: Motion perception, behavioral effects, and functional data. Proceedings from *2nd International Workshop on Interactive Sonification*, York.
- Effenberg, A. O., Fehse, U., Schmitz, G., Krueger, B., & Mechling, H. (2016). Movement sonification: Effects on motor learning beyond rhythmic adjustments. *Frontiers in Neuroscience*, 10, 219.
- Fagergren, E. (2012). "WA-UM-EI" *How a choreographer can use sonification to communicate with dancers during rehearsals*. Unpublished thesis. Linköping: Linköping University. Available at <http://www.diva-portal.org/smash/get/diva2:546262/FULLTEXT01.pdf>.
- Fusaroli, R., Perlman, M., Mislove, A., Paxton, A., Matlock, T., & Dale, R. (2016). Timescales of massive human entrainment. *PLoS One*, 10(4), e0122742.
- Fusaroli, R., Rączaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, 32, 147–157.
- Gangopadhyay, N. (2017). The future of social cognition: Paradigms, concepts, and experiments. *Synthese*, 194(3), 655–672.
- Geertz, C. (1973). *The interpretation of cultures*. London: Fontana.
- Gibson, J. J. (1979) *The ecological approach to visual perception*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Goffman, Erving. 1981. *Forms of talk*. Pennsylvania, PA: University of Pennsylvania Press, 124–159.
- Golonka, S. (2015). Laws and conventions in language-related behaviors. *Ecological Psychology*, 27, 236–250.
- Goodwin, C. (1994). Professional vision. *American Anthropologist*, 96(3), 606–633.
- Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, 32: 1489–1522.
- Goodwin, C. (2002). Time in action. *Current Anthropology*, 43, 19–35.
- Goodwin, C. (2003). The body in action. In J. Coupland and R. Gwyn (Eds.), *Discourse, the body and identity* (pp. 19–42). New York, NY: Palgrave/Macmillan.

- Goodwin, C. (2007). Language, culture, social organization and the material world: Why a five field approach is necessary. *Teaching Anthropology: SACC Notes*, 13(2), 5–9 34 pages.
- Goodwin, C. (2013). The co-operative, transformative organization of human action and knowledge. *Pragmatics*, 46, 8–23.
- Hagendoorn, I. (2004). Some speculative hypotheses about the nature and perception of dance and choreography. *Journal of Consciousness Studies*, 11(3–4), 79–110.
- Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, 51, 347–356.
- Harvey, M. (2015). Content in languaging: Why radical enactivism is incompatible with representational theories of language. *Language Sciences*, 48, 90–129.
- Harvey, M. (2017). Posing-solving can be explained with representations, because it is a form of perception-action. *Constructivist Foundations*, 13(1), 710–712.
- Harvey, M. (forthcoming). *Towards an account of linguistic interaction as vocal co-organizing in soundscapes*. PhD thesis. Odense: University of Southern Denmark.
- Hodges, B. H. (2009). Ecological pragmatics: values, dialogical arrays, complexity and caring. *Pragmatics & Cognition*, 17(3), 628–652.
- Hodges, B. H. (2014). Righting language: A view from ecological psychology. *Language Sciences*, 41, 93–103.
- Hodges, B. H., & Baron, R. M. (2007). On making social psychology more ecological and ecological psychology more social. *Ecological Psychology*, 19, 79–84.
- Hodges, B. H., Steffensen, S. V., & Martin, J. E. (2012). Caring, conversing, and realizing values: new directions in language studies. *Language Sciences*, 34, 499–506.
- Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.
- Jarvilehto, T. (2009). The theory of the organism-environment system as a basis for experimental work in psychology. *Ecological Psychology*, 21(2), 112–120.
- Jensen, T. W. (2014). Emotion in languaging: languaging as affective, adaptive, and flexible behaviour in social interaction. *Frontiers in Psychology*, 5(720), 1–14.
- Jewitt C. (2009). *The Routledge handbook of multimodal analysis*. London: Routledge.
- Kelso, J. A. S. (2009). Synergies: Atoms of brain and behavior. In D. Sternad (Ed.), *Progress in motor control: A multidisciplinary perspective*. New York, NY: Springer.

- Kennel, C., Streese, L., Pizzera, A., Justen, C., Hohmann, T., & Raab, M. (2015). Auditory reafferences: the influence of real-time feedback. *Frontiers in Psychology*, 6(69).
- Kimmel, M., & Preuschl, E. (2016). Dynamic coordination patterns in Tango Argentino: A cross-fertilization of subjective explication methods and motion capture. In J.-P. Laumond and N. Abe (Eds.), *Dance notations and robot motion, Springer Tracts in Advanced Robotics III*. Switzerland: Springer International Publishing.
- Kirby, R. (2009). Development of a real-time performance measurement and feedback system for alpine skiers. *Sports Technology*, 2, 43–52.
- Kirsh, D. (2010). Thinking with the body. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of The 32nd Annual Meeting of the Cognitive Science Society* (pp. 2864–2869). Austin, TX: Cognitive Science Society.
- Kirsh, D. (2012). Running it through the body. *Proceedings of The 34th Annual Cognitive Science Society*.
- Kirsh, D. (2013). Embodied cognition and the magical future of interaction design. *ACM Transactions on Computer-Human Interaction*, 20(1), 1–30.
- Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J. H., Miner, N., & Neuhoff, J. (2010). *Sonification report*. Department of Psychology, University of Nebraska at Lincoln.
- Laroche, J., Berardi, A. M., & Brangier, E. (2014). Embodiment of intersubjective time: Relational dynamics as attractors in the temporal coordination of interpersonal behaviors and experiences. *Frontiers in Psychology*, 5(1180).
- Latash, M. L. (2015). Bernstein's "Desired Future" and the physics of human movement. In M. Nadin (Ed.), *Anticipation: Learning from the past* (pp. 287–299). Zurich: Springer International Publishing.
- Linell, P. (2005). *The written language bias in linguistics*. New York, NY: Routledge.
- Linell, P. (2009) *Rethinking language, mind, and world dialogically: Interactional and contextual theories of human sense-making*. Charlotte, NC: Information Age Publishing.
- Linell, P. (2015). Dialogism and the distributed language approach: a rejoinder to Steffensen. *Language Sciences*, 50, 120–126.
- Love, N. (2004). Cognition and the language myth. *Language Sciences*, 26, 524–544.
- MacWhinney, B. (2005). The emergence of linguistic form in time. *Connection Science*, 17(3–4), 191–211.
- Malafouris, L. (2014). *How things shape the mind*. Cambridge, MA: MIT Press.
- Marmelat, V., & Delignières, D. (2012). Strong anticipation: Complexity matching in interpersonal coordination. *Experimental Brain Research*, 222, 137–148.

- McBeath, M. K., Schaffer, D. M., & Kaiser, M. K. (1995). How baseball outfielders determine where to run to catch fly balls. *Science*, 268(5210), 569–573.
- Mechner, F. (1995). *Learning and practice in skilled performance*. New York, NY: The Mechner Foundation.
- Mechsner, F., Kerzel, D., Knoblich, G., & Prinz, W. (2001). Perceptual basis of bimanual coordination. *Nature*, 414, 69–73.
- Mondada L (2008). Doing video for a sequential and multimodal analysis of social interaction: videotaping institutional telephone calls. *FQS (Forum: Qualitative Sozialforschung)*, 9(3). Retrived from <http://www.qualitative-research.net/index.php/fqs/article/view/1161/2571>.
- Mondada, L. (2012). Video analysis and the temporality of inscriptions within social interaction: The case of architects at work. *Qualitative Research*, 12(3), 304–333.
- Murgia, M., Sors, F., Vono, R., Muroi, A. F., Delitala, L., Di Corrado, D., & Agostini, T. (2012). Using auditory stimulation to enhance athletes' strength: An experimental study in weightlifting. *Review of Psychology*, 19(1), 13–16.
- Mylopoulos, M., & Farhat, W. (2014). "I can do better": Exploring purposeful improvement in daily clinical work. *Advances in Health Sciences Education. Theory and Practice*, 20(2), 237–283.
- Nolan, F., & Jeon, H.-S. (2014). Speech rhythm: A metaphor? *Philosophical Transactions of the Royal Society*, 369(1568), 1–11.
- Norris S. (2004). *Analyzing multimodal interaction: A methodological framework*. London: Routledge.
- Norris S. (2011). Identity in (inter)action: introducing multimodal (inter)action analysis. *Frontiers in Human Psychology*, 6, 69.
- Pedersen, S. B. (2012). Interactivity in health care: Bodies, values and dynamics. *Language Sciences*, 34(5), 532–542.
- Pedersen, S. B., & Steffensen, S. V. (2014). Temporal dynamics in medical visual systems. *Cybernetics & Human Knowing*, 21, 143–157.
- Pedersen, S.B. (2015). *The cognitive ecology of human errors in emergency medicine: An interactivity-based approach*. Ph.D. thesis. Odense: University of Southern Denmark.
- Phillips-Silver, J., & Trainor, L. J. (2005). Feeling the beat: Movement influences infant rhythm perception. *Science*, 308, 1430.
- Port, R. F. (2010). Rich memory and distributed phonology. *Language Sciences*, 32, 43–55.
- Port, R. F., Tajima, K., & Cummins, F. Speech and rhythmic behavior. In G. J. P. Savelsbergh, H. van der Maas, & P. C. L. van Geert (Eds.), *The non-linear analyses of developmental processes*. Amsterdam: Royal Dutch Academy of Arts and Sciences.

- Rączaszek-Leonardi, J. (2010). Multiple time-scales of language dynamics: An example from psycholinguistics. *Ecological Psychology*, 22(4), 269–285.
- Ramenzoni, V. C., Riley, M. A., Shockley, K., & Baker, A. A. (2012). Interpersonal and intrapersonal coordinative modes for joint and single task performance. *Human Movement Science*, 31, 1253–1267.
- Ramezanzade, H., Abdoli, B., Farsi, A., & Sanjari, M. A. L. (2014). The effects of sonification modelling on perception and accuracy of performing jump shot basketball. *International Journal of Sports Studies*, 4(11), 1388–1392.
- Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination: common ground and the coupling of eye movements during dialogue. *Psychological Science*, 18, 407–413.
- Riley, M. A., Richardson, M. J., Shockley, K., & Ramenzoni, V. C. (2011). Interpersonal synergies. *Frontiers in Psychology*, 2, 38.
- Rosenblum, L. D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science*, 17(6), 405–409.
- Rosenblum, L. D., Yakel, D. A., Baseer, N., Panchal, A., Nodarse, B. C., & Niehus, R. P. (2002). Visual speech information for face recognition. *Perception & Psychophysics*, 64(2), 220–229.
- Schaffer, D. M., & Mcbeath, M. K. (2002). Baseball outfielders maintain a linear optical trajectory when tracking uncatchable fly balls. *Journal of Experimental Psychology: Human Perception and Performance*, 28(2), 335–348.
- Schaffert, N., & Mattes, K. (2015). Effects of acoustic feedback training in elite-standard Para-Rowing. *Journal of Sports Sciences*, 33(4), 411–418.
- Schaffert, N., Mattes, K., Barrass, S., & Effenberg, A. O. (2009). Exploring function and aesthetics in sonification for elite sports. Proceedings from *Second International Conference on Music Communication*, Sydney, Australia.
- Shockley, K., Richardson, D. C., & Dale, R. (2009). Conversation and coordinative structures. *Topics in Cognitive Science*, 1, 305–319.
- Sigrist, R., Rauter, G., Riener, R., & Wolf, P. (2013). Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review. *Psychonomic Bulletin & Review*, 20, 21–53.
- Sloetjes, H., & Wittenburg, P. (2008). Annotation by category – ELAN and ISO DCR. In Proceedings of *The 6th International Conference on Language Resources and Evaluation (LREC 2008)*.
- Sors, F., Murgia, M., Santoro, I., & Agonstini, T. (2015). Audio-based interventions in sport. *The Open Psychology Journal*, 8(suppl. 3, M6), 212–219.
- Steffensen, S. V. (2012). Care and conversing in dialogical systems. *Language Sciences*, 34, 513–531.

- Steffensen, S. V., & Cowley, S.J. (2010). Signifying bodies and health: A non-local aftermath. In, S. J. Cowley, S. V. Steffensen, J. C. Major and A. Dinis (Eds.), *Signifying bodies: Biosemiosis, interaction and health* (pp. 331–356). Braga: Portuguese Catholic University Press.
- Steffensen, S. V., & Pedersen, S. B. (2014). Temporal dynamics in human interaction. *Cybernetics & Human Knowing*, 21, 80–97.
- Stepp, N., & Turvey, M. T. (2010). On strong anticipation. *Cognitive Systems Research*, 11, 148–164.
- Stienstra, J., Overbeeke, K., & Wensveen, S. (2011). Embodying complexity through movement sonification: Case study on empowering the speed-skater. Proceedings of *The 9th ACM Italian Chapter International Conference on Computer-Human Interaction*.
- Streeck, J. (2009). *Gesturecraft: The manufacture of meaning*. Amsterdam: Benjamins.
- Streeck, J. (2010). *New adventures in language and interaction*. Amsterdam: Benjamins.
- Streeck, J., Goodwin, C., & LeBaron, C. (2011). *Embodied interaction: Language and body in the material world*. Cambridge: Cambridge University Press.
- Thibault, P. J. (2011). First-order languaging dynamics and second-order language: The distributed language view. *Ecological Psychology*, 23(3), 210–245.
- Thibault, P. J., & King, M. E. (2016). Interactivity, values, and the microgenesis of learning in a tertiary setting. In C. C. Ng, R. Fox, & M. Nakano (Eds.), *Reforming learning and teaching in asia-pacific universities* (pp. 173–211). Springer Sciences and Business Media Singapore.
- Tilsen, S. (2009). Multiscale dynamical interactions between speech rhythm and gesture. *Cognitive Science*, 33, 839–879.
- Tilsen, S., & Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. *Journal of the Acoustical Society of America*, 134(1), 628–639.
- Trasmundi, S. B. (2016) Distribueret kognition og distribueret sprog: analyse af kognitive events i en akutmedicinsk social praksis. *Ny Kognitiv Sprogforskning*, 50, 55–85.
- Trasmundi, S. B., & Steffensen, S. V. (2016). Meaning emergence in the ecology of dialogical systems. *Psychology of Language and Communication*, 20(2).
- Trasmundi, S.B. (forthcoming). The educational value of vocalisations and divergence in multimodal dance choreography. *New Ideas in Psychology: Learning as Interactivity, Growth and Becoming: Explorations of Learners, Teachers and Teaching and Learning Environments*.

- Travis, C. (2008). *Occasion-sensitivity*. Oxford: Oxford University Press.
- Turvey, M. T. (1990). Coordination. *American Psychologist*, 45(8), 938–953.
- Turvey, M. T. (2007). Action and perception at the level of synergies. *Human Movement Science*, 26, 657–697.
- Uryu, M., Steffensen, S. V., & Kramsch, C. (2013). The ecology of intercultural interaction: timescales, temporal ranges and identity dynamics. *Language Sciences*, 41, 41–59.
- van den Herik, J. (2017). Linguistic know-how and the orders of language. *Language Sciences*, 61, 17–27.
- Warren, W. H. (2006). The dynamics of perception and action. *Psychological Review*, 113, 358–389.
- Wilson, A. D., & Golonka, S. (2013). Embodied cognition is not what you think it is. *Frontiers in Psychology*, 4, 58. doi:10.3389/fpsyg.2013.00058
- Wilson, A. D., Weightman, A., Bingham, G., & Zhu, Q. (2016). Using task dynamics to quantify the affordances of throwing for long distance and accuracy. *Journal of Experimental Psychology: Human Perception and Performance*, 42(7), 965.