Harmony in Motion: Real-time Sonification Strategies for Joint Action Research

Linus Backström[1] **and Luke Ring**[1]

1 Aarhus University

Abstract

Placeholder: For any time-sensitive task making use of auditory feedback, having low latency is vital to ensure the sonification feels connected to the action, rather than disjointed…

Programme
BSc Cognitive Science

Course
Bachelor's Project (147201E020)

Supervisor
Anna Zamm, Assistant Professor

Faculty
Faculty of Arts
Aarhus University

Submitted: 15 February 2023

Student Details

- Linus Backström
  ID: 202004875
  Initials: LB
- Luke Ring
  ID: 202009983
  Initials: LR

Software

- Repository

License
Authors of papers retain copyright and release the work under a MIT Licence (MIT).

Contents

1 Harmony in Motion: Real-time Sonification Strategies for Joint Action Research

Joint action tasks form an integral part of everyday life for humans (van der Wel et al., 2021) and other species (Ferrari-Toniolo et al., 2019). Examples include games and sports, such as football, where it is vital to work with other team members to outplay opponents; construction work, where people may be holding a wall panel up while others fix it to a frame; and music and dancing, where pairs of people may interact in elaborate ways to the rhythm of a song, creating a joint performance from their individual movements. The mechanisms underlying this cooperative ability to work together towards a common goal are of particular interest for research in cognition, creativity, and learning. A huge part of humanity's progress can be attributed to joint action, which has allowed us to build our modern society with all of its infrastructure and technological advancements. An essential part of successful human cooperation is made up of our unique ability of speech, and many types of cooperation involve perception and production of sound as a key aspect. Auditory perception, or simply hearing, in humans refers to our ability to perceive changes in air pressure as sound by detecting vibrations with our ears and interpreting them using our brains. The topic is particularly interesting for cognitive science because of the large amount of perceptual-cognitive processing that occurs from the moment our ears pick up vibrations to when a perception arises.

Sound as a key aspect in cooperation usually belongs to one of two categories: either as the focus of the task, as is the case for musicians in a band, or as a component that can be leveraged for increasing situational awareness or synchronization, for example with a steady beat that members of military corps lock step to. This study investigates the relation between joint actions and sounds by utilizing sonification as a way to facilitate monitoring of individual and joint outcomes during joint action. These two primary concepts of the current paper – joint action and sonification – are introduced and briefly defined here, while a more in-depth discussion of each concept, terminology around them, and previous research into them, follows in the Background-section. We refer to joint action as any situation where two or more people synchronize their actions in pursuit of a shared goal (Knoblich et al., 2011). Sonification is defined as "the use of nonspeech audio to convey information" (Kramer et al., 1999, p. 4).

Previous research indicates two basic features of auditory perception that provide good arguments for representing data as sound (Kramer et al., 1999). First, auditory perception is especially useful for detecting temporal characteristics, i.e. variations in sound over time (Hildebrandt, 2014). Sonification can thus be useful for monitoring or understanding complex temporal data. Second, our sense of hearing does not require us to be oriented towards the sound source. Unlike visual perception, which allows us to perceive approximately 180 degrees of our environment in front of us while we remain blind to the other 180 degrees behind us, auditory perception allows perception of 360 degrees. This makes auditory signals particularly useful for situations where our visual system is occupied with another task and we cannot afford to look around constantly, such as surveillance and alarm applications. Other benefits of auditory perception that speak for sonification are parallel listening (the ability to monitor and process multiple audio sources), affective response (increased learning and engagement), and finally rapid detection – humans are able to react faster to sound than to any other type of stimulus, achieving reaction times of around 160 ms in simple reaction time experiments (Kosinski, 2008; Kramer et al., 1999).

This study explores how the learning of a novel joint action task is affected by different methods of sonification. When performing joint actions an individual can either focus on themselves and their partner as separate entities, which we refer to as self-other representations, or instead focus primarily on the effect that their combined actions have, which we call joint outcome representations. The purpose of the current study is to investigate whether learning and synchronization during joint action can be optimized by enhancing attention towards one of these representations using sonification. Movement sonification can be used to facilitate synchronization by providing auditory feedback of actions, allowing individuals to adjust their movements in real-time to achieve a more synchronized state. More specifically, the use of sonification can help individuals to better perceive and coordinate their movements, leading to improved joint performance and increased levels of synchrony (Dotov & Froese, 2018). Our research question is thus the following:

> Is synchrony optimized when focusing on self-other representations or joint outcome representations?

When using auditory feedback in a joint action context, latency is particularly important due to the fact that there is a relatively small window where an event and a related sound are perceived as synchronous. Although studies report varying results (Keetels & Vroomen, 2012), asynchrony is detectable at as little as 6 ms, and more likely around 30 ms for continuous movement (McPherson et al., 2016), meaning any pipeline with a higher latency is likely to introduce confounding variables in measurements. Only a relatively limited number of studies have investigated

the effects of sonification on joint action[1], and this thesis aims to expand the current body of research by presenting a flexible low latency sonification framework that uses real-time positional data for joint action research. To this end, the present study implements a novel method for sonifying joint actions in a pilot study investigating how different representations affect synchronization. By comparing subject synchronization during an experiment employing self-other represented (task-oriented) or joint outcome represented (synchronization-oriented) strategies, we attempt to show differences that highlight the importance of selecting appropriate mapping patterns for sonification, and provide a pathway for further investigation.

2 Background

2.1 Sonification

The current study investigates whether sonification can be used to optimize synchronization during joint action by enhancing attention towards either self-other or joint outcome representations. Sonification is defined as "the use of nonspeech audio to convey information" (Kramer et al., 1999, p. 4). 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., 1999, p. 4). According to Dubus (2013) sonification is the use of sound to communicate, interpret and perceive data. Sonification is especially suitable for tasks with time constraints, such as monitoring and synchronizing (Dubus, 2013). Sonification can also be characterized as a segment of augmented reality that reveals information otherwise hidden, with the help of sound (Dubus, 2013). According to Dubus (2013) that is done through clear connections between data dimensions and auditory dimensions of the sonification display. The layperson may confuse sonification with music, but according to Dubus (2013) there is a clear difference between the two: sonification is meant to communicate objective data, and music is instead often used to communicate more subjective things, such as emotions. Nevertheless, the differences between music and sonification are not fully agreed upon among researchers, and thus there is no clear consensus on the distinction in the academic discourse (Dubus, 2013). As mentioned above, sonification is a relatively young field of research. Sonification studies are plagued by a lack of consistency in terminology and the arbitrary nature of sonification mappings (Dubus, 2013; Dubus & Bresin, 2013).

[1] A Google Scholar search (14 February 2023) for '+"joint action" +sonification' only yielded 193 results, compared to over 326,000 results for '+"joint action"' alone

Although concepts around sonification and audification were not formalized until around the year 1992, when the first International Conference on Auditory Display (ICAD) was held (Dubus & Bresin, 2011), practical examples of sonification can be found throughout history (Dubus, 2013). Water clocks in ancient Greece and medieval China were sometimes constructed to produce sounds and thereby provide auditory information about the passage of time (Dubus & Bresin, 2011). The stethoscope, which is used for listening to sounds made by the heart and lungs as well as other internal sounds of the body, was invented in 1816 by French physician and amateur musician Rene Laënnec (Roguin, 2006). The Geiger counter developed in 1928 provides perhaps the most characteristic example of sonification through its function of sonifying levels of radiation. The device detects ionizing radiation and translates it into audible clicks, where a faster tempo signifies a higher level of radiation (Figure 1). Dubus and Bresin (2011) describe the value of the Geiger counter as "transposing a physical quantity which is essentially non-visual and pictured in everyone's imagination as very important because life-threatening, to the auditory modality through clicks with a varying pulse" (Dubus & Bresin, 2011, p. 1).

Figure 1: Photograph depicting a Geiger counter being used to detect levels of radiation. Geiger counters use sonification to represent radioactivity by producing audible clicks that increase in frequency as the level of measured ionizing radiation increases (Dobson, 1963).

2.1.1 Monitoring via auditory feedback

In a review of mapping strategies for sonification Dubus and Bresin (2013) identify a number of applications for sonification, including monitoring, motion perception, data exploration, accessibility, art and aesthetics, the study of psychoacoustics, and as a complement to visualization. Debashi and Vickers (2018) specify that sonification is a particularly useful tool for conveying

the type of information that changes over time. Kramer et al. (1999) point out that sonification can allow the user to make sense of large amounts of data by utilizing modern powerful media technologies. Out of the various applications for sonification, monitoring of external information is the most relevant for the current study.

Using sonification for external monitoring can for example mean that there is a sound that the user listens to while simultaneously working on something else, such as when medical staff in operating rooms rely on auditory cues from their equipment to monitor the patient's vital signs (Dubus, 2013). In such instances, a change to the monitored state causes a corresponding change in the sound, allowing the user to quickly become aware of the change and react as needed. One of the clear advantages with using sonification for external monitoring is then that the user is free to work on a different task than the monitoring while still maintaining the ability to detect and react to changes (Vickers, 2011).

Compared to visualization, sonification can have certain advantages that make it suitable as a complement or replacement to visualization. This can be observed in practice in the health sector, where real-time sonification using parameter mapping methods is used; one study identified a high potential and found positive results for the use of real-time auditory feedback-oriented training devices in physical rehabilitation and fitness training to increase awareness of physiological responses (Yang & Hunt, 2015). The fact that humans are very sensitive to changes in rhythm or sequences of sounds lends further support to the idea of complementing visualizations with sonification (Hildebrandt, 2014). A recent study by Debashi and Vickers (2018) comparing sonification and visual methods of monitoring found that the visual method alone performed significantly worse than a combination of both, and further that using sonification resulted in reduced visual fatigue rates. In summary, the scientific literature clearly indicates that sonification has an important part to play in the context of monitoring external information, and furthermore its use cases extend to several different sectors and should be researched further.

2.1.2 Movement sonification

As previously mentioned, sonification involves the transformation of all types of data into sound (Kramer et al., 1999). The term movement sonification specifically refers to the transformation of movement – typically that of a human – into sound (Vinken et al., 2013). A. Effenberg (2005) states that perception and reproduction accuracy of gross motor patterns can be improved with the help of movement sonification, indicating a wide range of potential applications for artificial auditory movement information in sports and rehabilitation. Based on the idea that perceiving gross motor patterns is facilitated when more senses are active, Sport Scientist in particular have tried to take advantage of this effect by creating and conveying an increased amount of auditory movement information [refs]. In order to achieve multisensory integration benefits, it is vital that the additional auditory movement information corresponds to the structure of the perceptual features of another modality (visual, kinesthetic, or tactile) (Schmitz & Effenberg, 2017). When visual motion perception is the reference with which bi- or multimodal convergence is to be achieved, movement sonification needs to be based on kinematic parameters (Schmitz & Effenberg, 2017). These kinematic parameters refer to the spatiotemporal features of a movement pattern or pose. This acoustic enhancement of motor perception became known as "movement sonification" when A. Effenberg (2005) took the sonification approach of the early 1990s and adapted it to the kinematics and dynamics of human motor actions.

In the empirical section of the current study, we describe how we used movement sonification to emphasize different joint action strategies and manipulate synchronization during a joint action task. The use of movement sonification in joint action research is supported by the finding that movement sonification enhances perception of movement and improves motor performance (Schmitz et al., 2013). Other studies in Sport Science have found that when movements are mapped onto sound, i.e. sonified, predictions can be facilitated (A. Effenberg, 2005; Schmitz & Effenberg, 2012). Movement sonification may also support synchronization in joint action by addressing central motor representations, more precisely by making the movements of athletes more predictable to their teammates (Schmitz & Effenberg, 2012). Furthermore, sonification is well suited to support applications for physical training, as seen in a study by Dubus (2012) where professional rowers were able to use kinetic and kinematic cues to optimize their rowing speed. The author concluded that rowing performance could be improved with the help of interactive augmented feedback (Dubus, 2012). Finally, Schmitz and Effenberg (2017) found that complementing visualizations of a swimmer with kinematic sonification allowed for more accurate perceptions of differences in swimming stroke frequency.

With sonification being such a recent field of research, its subfield movement sonification has had even less time to be researched (Vinken et al., 2013). As such, the question of how to map movement parameters onto sound in an optimal way remains uncertain due to a lack of an adequate theoretical background (A. O. Effenberg & Schmitz, 2018). With this uncertainty in mind, A. O. Effenberg and Schmitz (2018) suggest that movement sonification can function as an accessible form of information similar to visual information when coded properly. Along the

same lines, Vinken et al. (2013) state that movement sonification can improve motor processes, as well as adding information to parts of movements that are typically silent. By contrast, Vinken et al. (2013) also explain that despite these potential use cases there is hardly any empirical data from scientific research that clarifies how to sonify gross motor human movement in order to achieve information rich sound sequences. For these reasons it is important to gather more data about both sonification in general, as well as movement sonification specifically.

Vinken et al. (2013) identify three main areas for movement sonification that are lacking in empirical proof: the selection of appropriate movement features, the optimal mapping patterns between kinetic and acoustic features, and the appropriate number of dimensions for sonification. The current study adds to the existing body of research in movement sonification by implementing a flexible low latency sonification pipeline and describing our strategy selection based on movement and acoustic features, how they were mapped, and which dimensions were used. In addition to the aforementioned contribution to movement sonification research, this study also furthers adds to the literature by investigating sonification in the context of learning a novel joint action task and improving performance as measured by synchronization. Music belongs to the relatively small set of joint action behaviors that support a high degree of coordination, because of how suitable our auditory system is for temporal coordination (Hildebrandt, 2014). This raises the question of whether we can optimize temporal coordination by using movement sonification, or said differently: can sonification help with joint action? The existing body of research falls short on this question, and the present study aims to open a discussion with the help of a pilot experiment conducted at Aarhus University. In the next section we will present the theory behind joint action with a focus on representations, action monitoring, and action prediction. Then, before presenting our pilot experiment, we will further discuss how and why the two concepts of sonification and joint action are integrated.

2.2 Joint action

Joint actions, where two or more people synchronize their actions in pursuit of a shared goal (Knoblich et al., 2011), are a regular part of human behavior. A longer definition refers to joint action as "any form of social interaction whereby two or more individuals coordinate their actions in space and time to bring about a change in the environment" (Sebanz et al., 2006, p. 1). With these definitions in mind, examples of joint action can include an extremely wide range of activities, such as handshakes, conversations, musical performances and partner dances, but also bank robberies and the building of the pyramids. To avoid overwhelming the reader, a typical example that is found in the literature (Sebanz et al., 2006) and that fits both definitions of joint action is when two people work together to carry a table from point A to point B.

As mentioned in the introduction, the progress of human civilization is largely based on working together. Joint actions constitute a significant part of the human experience, and they form an important and intriguing research topic for the field of cognitive science. Studies of joint actions such as putting together furniture or playing a piano duet, for instance, have shed light on how speech is used to establish who will do what and to agree on the details of the joint performance (Clark, 2005). Additionally, research on how people solve problems of spatial coordination has shown that humans are capable of creating new symbol systems to coordinate their actions when conventional communication is not available (Galantucci, 2009).

Joint actions can be divided into two categories: emergent and planned coordination (Knoblich et al., 2011). Emergent coordination describes coordinated behavior arising from perception-action connections that lead to similar actions among individuals, independently of prior planning; an example of this is when pedestrians end up walking in step with each other without explicitly planning to do so (Knoblich et al., 2011). With regards to planned coordination, the behavior of agents is driven by representations that describe the desired outcomes of joint action and the respective role of the agent in achieving these outcomes (Knoblich et al., 2011). Next we will describe the cognitive processes that are involved in the more relevant, planned coordination, category of joint action. In addition to representations, recent theory identifies action monitoring and action prediction as the other main cognitive processes involved in planned joint action (Loehr et al., 2013; Sebanz et al., 2006; Vesper et al., 2010).

2.2.1 Representations

According to the minimal architecture for joint action proposed by Vesper et al. (2010), an agent involved in joint action must, at a minimum, have a representation of their own task and the shared goal. An assumption is also that the shared goal cannot be achieved without the contribution of both parties (Vesper et al., 2010). In the model developed by Vesper et al. (2010), the shared goal is expressed as "ME + X", where "ME" stands for the agent's own contribution, and "X" stands for the contribution that is not produced by the agent themselves. A study by Loehr and Vesper (2016), which had piano novices practice a duet with a more experienced pianist, found that the novices' representations consisted of the duet participants' shared goal, and to a small extent of the novices' own personal goal.

A minimal version of including the other in one's representations is theorized to be the understanding that the source of "X" – that which is not produced by an agent themselves – is the joint action partner (Loehr & Vesper, 2016; Vesper et al., 2010) . Although not required, it is often helpful to also represent the other's task, as it allows for more precise predictions of what the other will do next (Bolt & Loehr, 2021; Wenke et al., 2011). As an example, consider two singers performing a duet together. Each singer must fully know their own part, while also representing the shared goal of synchronized singing. Although these two main representations can be sufficient for performing a duet, professional singers typically familiarize themselves with their singing partner's part in addition to their own, as it allows for a more polished and cohesive musical performance. The benefits of representing the other's task were demonstrated by a study (Keller et al., 2007) in which pianists were asked to record one part from a selection of piano duets, and then play the complementary part in synchrony with either their own or other participants' recordings. The results showed that the pianists synchronized better with recordings of themselves than those of others, indicating that synchronization is facilitated by having a more precise representation of the auditory stimuli with which one is coordinating actions. Further insight into the role of representations in joint action comes from an EEG study by Kourtis et al. (2012), which found that partners represented each other's actions in advance when passing an object, and doing so facilitated coordination. Having these shared representations of actions and their underlying goals allows individuals to establish a procedural common ground for joint action without needing to rely on symbolic communication (Sebanz et al., 2006).

Two music-related studies have explored both shared and individual goals. First, Keller and Burnham (2005) demonstrated that musicians performing duets can attend to and recall both their own part and a combination of their own and a complementary part. More recently, Loehr et al. (2013) reported that duetting pianists prioritize shared goals (the musical harmony arising from both pianists' combined pitches) over individual action goals (the individual pitches played by each pianist), as demonstrated by stronger neural responses to pitch errors that impact the former compared to the latter. Based on their research, Loehr and Vesper (2016) argue that shared goals are more salient compared to individual goals in novel joint actions performed by non-experts.

Empirical research of representations has largely focused on how people represent each individual's contributions to the joint action (Knoblich et al., 2011; Loehr & Vesper, 2016). The details of *how* people represent shared goals remain mostly unclear, however (Loehr & Vesper, 2016). Findings by Loehr and Vesper (2016) indicate that novices in joint action contexts that promote minimal representations represent their actions in relation to the shared goal, which supports the argument that joint action participants represent the shared goal of the task (Vesper et al., 2010). Still, researchers highlight the need for further research, for instance by pointing out the lack of joint action studies teasing apart representations of shared and individual goals (Loehr & Vesper, 2016). The present study thus aims to fill this gap in joint action research by addressing both representations in the form of self-other- and joint outcome -strategies as separate experimental conditions.

2.2.2 Action monitoring

Another cognitive process involved in joint action is known as action monitoring, or simply monitoring. Representations, specifically shared task representations, are intrinsically linked with both monitoring and predicting processes, with all of them working together to enable interpersonal coordination in real time (Knoblich et al., 2011). Knoblich et al. (2011) describe this interplay of cognitive processes by stating that shared task representations determine how agents monitor and plan their actions. A simple way to consider this is that in order to effectively monitor an action, a basic idea of what it should resemble – i.e., a task representation – is required.

Monitoring processes are used to assess the extent to which a task or goal is being accomplished and whether actions are proceeding as intended (Botvinick et al., 2001). In terms of assessing task and goal progress, three things can be monitored: the agent's own task, the other's task, and the shared goal. The agent must at least monitor the progress of their own task and the shared goal. It is not strictly necessary to monitor the other's task, and it depends on the type of joint action that is performed. For example, consider a very simple task such as lifting an object straight up in the air together with a partner. It is entirely possible to do so successfully even if both agents only monitor their own task ("lift this side of the object") and the shared goal ("lift this object together"). Nevertheless, it is likely true that monitoring what one's partner is doing will improve joint action performance – especially for tasks that require precise synchronization (Vesper et al., 2010).

With respect to monitoring the sensory consequences or outcomes of joint actions, a distinction can be made between monitoring the individual outcomes vs joint outcomes. A study by Loehr et al. (2013) distinguished between individual and joint outcomes of actions with the help of a clever experiment, where experienced pianists played a pre-rehearsed duet on a digital piano while the outcomes of certain keypresses were manipulated by the researchers. In the individual outcome condition, the pro-

duced tones of keypresses were manipulated so that the harmony of the resulting chord remained the same. In the joint outcome condition, the produced tones were manipulated so that the harmony of the chord changed. The researchers found that the musicians in their study were able to monitor both individual and joint outcomes, while maintaining a distinction between the two. Furthermore, the musicians were able to monitor the outcomes of both their own and their partner's actions in parallel, while also differentiating between the two (Loehr et al., 2013). To summarize, it appears that agents involved in joint action are able to represent and monitor their own and their partners' actions, as well as the joint outcome of their actions. Nevertheless, research into how individual vs joint outcomes in joint action are monitored is extremely scarce, and the present study sheds light on this particular issue by conducting an experiment where individual and joint outcomes are sonified separately.

2.2.3 Action prediction

The crucial final feature of joint action relates to the manner in which individuals adapt their own actions to those of others in time and space, and doing so requires making predictions of the other's actions. In order to avoid constantly being one step behind during joint action, interacting partners cannot simply respond to observed actions, but must rather plan their own actions in relation to what they predict their partner will do (Sebanz et al., 2006). This prediction process is achieved through motor simulation, which uses internal models to determine the sensory consequences of actions as well as their effect on the environment (Schmitz & Effenberg, 2017; Vesper et al., 2010). Simulating the actions of others as they occur may be especially beneficial when engaging in joint action, and it has been suggested that such motor simulation influences perception and assists in predicting the consequences and timing of others' actions (Vesper et al., 2010). The idea that internal predictive models contribute to the ability to anticipate others' actions is supported by findings that short-term predictions of others' actions are based on one's own motor experience (Aglioti et al., 2008; Calvo-Merino et al., 2005). The data from Loehr and Vesper (2016) complement previous joint action research by strengthening the notion that agents can predict the consequences of others' actions in parallel with their own (Loehr et al., 2013; van der Steen & Keller, 2013; Vesper et al., 2014; Wolpert et al., 2003) and also incorporate these predictions of other's actions when planning and executing their own actions (Knoblich & Jordan, 2003; Kourtis et al., 2012; Loehr & Palmer, 2011; Vesper et al., 2013).

It is not fully clear yet whether similar mechanisms as those mentioned above exist specifically for predicting the joint outcome of an agent's and their partner's actions. Some support for predicting joint outcomes comes from a study by Knoblich and Jordan (2003), which demonstrated the ability to predict combined outcomes through improved joint task performance with practice. The results showed that participants initially struggled with the joint task of controlling a cursor together to track a moving target on a computer screen, but with practice, performance reached the level of individual performance. Furthermore, participants who were provided with an external cue tone about the state of their partner's action were more successful at the task, indicating that auditory feedback can facilitate coordination (Knoblich & Jordan, 2003). This is particularly interesting in the context of the present study, as it suggests a potential benefit with using sonification in the context of joint action.

2.3 Integrating sonification and joint action

In this section, we will focus on how the concepts of sonification and joint action relate to each other. To briefly restate what has been previously discussed, sonification is defined as the transformation of data into sound (Kramer et al., 1999), and joint action refers to situations where two or more people synchronize their actions to achieve a shared goal (Knoblich et al., 2011). The cognitive processes related to joint action include representation, monitoring and prediction (Loehr et al., 2013; Sebanz et al., 2006; Vesper et al., 2010). We will now discuss how sonification can make use of these three different processes in the light of previous research.

The first cognitive process in joint action to be considered is representation. There is no clear consensus in the academic literature on the details of representations for an agent involved in joint action, but previous research indicates that the agent must, at the very least, represent their own task and the shared goal (Vesper et al., 2010). There are also several studies supporting the idea that representing the other's task can be beneficial for joint action by making prediction and synchronization easier (Bolt & Loehr, 2021; Keller et al., 2007; Kourtis et al., 2012; Sebanz et al., 2006; Wenke et al., 2011). Loehr and Vesper (2016) point out the need for future research to investigate which factors influence whether or not an agent represents their partner as the other source contributing to the shared goal, and how those representations of their partner may change while learning a joint action. In their study, Loehr and Vesper (2016) found that novices have the ability to integrate the auditory effects of their partner's actions into their sensorimotor action representations while learning to play musical pieces together. The notion that such an ability is not limited only to experts is pertinent to the current study because it allows for the exploration of learning novel joint action tasks using sonification. Specifically, sonification can be used to sonify an

agent's own actions, their partner's actions, or the joint outcome of both participants. This allows us to direct participants' attention towards specific types of representations, namely self-other and joint outcome representations.

The next cognitive process to discuss in the context of joint action and sonification is monitoring. As stated earlier, a popular and useful application for sonification is the monitoring of external information (Dubus & Bresin, 2013). Designing a sonification system for monitoring purposes requires careful consideration of various conditions and requirements. The sound must be capable of supporting extended periods of listening, changes in status have to be salient, and unexpected events have to be immediately apparent (Kimoto & Ohno, 2002). In joint action, monitoring one's own actions and the progress towards the shared goal is crucial for success, and the other's actions must also be monitored when precise synchronization is required (Vesper et al., 2010). When discussing monitoring in joint actions, it is important to take into account the divided nature of joint action and identify the challenges that come with it. One of the main challenges arises from the fact that joint actions often require simultaneous actions by the participants, which may create the need for agents to monitor both their own and their partner's actions in parallel (Loehr et al., 2013). A closely related challenge is that monitoring an action, whether one's own or the other's, is dependent on having a representation of that action (Knoblich et al., 2011). The other main challenge relates to the fact that joint action outcomes are often more than the sum of individual action outcomes (Loehr et al., 2013). An example of this is how the same tones played by one musician can take on different qualities and become part of different harmonies, depending on what tones another musician is simultaneously playing (Loehr et al., 2013). This leads to the consideration of whether agents monitor their own or their partner's actions in relation to individual action goals (those required to achieve each individual's own task) or in relation to shared action goals (the joint outcome of their actions) (Loehr et al., 2013). In the current study, we attempt to address this challenge of individual vs joint outcomes by creating two different sonification schemes, where one scheme sonifies the individual action outcomes, and the other scheme sonifies the joint outcome of both participants' combined actions. We can therefore use sonification to encourage and facilitate monitoring of either individual (self-other) or joint outcomes. We theorize that the previously identified benefits of sonification in monitoring, such as the ability to work on another task while monitoring with one's ears (Vickers, 2011) and the human auditory system's sensitivity to changes in sequences of sound (Hildebrandt, 2014), should improve performance of related joint action tasks by reducing the cognitive load required for monitoring. This reduction in cognitive load would then allow joint action participants to also focus on other points of interests that support progress towards the shared goal, such as fine motor control and planning their next actions. For these reasons, we postulate that sonification holds substantial promise for both practical applications and academic research of joint action monitoring.

The third and final cognitive process involved with joint action is prediction. Previous research in the field of Sport Science has revealed that sonification can improve perception accuracy of movements (A. Effenberg, 2005; Schmitz et al., 2013), revealing one potential mechanism by which predictions in joint action may be facilitated using sonification. As predictions play an important role in joint action (Sebanz et al., 2006), especially for tasks that require a high degree of synchronization (Vesper et al., 2010), we suggest that joint action performance can be improved by facilitating action prediction with the use of sonification. This is substantiated by the findings of Knoblich and Jordan (2003), which revealed that joint action performance in a task requiring participants to predict joint outcomes improved when participants were provided with an external cue tone relating to their partner's actions. Further research needs to be conducted in order to determine whether auditory cues using sonification can facilitate prediction in other joint action contexts, and the present study aims to contribute to this body of research by using a joint action task that emulates the need for a high degree of synchronization.

Some of the questions that have been investigated in recent joint action research concern the aforementioned cognitive processes (representations, action monitoring and action predicting) and how they relate to agency (self vs other) and outcome (individual vs joint). Researchers have studied whether agents involved in joint action represent both their own task and their partner's task (Loehr & Vesper, 2016), whether they monitor individual outcomes or joint outcomes (Loehr et al., 2013), and how predictions of other's actions are incorporated when planning and executing actions (Knoblich & Jordan, 2003; Kourtis et al., 2012; Loehr & Palmer, 2011; Vesper et al., 2013). Based on the literature, a common denominator between sonification and joint action appears to be synchronization. For this reason, we identify a potential application for sonification particularly in joint action tasks that require a high degree of synchronization. Previous research has found that sonification can improve synchronization in joint action by addressing central motor representations (Schmitz & Effenberg, 2012) and monitoring (Vesper et al., 2010). More generally, several researchers have argued that having more precise representations may be key to improving synchronization in joint action (Bolt & Loehr, 2021; Keller et al., 2007; Kourtis et al., 2012; Sebanz et al., 2006;

Wenke et al., 2011). Sonification also appears to be useful for sensorimotor learning by providing auditory feedback of movements (Bevilacqua et al., 2016) and in the present study we address this aspect by giving subjects a novel joint action task with varying methods of auditory feedback.

In summary, this study adds to the discussion about strategies relating to joint action representations, namely self-other and joint outcome representations, by investigating their potential effect on synchronization. Participants in our study performed a novel joint action task under three different conditions – individual outcome, joint outcome, and a control condition – where the sensory consequences were manipulated through real-time sonification of movement. The sonification was used to prime the participants' attention towards either individual or joint, while joint task synchronization was recorded.

2.3.1 Research question

Is synchrony optimized when focusing on self-other representations or joint outcome representations?

3 Low Latency Motion Capture Sonification Validation Experiment

A pilot experiment was conducted to assess the viability of the sonification framework in a laboratory setting. This experiment required blindfolded subjects to move their assigned sleds along parallel tracks and use sounds they hear to remain as spatially synchronized on their tracks as possible.

3.1 Participants

An availability sample of ten subjects (age range 20-29 years; 5 female, 4 male, 1 gender-fluid; 7 right-handed, 2 left-handed, 1 ambidextrous) were recruited to participate in pairs. Subjects optionally reported basic demographic information regarding age range (intervals of 10, i.e. 10-19, 20-29, ..., 90-99), gender, handedness, years of formal music training and reported if they were known to be tone-deaf (6 not tone-deaf, 4 unknown). Subjects reported a mean of 2.4 years of formal music training (SD=4.2; min=0.0; max = 12.0 years). Due to this experiment being a pilot, five subject pairs were regarded as sufficient to validate the experimental setup as well as gather preliminary data on movement synchronization for the three conditions. Additionally, the number of possible participants was constrained by limited access to the motion capture system laboratory, which is shared by other researchers, meaning that subjects needed to be available at the scheduled lab times within the study timeframe.

3.2 Track and Sleds

Two parallel tracks were designed with a sigmoid curve shape, surfaced with a smooth veneer that allowed for free movement along the length of the track. Two identical sleds were constructed from LEGO parts, and three felt adhesive pads were attached to the underside of each sled to reduce resistance during movement. The sleds were coated with a matte black paint to limit near-infrared reflectivity (Benedict et al., 2016), as prior tests with unpainted LEGO bricks introduced artifacts into the motion capture system that were incorrectly identified as markers.

3.3 Frequency Range Selection

Two distinct, continuous frequency ranges were selected for application in the sonification conditions. These ranges are offset by a perfect fifth and span eight semitones. The *overtone* (the tone with the higher frequency) range was chosen based on a center frequency of 220 Hz (A3) and the range was limited to avoid a large overlap with the *undertone* (the tone with the lower frequency) range during normal operation (Table 1). Consideration was also given to creating ranges that were not sufficiently high to cause discomfort at a consistent amplitude. Sethares (2005) shows that the perceived dissonance between two tones varies by the lower tone's frequency, indicating that the selected undertone range of approximately 116 – 227 Hz would result in an increased perceived dissonance as the interval distance decreased (see Figure 2). The selected ranges were tested during various simulation trials and were harmonious sounding when they were at the perfect fifth interval, and were dissonant as the tones deviated from a perfect fifth.

Table 1: Frequency ranges for the two tones used in the sonification conditions.

	Lower Bound		Center		Upper Bound	
	Freq	Note[a]	Freq	Note[a]	Freq	Note[a]
Overtone	174.614	F3	220.000	A3	277.183	C#4
Undertone	116.409	A#2	146.666	D3	184.788	F#3

[a]Note names are in International Pitch Notation, and are the closest approximation to the frequencies used

Note. Overtone frequencies were calculated to have a center frequency of 220Hz, and undertone frequencies are two-thirds of their overtone counterparts.

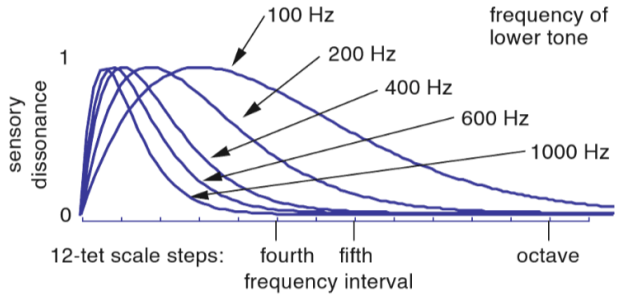

Figure 2: Sensory dissonance of sine waves by interval for five frequencies. Figure (Sethares, 2005, p. 47)

4 Task and Procedure

Participants were asked to sit on opposite sides of the track structure and familiarize themselves with the movement of the sleds along the tracks. They were instructed to continuously move the sleds along the track from end to end, as rapidly as possible while remaining spatially synchronized with their partner's position on their respective track, using sounds they may hear during the various conditions to assist them. Once the participants had given their informed consent and been briefed, they would indicate when they were ready and were blindfolded for the duration of all trials within each condition. The experiment flow control was automated, starting with a practice trial of 30 seconds, then a 15 second break, followed by three experimental trials running for 90 seconds each with 15 second breaks between trials. Three tones played immediately prior to each trial to indicate the start of the trial, and a single sustained tone played at the end of the trials to indicate the completion, after which participants were asked to return their sleds to the start of the track. Sonification and recording were paused between conditions to allow sufficient time for subjects to rest. When the subjects were ready, a hardware button on the Bela was pressed to begin the next condition.

(prioritized comfort, allowed choice of arm despite handedness) [why this many, why not more, etc, ethics] Subjects were given and signed an informed consent form [appendix...] and information sheet [appendix...]... and were under the umbrella project...

4.1 Sonification Strategy Conditions

The experiment consisted of three conditions that employed different sonification mapping strategies, namely: a no sonification control condition, a task-oriented sonification strategy and a synchronization-oriented sonification strategy. Each condition consisted of one practice trial of 30 seconds duration, and three main trials of 90 seconds each. Before each practice trial, subjects were reminded that it was a shorter trial and that they could use it to experiment with the sonification.

4.1.1 No sonification

In the no sonification condition, only the motion capture data from participants' sleds were recorded, and subjects could use the audible sounds of the sleds moving along the track to align themselves with their partner.

4.1.2 Task-oriented sonification strategy

The task-oriented sonification represented the position of each sled along the length of the track as a synthesized tone that varied in frequency from highest to lowest at the start and end of the track respectively. One sled produced a higher frequency overtone, while the other produced a lower frequency undertone. If the sleds were at the exact same x-coordinate, the two tones would be a perfect fifth apart, creating a harmonious interval; if the sleds drifted further apart, the frequency difference would deviate from the perfect fifth and create a more dissonant sound. Figure 3 illustrates the implementation of the task-oriented sonification strategy. This strategy was selected for sonifying the movement along the track, i.e. the task required of subjects.

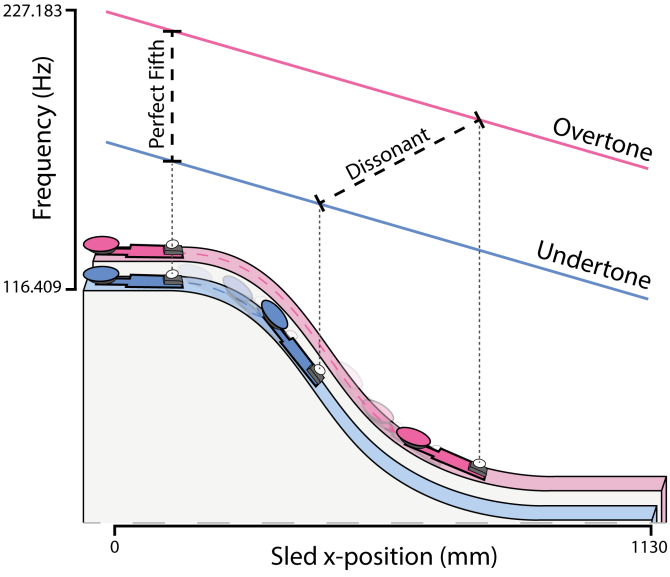

Figure 3: Task

4.1.3 Synchronization-oriented sonification strategy

The sonification strategy oriented around synchronization represented the position of the sleds relative to each other, so that the two sleds at the same x-coordinates would create a harmonious perfect fifth interval. If sleds drifted apart, the overtone amplitude decreased, and the undertone frequency changed based on the distance between the two sleds. Figure 4 illustrates the implementation of the synchronization-oriented sonification strategy.

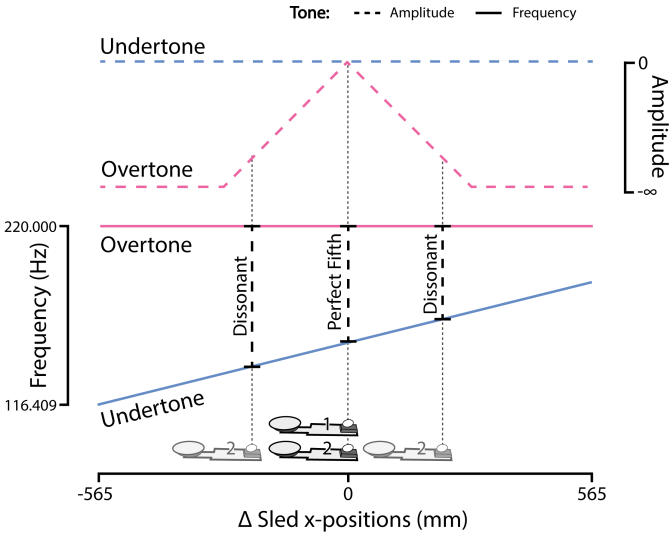

Figure 4: Sync

5 Hardware and Software Implementation

5.1 Motion Capture

Motion capture data were collected using a 9 camera (8 Qualisys Miqus M3 and 1 Qualisys Miqus Video) system connected to a Qualisys Camera Sync Unit. Marker data were acquired at a sampling rate of 300 Hz and video data were acquired at a sampling rate of 25 Hz. Qualisys Track Manager (QTM) software version 2022.2 (build 7700) was used to collect and process the data with real-time 3D tracking data output. QTM options for 'processing of every frame' and '2D data preprocessing' were disabled for real-time output to ensure minimal latency. Figure 5 outlines the flow of data from motion capture to sonification.

5.1.1 Markers

For the experimental setup, one passive marker was placed on each car, and two additional passive reference markers

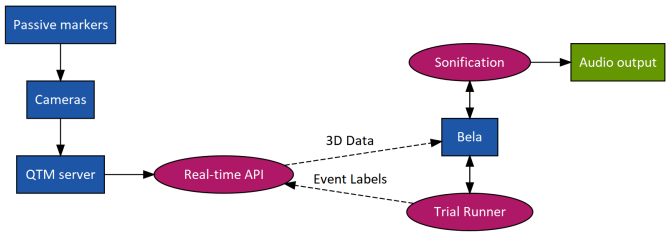

Figure 5: Low-latency sonification pipeline

were placed on the front corners of the track (see Figure 6 for a visual representation of the track and marker placement). These additional markers provided reference points for 3D orientation of the track and the cars across trials in case of accidental track movement. Four preliminary sessions of two minutes were recorded of variable speed sled movements in QTM, and unique labels were given to the four passive markers. The recordings were used for training a QTM Automatic Identification of Marker (AIM) model. AIM models were applied to recordings and real-time output to apply known labels to marker data, allowing the sonification to read the current position of both sleds.

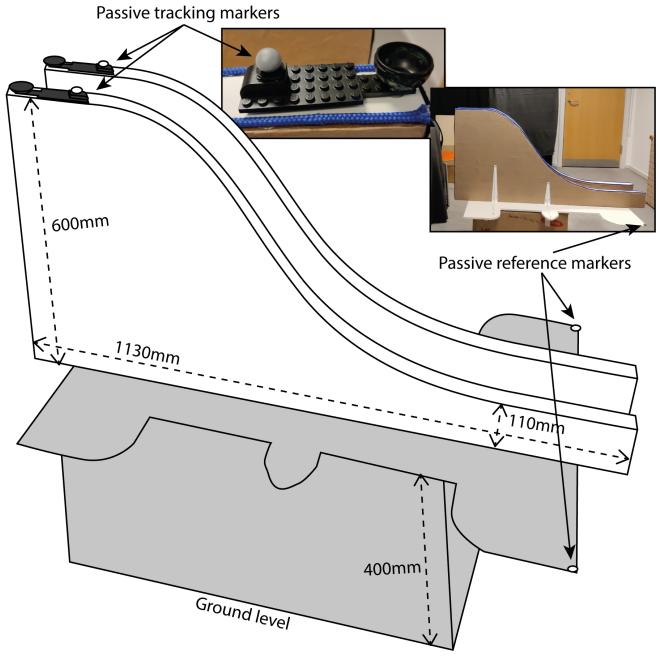

Figure 6: Illustration depicting the experimental track setup and dimensions. Photos depict the constructed track as well as the sleds with the passive motion tracking markers.

5.2 Sonification

Motion capture data were sent via UDP packets over USB networking to a Bela Mini device running version 0.3.8g

running a custom C++ program[2]. The main program loop was configured to execute every 32 samples, with an output sample rate of 4.41×10^4 Hz for 2 audio channels. The two audio output channels were connected to a pair of Genelec G Two active speakers. The main program used the latest available Bela platform framework [3].

Two 16-bit 44.1 kHz wave files were prepared from the output of a MIDI synthesizer at the lowest frequency for both of the tones, and were cut off at 113145 samples where the zero-crossing of both files aligned. A 5 ms fade-in and fade-out was applied to the start and end of the files to minimize DC pop. To allow for dynamic frequency changes, sound file playback used a floating point read pointer which was incremented sequentially as the 32 ms buffer was populated. When the playback frequency was increased, the step size of the read position would increase by change in frequency, and would proportionally interpolate between samples to provide a smooth sounding frequency transition. This method was selected to ensure that the audio resolution was never below the original file's resolution.

5.2.1 Real-time 3D Data

A version of the Qualisys C++ SDK using protocol version 1.23 was modified to be compatible with the Bela platform and was used for communicating with QTM. To reduce latency, connection to the QTM server was made over UDP, and round-trip communication latency was verified by performing 1000 requests to the QTM server and logging the elapsed round-trip time, resulting in a mean latency of 0.25 ms (SD 0.03 ms, min 0.23 ms, max 0.43 ms).

Using the SDK, 3D streaming was initiated at the start of each sonification condition, and labelled markers were used to obtain the current position of each sled. The coordinates of the sleds were stored in a buffer containing the current and last recorded coordinates.

5.3 Experiment

5.3.1 Workflow

The experiment flow control was automated via the main C++ application running on the Bela mini. Before each experiment started, the condition order was configured in the application, and after compilation the suite of conditions and trials would run. Prior to commencement of each condition, the execution of the application halted to allow sufficient time for subjects to rest, after which a hardware button on the Bela could be pressed to continue. After commencement of a condition, all trials for that condition were run consecutively with 15 second breaks between them.

5.3.2 Event Labels

From the main Bela application, event labels indicating the start of an experiment suite, start and end of a condition and the start and end of individual trials were sent to the QTM server. These labels appear in the recorded 3D data and were exported alongside the marker positions for use in analysis and enable data to be segmented into their respective conditions and trials.

6 Analyses

6.1 Data Preprocessing

6.1.1 QTM

Each session recorded had the AIM model applied to the duration of the recording, and labelled markers were manually verified and adjusted as required to ensure that for each completed trial, there was 100% coverage of the marker data.

6.1.2 3D Data

3D data were exported from QTM and several preprocessing scripts were developed using the R programming language. Data were imported and collated by unique subject pair, condition and trial using the indices of the associated event labels. Subsequently, practice trials, data outside of trials and invalid trials were removed. Invalid trials were defined as trials that did not have both a start and an end event label. Trajectories were then created from marker x-coordinate time series using the R package `mousetrap` (Wulff et al., 2021), which was designed to aid analyses of mouse movement trajectories, and is able to be applied to arbitrary spatial data. The starting position of trajectories were aligned to account for track movement between trials as well as axis misalignment, and x-axis trajectories were standardized within trials to have a mean of zero and a range from -1 to 1 allowing comparison between subject pairs, conditions and trials. Visual inspection of trajectory data was performed, and six trials where participants had lost control of the sleds were truncated to the time of the incident. This left a total of 44 experimental trials (38 complete and 6 truncated, partial trials), meaning

[2]Source, data and analyses are available at https://github.com/zeyus/QTM_Bela_Sonification

[3]Commit ID `42bbf18c3710ed82cdf24b09ab72ac2239bf148e` from 10 August 2022: https://github.com/BelaPlatform/Bela/commit/42bbf18c3710ed82cdf24b09ab72ac2239bf148e

observation data were available for all subject pairs in all conditions, with a single trial unavailable for the no sonification condition from one subject pair.

6.2 Subject Synchronization

Three methods were used on data collected from the experiment to assess the level of synchrony between the participants in the various sonification conditions. The first method, absolute spatial distance (delta), provides a simple and straightforward measure of the overall synchrony between the sleds, but does not account for the temporal dynamics of the movement. The second method, Instantaneous Phase Angle Difference, provides real-time information about the synchrony of the movement, but may be sensitive to noise in the data and requires a good understanding of the mathematics involved. The third method, Dynamic Time Warping, offers a detailed analysis of the temporal dynamics of the movement and can handle differences in the speed of the movement, but may be computationally intensive and sensitive to the choice of the time-warping parameter. Each of the three methods has its own strengths and limitations, and the choice of which method to use will depend on the specific goals and requirements of the analysis.

6.2.1 Absolute Spatial Distance

Distance between subject sleds is a useful proxy for determining the level of success of synchronization, where a trial where subjects move perfectly together would result in a delta of zero for each time point, and large distances would indicate that they were unable to synchronize their sled movements. Absolute distance deltas between the standardized x-coordinates of subject pairs were calculated for each time point by trial and condition. These delta values were used to compute mean and standard deviation values for the experimental conditions. Furthermore, a linear mixed model was fit to the data.

6.2.2 Instantaneous Phase Angle of Trajectories

Instantaneous phase is a commonly used method for assessing synchrony between two or more signals. It involves analyzing the time-varying phase of signals and computing the phase difference between them at each point. It is often used to analyze neural data such as EEG signals and other data where rhythmic fluctuations are present, including movement (Varlet & Richardson, 2011). The phase relationship between signals can provide insight into the degree of synchronization between the signals. For the present study trajectory data were processed using a Hilbert transform, which resulted in the calculation of the instantaneous phase angle for each subject at each time point. The angle difference between subject pairs were calculated for each time point and the absolute difference in instantaneous phase angles were used to determine the mean angle difference and standard deviation for each condition. Although using the absolute value in the calculations removes information about the leader-follower dynamics, it allows a more useful mean value to be calculated from these data due to the continuous change in direction at the ends of the track creating jumps from 0 to 180 degrees and vice-versa, making the mean values converge towards 90 degrees. Additionally, a linear mixed model was fit to resulting angles.

6.2.3 Dynamic Time Warping

Dynamic Time Warping (DTW) is a method used to compare the similarity of two or more sequences of data. DTW works by finding the optimal alignment between the two sequences by stretching or compressing one sequence so that the difference between the sequences is minimized ("Dynamic Time Warping", 2007). This alignment is represented by a mapping function that indicates the relationship between the two sequences at each point in time. DTW has found applications in a wide range of fields, including speech recognition, music analysis, and in joint action (Hoch et al., 2021). For the present study, DTW was used to compare the trajectories of the two sleds and assess the level of synchrony between them over time, allowing for a more detailed analysis of the movement patterns. For the analysis, data were decimated, resulting in a sample rate of 30 Hz which reduced the computation time significantly and comparisons between several full-rate and down-sampled DTW analyses resulted in comparable normalized path distances. DTW path distances were calculated with the R package `dtw` (Giorgino, 2022) using the Sakoe-Chiba windowing method (Geler et al., 2019) with a window size of 90 samples (3 seconds) and using the `symmetric2` step pattern. Furthermore, a linear mixed model was fit to the normalized path distances.

7 Results

7.1 Spatial Difference

Analysis of normalized x-coordinate deltas showed 0.174 $\pm$ 0.196 for the no sonification condition, 0.237 $\pm$ 0.297 for the task condition, and 0.162 $\pm$ 0.16 for the sync condition. (Figure 7).

We fitted a linear mixed model (estimated using REML and nloptwrap optimizer) to predict Position Delta with Condition (formula: `Position Delta` ~ Condition). The

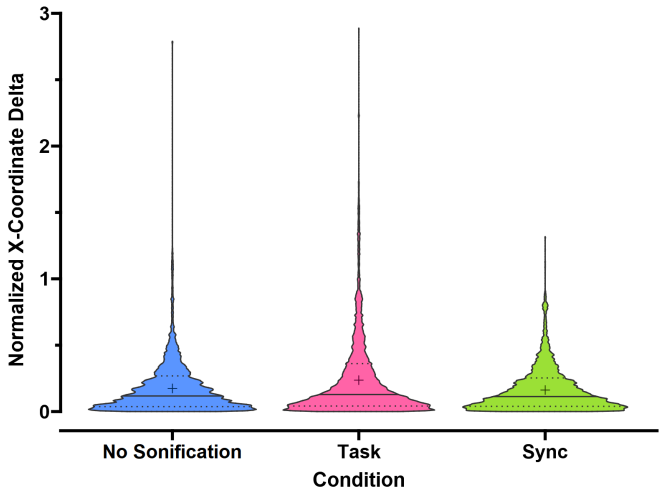

Figure 7: Distribution density plot of normalized x-coordinate deltas between subject pairs for each condition. The mean normalized distance is shown as a point (+), the 20th and 80th percentiles are shown as dotted lines and the 50th percentile is shown as a solid line. No Sonification condition normalized delta mean = 0.174 (sd = 0.196), task sonification condition normalized delta mean = 0.237 (sd = 0.297), sync sonification condition normalized delta mean = 0.162 (sd = 0.16).

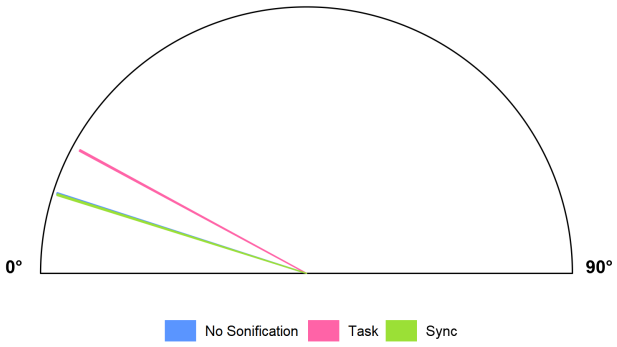

Figure 8: Plot of mean absolute instantaneous phase angles of experimental conditions with the length of the needles representing the standard deviation as a percentage of 90 degrees. No Sonification condition mean phase angle = 8.861 (sd = 0.067) degrees, task sonification condition mean phase angle = 14.211 (sd = 0.216) degrees, sync sonification condition mean phase angle = 8.749 (sd = 0.075) degrees.

model included Subject Pair as random effects (formula: list(~1 | Subject Pair, ~1 | Trial)). The model's total explanatory power is moderate (conditional R2 = 0.19) and the part related to the fixed effects alone (marginal R2) is of 0.02. The model's intercept, corresponding to Condition = No Sonification, is at 0.17 (95% CI [0.08, 0.25], t(2370020) = 3.78, p < .001). Within this model:

- The effect of Condition [Task] is statistically significant and positive (beta = 0.06, 95% CI [0.06, 0.06], t(2370020) = 190.11, p < .001; Std. beta = 0.27, 95% CI [0.27, 0.28])
- The effect of Condition [Sync] is statistically significant and positive (beta = 1.19e-03, 95% CI [5.29e-04, 1.85e-03], t(2370020) = 3.54, p < .001; Std. beta = 5.19e-03, 95% CI [2.31e-03, 8.07e-03])

Standardized parameters were obtained by fitting the model on a standardized version of the dataset. 95% Confidence Intervals (CIs) and p-values were computed using a Wald t-distribution approximation.

7.2 Instantaneous Phase Angle

Analysis of the mean absolute instantaneous phase angle difference between subject pair Hilbert transformed trajectories gave 8.861° ± 0.067° (SD) for the no sonification condition, 14.211° ± 0.216° (SD) for the task condition, and 8.749° ± 0.075° (SD) for the sync condition. (Figure 8).

We fitted a linear mixed model (estimated using REML and nloptwrap optimizer) to predict Relative IPA with Condition (formula: `Relative IPA` ~ Condition). The model included Subject Pair as random effects (formula: list(~1 | Subject Pair, ~1 | Trial)). The model's total explanatory power is weak (conditional R2 = 0.10) and the part related to the fixed effects alone (marginal R2) is of 0.01. The model's intercept, corresponding to Condition = No Sonification, is at 11.20 (95% CI [4.48, 17.91], t(2370020) = 3.27, p = 0.001). Within this model:

- The effect of Condition [Task] is statistically significant and positive (beta = 6.65, 95% CI [6.58, 6.73], t(2370020) = 168.83, p < .001; Std. beta = 0.25, 95% CI [0.25, 0.26])
- The effect of Condition [Sync] is statistically significant and positive (beta = 0.54, 95% CI [0.46, 0.62], t(2370020) = 13.44, p < .001; Std. beta = 0.02, 95% CI [0.02, 0.02])

Standardized parameters were obtained by fitting the model on a standardized version of the dataset. 95% Confidence Intervals (CIs) and p-values were computed using a Wald t-distribution approximation.

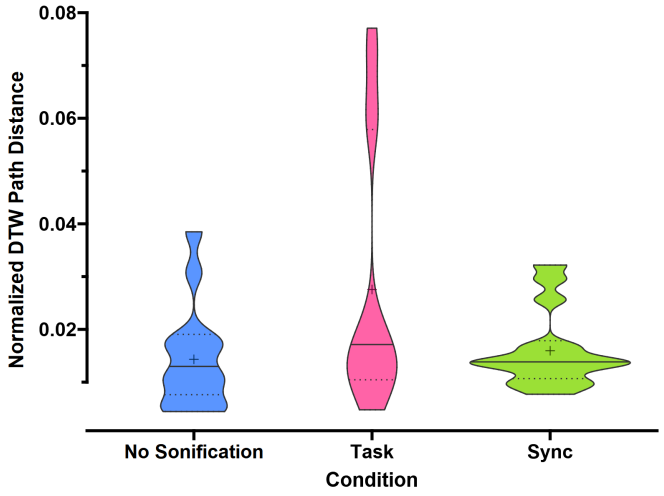

Figure 9: Distribution density plot of normalized DTW path distance between paired trajectories by condition. The mean normalized path distance is shown as a point (+), the 20th and 80th percentiles are shown as dotted lines and the 50th percentile is shown as a solid line. No Sonification condition normalized path distance mean = 0.014 (sd = 0.01), task sonification condition normalized path distance mean = 0.028 (sd = 0.025), sync sonification condition normalized path distance mean = 0.016 (sd = 0.007).

7.3 Dynamic Time Warping

Analysis of DTW results gave a mean normalized path distance of 0.014 ± 0.01 for the no sonification condition, 0.028 ± 0.025 for the task condition, and 0.016 ± 0.007 for the sync condition. (Figure 9).

We fitted a linear mixed model (estimated using REML and nloptwrap optimizer) to predict Normalized Distance with Condition (formula: `Normalized Distance` ~ Condition). The model included Subject Pair as random effect (formula: ~1 | `Subject Pair`). The model's total explanatory power is substantial (conditional R2 = 0.34) and the part related to the fixed effects alone (marginal R2) is of 0.12. The model's intercept, corresponding to Condition = No Sonification, is at 0.01 (95% CI [3.01e-03, 0.02], t(39) = 2.58, p = 0.014). Within this model:

- The effect of Condition [Task] is statistically significant and positive (beta = 0.01, 95% CI [2.81e-03, 0.02], t(39) = 2.55, p = 0.015; Std. beta = 0.80, 95% CI [0.16, 1.44])
- The effect of Condition [Sync] is statistically non-significant and positive (beta = 2.06e-03, 95% CI [-8.80e-03, 0.01], t(39) = 0.38, p = 0.703; Std. beta = 0.12, 95% CI [-0.52, 0.76])

Standardized parameters were obtained by fitting the model on a standardized version of the dataset. 95% Confidence Intervals (CIs) and p-values were computed using a Wald t-distribution approximation.

8 Discussion

Although we could not completely isolate the joint action representations, partially due to sound created from the movement of the sleds, our sonification strategies primed attention towards either self-other monitoring by sonification of both sled positions independently, or joint outcome by sonification of the positions of each sled relative to the other. The results of the models fit for calculated synchrony measures all showed that the No sonification performed best, with the lowest distance delta, relative instantaneous phase angle and DTW path length, and the task condition performed the worst and the model estimates were reported as statistically significant for all three measures. The Sync condition was marginally worse than the no sonification condition, but the model results were only statistically significant for the models applied to distance delta and relative instantaneous phase angle.

While these results may initially seem surprising, we have identified potential contributing factors that may have affected the outcome. All participants reported during debriefing that they subjectively felt that the no sonification condition was the easiest for them, despite the fact that they were blindfolded for all the conditions. One of the most obvious of the possible reasons is the sound of the sleds sliding on the track, although not very loud, provides spatial information to the subjects in a familiar way – that is, binaural input that allows auditory localization – and additional information such as the track length was already known to subjects. Individuals with unimpaired hearing are able to use auditory cues to discriminate the spatial location of movement, Carlile and Leung (2016) summarize that the level of accuracy is dependent on the velocity of movement (degrees/second), with lower velocity generally resulting in higher accuracy, and in our experiment, the velocity of the sleds in degrees/second would be relatively low[4]. In the experimental conditions, the auditory localization is masked by the sonification, which comes from speakers either in front or behind the subject depending on the side of the track they were seated at. This issue could be mitigated by the addition of a condition where subject hearing was also restricted, giving a baseline level of synchrony, as well as using some form of spatial cues in the sonification conditions in the form of stereo separation that mirrors track position, or requiring

[4]Given the track length of 1130mm and a head positioned 700mm perpendicular from the track, there is approximately 78 degrees between the start and end of the track, giving an approximate velocity of 26 degrees/second and 4 degrees/second for the highest and lowest frequency subjects respectively.

headphones for all the conditions and implementing a real time binaural synthesis (Tommasini et al., 2019).

Another potential confounding factor is the implementation of the sonification strategies. We initially planned for them to have at least a stereo component, but due to hardware constraints and time limitations, the sonification was restricted to using speakers situated in the room. Both subjects would hear the same audio during the experiment which makes it impossible to sonify the position of a single subject relative to the other (i.e. they are to the left or right of the other person). In the task condition, this is less important, as their absolute position on the track is sonified, while in the sync position, this would provide subject-specific information about their relative location. Although the mapping of pitch was consistent across all trials because participants did not change seats during an experiment, this limitation meant that subjects would have to learn the mapping of the undertone to their sled.

One further limitation is the frequency range that was used in the experiment, although during testing, it was harmonious and not irritating, due to a mistake in the software synthesizer tone generation process, the output frequency was one octave lower than the selected note[5] (see Figure 10), and as such, may have resulted in more generally dissonant sounding sonification than the originally intended center frequency of 440Hz (Figure 2).

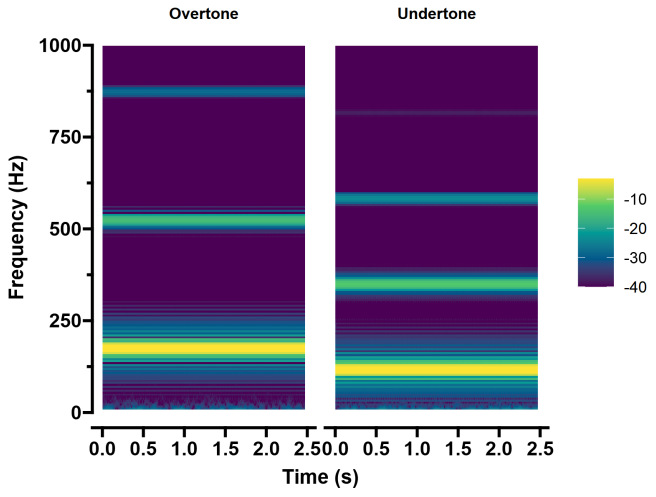

Figure 10: Spectral analysis of the two base frequency audio files

Examples illustrating the different sonification strategies: same x coordinate resulted in a perfect fifth for both strategies, but a flawlessly performed task would sound much different between the conditions. The sync-oriented would produce a stable, unwavering sound interval; the task-oriented would produce a sound that slides up and down in parallel with the track, the constant being the interval that remains a fifth between the two tones.

How might our task be affected by musical training? "Successful music performance requires that musicians monitor the auditory consequences of their actions. Years of training on an instrument lead to strong associations between a given movement or set of movements and a given auditory outcome." (Loehr et al., 2013)

ML model applications in this project? Could consider another sonification to compare to: besides no sonification / dealing with spatial audio, could also have a regular beep to mark when they should be at the ends, or even a constant range representing where they should be (i.e. directing their movements with the sound rather than sonification of their movement, this could be another type of control and represent some more traditional methods) Loudness and brightness may be better choices for modulation targets than pitch changes: (McDermott et al., 2010)

Mention the marker placement as being problematic. Also the number of reference markers…also that we wanted to use headphones… Additionally, we could have had start and end checkpoints where it could be used to segment the data into sub-trials or maybe even make participants stop and wait (but that ruins flow)

Mention analysis of learning effect, not enough data here, too many invalid trials, but would be interesting to see if participants "learn" the sonification schemes

• How does our task compare to e.g. Loehr et al. piano duet task? • What can and can't our task reveal? o Can't say that only one of the participants makes a mistake, since the goal is to in sync with each other, without a general "tempo" to follow • Individual representations vs joint outcome representations? o Often can't mirror the other person exactly

if someone wanted to reproduce we can talk about things like the height of the track, the width of the base making it less comfortable, the range in participant music experience, also the fact that the mocap lab is small and there were problems making camera trakcing more difficult.

and of course all the stuff about different sonification options and testing ranges of values to see what works, and whole new strategies we haven't thought of using headphones to stop the track noise which subjects noted they used…

and of course we can talk about that in relation to the results, lack of results, why we might be seeing the data we see,

[5] The selected software synthesizer was a bass virtual instrument which output a midi note one octave below (i.e. A4, 440Hz would be rendered as A3, 220Hz)

and especially that we actually managed to create a low-latency method for sonification of /arbitrary/ data, and we used it for an expensive mocap setup but this could be applied to outputs from machine learning models (i.e. webcam object tracking) or even other low-cost hardware like wiimotes or something

we also don't use pure sine tones, so the more complex tones might make it more difficult to distinguish.,..we could try also octaves, perhaps it is easier to distinguish for nonmusicians.

Maybe we can add more dimensions

Learning: by trial (some initial result)

Applications

Movement sonification + joint action is useful for training sports, e.g. rowing, but also track relay, where two runners try to optimize the timing when passing the baton

References

Aglioti, S. M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience*, *11*(9), 1109–1116. https://doi.org/10.1038/nn.2182

Benedict, T., Barrick, G. A., & Pazder, J. (2016, August 9). Survey of materials and coatings suitable for controlling stray light from the near-UV to the near-IR. In C. J. Evans, L. Simard, & H. Takami (Eds.). https://doi.org/10.1117/12.2231348

Bevilacqua, F., Boyer, E. O., Françoise, J., Houix, O., Susini, P., Roby-Brami, A., & Hanneton, S. (2016). Sensori-Motor Learning with Movement Sonification: Perspectives from Recent Interdisciplinary Studies. *Frontiers in Neuroscience*, *10*. https://doi.org/10.3389/fnins.2016.00385

Bolt, N. K., & Loehr, J. D. (2021). Sensory Attenuation of the Auditory P2 Differentiates Self- from Partner-Produced Sounds during Joint Action. *Journal of Cognitive Neuroscience*, *33*(11), 2297–2310. https://doi.org/10.1162/jocn_a_01760

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. https://doi.org/10.1037/0033-295X.108.3.624

Calvo-Merino, B., Glaser, D., Grèzes, J., Passingham, R., & Haggard, P. (2005). Action Observation and Acquired Motor Skills: An fMRI Study with Expert Dancers. *Cerebral Cortex*, *15*(8), 1243–1249. https://doi.org/10.1093/cercor/bhi007

Carlile, S., & Leung, J. (2016). The Perception of Auditory Motion. *Trends in Hearing*, *20*, 233121651664425. https://doi.org/10.1177/2331216516644254

Clark, H. H. (2005). Coordinating with each other in a material world. *Discourse Studies*, *7*(4-5), 507–525. https://doi.org/10.1177/1461445605054404

Debashi, M., & Vickers, P. (2018). Sonification of network traffic flow for monitoring and situational awareness (R. Mankin, Ed.). *PLOS ONE*, *13*(4), e0195948. https://doi.org/10.1371/journal.pone.0195948

Dobson, W. (1963). Details - Public Health Image Library(PHIL). Retrieved February 14, 2023, from https://phil.cdc.gov/details.aspx?pid=12020

Dotov, D., & Froese, T. (2018). Entraining chaotic dynamics: A novel movement sonification paradigm could promote generalization. *Human Movement Science*, *61*, 27–41. https://doi.org/10.1016/j.humov.2018.06.016

Dubus, G. (2012). Evaluation of four models for the sonification of elite rowing. *Journal on Multimodal User Interfaces*, *5*(3-4), 143–156. https://doi.org/10.1007/s12193-011-0085-1

Dubus, G. (2013). Interactive sonification of motion : Design, implementation and control of expressive auditory feedback with mobile devices. Retrieved September 5, 2022, from http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-127944

Dubus, G., & Bresin, R. (2011). Sonification of Physical Quantities Throughout History: A Meta-Study of Previous Mapping Strategies. *International Conference on Auditory Display, 2011*, 8.

Dubus, G., & Bresin, R. (2013). A Systematic Review of Mapping Strategies for the Sonification of Physical Quantities. *PLOS ONE*, *8*(12), e82491. https://doi.org/10.1371/journal.pone.0082491

Dynamic Time Warping. (2007). In *Information Retrieval for Music and Motion* (pp. 69–84). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-74048-3_4

Effenberg, A. O., & Schmitz, G. (2018). Acceleration and deceleration at constant speed: Systematic modulation of motion perception by kinematic sonification. *Annals of the New York Academy of Sciences*, *1425*(1), 52–69. https://doi.org/10.1111/nyas.13693

Effenberg, A. (2005). Movement Sonification: Effects on Perception and Action. *IEEE Multimedia*, *12*(2), 53–59. https://doi.org/10.1109/MMUL.2005.31

Ferrari-Toniolo, S., Visco-Comandini, F., & Battaglia-Mayer, A. (2019). Two brains in action: Joint-action coding in the primate frontal cortex. *The Journal of Neuroscience*, 1512–18. https://doi.org/10.1523/JNEUROSCI.1512-18.2019

Galantucci, B. (2009). Experimental Semiotics: A New Approach for Studying Communication as a Form of Joint Action. *Topics in Cognitive Science*, *1*(2), 393–410. https://doi.org/10.1111/j.1756-8765.2009.01027.x

Geler, Z., Kurbalija, V., Ivanovic, M., Radovanovic, M., & Dai, W. (2019). Dynamic Time Warping: Itakura vs Sakoe-Chiba. *2019 IEEE International Symposium on INnovations in Intelligent SysTems and Applications (INISTA)*, 1–6. https://doi.org/10.1109/INISTA.2019.8778300

Giorgino, T. (2022). *Dtw: Dynamic time warping algorithms* [R package version 1.23-1]. https://CRAN.R-project.org/package=dtw

Hildebrandt, T. (2014). Short Paper: Towards Enhancing Business Process Monitoring with Sonification. In N. Lohmann, M. Song, & P. Wohed (Eds.), *Business Process Management Workshops* (pp. 529–536, Vol. 171). Springer International Publishing. https://doi.org/10.1007/978-3-319-06257-0_42

Hoch, J. E., Ossmy, O., Cole, W. G., Hasan, S., & Adolph, K. E. (2021). "Dancing" Together: Infant–Mother Locomotor Synchrony. *Child Development*, *92*(4), 1337–1353. https://doi.org/10.1111/cdev.13513

Keetels, M., & Vroomen, J. (2012). Perception of synchrony between the senses. In Murray, MM & Wallace, MT (Eds.), *The neural bases of multisensory processes*. CRC Press/Taylor & Francis. https://www.ncbi.nlm.nih.gov/books/NBK92837/

Keller, P. E., & Burnham, D. K. (2005). Musical Meter in Attention to Multipart Rhythm. *Music Perception*, *22*(4), 629–661. https://doi.org/10.1525/mp.2005.22.4.629

Keller, P. E., Knoblich, G., & Repp, B. H. (2007). Pianists duet better when they play with themselves: On the possible role of action simulation in synchronization. *Consciousness and Cognition*, *16*(1), 102–111. https://doi.org/10.1016/j.concog.2005.12.004

Kimoto, M., & Ohno, H. (2002). Design and implementation of stetho: Network sonificiation system. *ICMC*.

Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological Research on Joint Action. In *Psychology of Learning and Motivation* (pp. 59–101, Vol. 54). Elsevier. https://doi.org/10.1016/B978-0-12-385527-5.00003-6

Knoblich, G., & Jordan, J. S. (2003). Action coordination in groups and individuals: Learning anticipatory control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(5), 1006–1016. https://doi.org/10.1037/0278-7393.29.5.1006

Kosinski, R. J. (2008). A Literature Review on Reaction Time. *Clemson University*, *10(1)*, 337–344.

Kourtis, D., Sebanz, N., & Knoblich, G. (2012). Predictive representation of other people's actions in joint action planning: An EEG study. *Social neuroscience*, *8*. https://doi.org/10.1080/17470919.2012.694823

Kramer, G., Walker, B., Bonebright, T., & Cook, P. (1999). Sonification Report: Status of the Field and Research Agenda.

Loehr, J. D., Kourtis, D., Vesper, C., Sebanz, N., & Knoblich, G. (2013). Monitoring Individual and Joint Action Outcomes in Duet Music Performance. *Journal of Cognitive Neuroscience*, *25*(7), 1049–1061. https://doi.org/10.1162/jocn_a_00388

Loehr, J. D., & Palmer, C. (2011). Temporal Coordination between Performing Musicians. *Quarterly Journal of Experimental Psychology*, *64*(11), 2153–2167. https://doi.org/10.1080/17470218.2011.603427

Loehr, J. D., & Vesper, C. (2016). The sound of you and me: Novices represent shared goals in joint action. *Quarterly Journal of Experimental Psychology*, *69*(3), 535–547. https://doi.org/10.1080/17470218.2015.1061029

McDermott, J. H., Keebler, M. V., Micheyl, C., & Oxenham, A. J. (2010). Musical intervals and relative pitch: Frequency resolution, not interval resolution, is special. *The Journal of the Acoustical Society of America*, *128*(4), 1943–1951. https://doi.org/10.1121/1.3478785

McPherson, A. P., Jack, R. H., Moro, G., & Proceedings of the International Conference on New Interfaces for Musical Expression, B. (2016). Action-Sound Latency: Are Our Tools Fast Enough? Retrieved October 31, 2022, from https://qmro.qmul.ac.uk/xmlui/handle/123456789/12479
Accepted: 2016-05-24T10:51:01Z.

Roguin, A. (2006). Rene Theophile Hyacinthe Laennec (1781-1826): The Man Behind the Stethoscope. *Clinical Medicine & Research*, *4*(3), 230–235. https://doi.org/10.3121/cmr.4.3.230

Schmitz, G., & Effenberg, A. O. (2012). Perceptual effects of auditory information about own and other movements. *Proceedings of the 18th International Conference on Auditory Display*.

Schmitz, G., & Effenberg, A. O. (2017). Sound Joined Actions in Rowing and Swimming. In C. Meyer & U. V. Wedelstaedt (Eds.), *Moving bodies in interaction– interacting bodies in motion: Intercorporeality, interkinaesthesia, and enaction in sports* (pp. 193–214). John Benjamins Publishing Company.

Schmitz, G., Mohammadi, B., Hammer, A., Heldmann, M., Samii, A., Münte, T. F., & Effenberg, A. O. (2013). Observation of sonified movements engages a basal ganglia frontocortical network. *BMC Neuroscience*, *14*(1), 32. https://doi.org/10.1186/1471-2202-14-32

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76. https://doi.org/10.1016/j.tics.2005.12.009

Sethares, W. A. (2005). Sound on Sound. In *Tuning, Timbre, Spectrum, Scale* (pp. 39–50). Springer-Verlag. https://doi.org/10.1007/1-84628-113-X_3

Tommasini, F. C., Ramos, O. A., Hüg, M. X., & Ferreyra, S. P. (2019). A Computational Model to Implement Binaural Synthesis in a Hard Real-Time Auditory Virtual Environment. *Acoustics Australia*, *47*(1), 51–66. https://doi.org/10.1007/s40857-019-00152-7

van der Steen, M. C. (, & Keller, P. E. (2013). The ADaptation and Anticipation Model (ADAM) of sensorimotor synchronization. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00253

van der Wel, R. P., Becchio, C., Curioni, A., & Wolf, T. (2021). Understanding joint action: Current theoretical and empirical approaches. *Acta Psychologica*, *215*, 103285. https://doi.org/10.1016/j.actpsy.2021.103285

Varlet, M., & Richardson, M. J. (2011). Computation of continuous relative phase and modulation of frequency of human movement. *Journal of Biomechanics*, *44*(6), 1200–1204. https://doi.org/10.1016/j.jbiomech.2011.02.001

Vesper, C., Butterfill, S., Knoblich, G., & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks*, *23*(8-9), 998–1003. https://doi.org/10.1016/j.neunet.2010.06.002

Vesper, C., Knoblich, G., & Sebanz, N. (2014). Our actions in my mind: Motor imagery of joint action. *Neuropsychologia*, *55*, 115–121. https://doi.org/10.1016/j.neuropsychologia.2013.05.024

Vesper, C., van der Wel, R. P. R. D., Knoblich, G., & Sebanz, N. (2013). Are you ready to jump? Predictive mechanisms in interpersonal coordination. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 48–61. https://doi.org/10.1037/a0028066

Vickers, P. (2011). Sonification for process monitoring. In *The sonification handbook* (pp. 455–492).

Vinken, P. M., Kröger, D., Fehse, U., Schmitz, G., Brock, H., & Effenberg, A. O. (2013). Auditory Coding of Human Movement Kinematics. *Multisensory Research*, *26*(6), 533–552. https://doi.org/10.1163/22134808-00002435

Wenke, D., Atmaca, S., Holländer, A., Liepelt, R., Baess, P., & Prinz, W. (2011). What is Shared in Joint Action? Issues of Co-representation, Response Conflict, and Agent Identification. *Review of Philosophy and Psychology*, *2*(2), 147–172. https://doi.org/10.1007/s13164-011-0057-0

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction (C. Frith & D. Wolpert, Eds.). *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1431), 593–602. https://doi.org/10.1098/rstb.2002.1238

Wulff, D. U., Kieslich, P. J., Henninger, F., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (2021). Movement tracking of cognitive processes: A tutorial using mousetrap. https://doi.org/10.31234/osf.io/v685r

Yang, J., & Hunt, A. (2015). Real-time sonification of biceps curl exercise using muscular activity and kinematics OCLC: 951458301.