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: 4 January 2023

Student Details

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

Software

- Repository ⧉

Contents

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

Joint action tasks form an integral part of the everyday life of humans and many other species [ref], and the mechanisms underlying this cooperative ability to work together towards a common goal have been the subject of an increasing number of publications [ref, scholar search?]. Many types of cooperation involve auditory perception as a key aspect, 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 synchronization, for example with a steady beat to set the pace for rowers. 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 disjoint [ref]. While there has been research into the effects of sonification on joint action [ref], this thesis presents a flexible low latency sonification framework that uses real-time positional data for joint action research. This framework has been implemented in a pilot study to investigate the utility of a novel joint action synchrony paradigm that puts the focus of the sonification strategy as a core aspect of study design in this field. By comparing subject synchronization during experiments employing task-oriented or synchronization-oriented strategies, we have attempted to show differences that highlight the importance of strategy selection for sonification and provide a pathway for further investigation.

2 Background

2.1 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 behaviour. Examples include handshakes, conversations, musical performances, dancing...

What are the cognitive processes involved in joint action? Recent theory suggests that the main cognitive processes involved include representations, action monitoring and action prediction (Loehr et al., 2013; Sebanz et al., 2006; Vesper et al., 2010).

2.1.1 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. While it is not required, it is usually 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). 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. This is supported by a study (Keller et al., 2007) where pianists were asked to record one part from a selection of piano duets and then, at a later date, 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 having more accurate representations of the other's actions facilitated synchronization.

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).

2.1.2 Action Monitoring

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: one's own task, the other's task, and the shared goal. As with representations, one must at least monitor the progress of one's 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. 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 pianists played a pre-rehearsed duet on a digital piano while the outcomes of

keypresses were manipulated by the researchers. In the individual outcome condition, the produced tones 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. {More text to follow, incorporating these findings: "musicians engaged in joint actions monitor their own and their partner's actions as well as their combined action outcomes, while at the same time maintaining a distinction between their own and others actions and between individual and joint outcomes.[…] The first goal of the current study was to provide further evidence that people monitor their own and their partners action outcomes in parallel. The second goal of the current study was to investigate whether people monitor each person s individual part in a joint action and/or the combined outcome of their coordinated actions. […] These findings indicate that skilled performers are able to monitor the outcomes of their own actions, their coperformers actions, and their combined actions when they perform joint actions together. They also indicate that performers nevertheless differentiate between their own and others action outcomes and between individual and joint action outcomes."}

2.1.3 Action Predicting

In order to coordinate effectively, it is often necessary to make predictions about future events. 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 (Gerd Schmitz & Alfred O. Effenberg, 2017; Vesper et al., 2010). Simulating the actions of others as they occur may be especially beneficial when engaging in joint action, as it is believed to influence perception and assist 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). It is not fully clear yet whether similar mechanisms exist specifically for predicting the joint outcome of an agent and their partner. Some support for predicting joint outcomes comes from a study by Knoblich & 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.

2.1.4 Integrating Predicted Outcomes of One's Own and Others' Actions

The crucial final feature of joint action is the manner in which individuals adapt their own actions to those of others in time and space. The question concerns the decision-making process that occurs after a prediction of another person's action is made, i.e. choosing a suitable complementary action and performing it at the appropriate time. 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). A study by Knoblich and Jordan [Sebanz 43-34] using a joint action tracking paradigm found that partners planned their actions based on a prediction of what the joint effect of both their own and their partner's actions would be.

Some of the questions that have been investigated in recent joint action research concern the aforementioned 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 (citation), whether they monitor individual outcomes or joint outcomes, and… The current study primed the participants' attention towards individual and joint outcomes in different conditions while investigating joint task synchronization.

2.1.5 Current Study

The current study had participants perform a joint action task under three different conditions, where the sensory consequences were manipulated through real-time sonification of movement.

The current study focuses on what happens during learning of joint action. Through movement sonification, we can enhance attention towards specific features of joint action. Is learning optimized when focusing on self-other representations or joint outcome representations?

While we could not completely separate the conditions, our sonification primes attention towards either self-other monitoring, or joint outcome.

2.2 Sonification

Sonification is defined as the use of nonspeech audio to convey information. 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).

While 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. 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. 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".

2.2.1 What is sonification useful for?

{this section is just raw, copypasted quotes for now} Making sense of large amounts of data, and utilizing modern powerful media technologies. "Sonification research is well positioned to provide technology to assist scientists in comprehending the information and data-rich world of today and of the future.". "The wide availability of audio technology (e.g., in multimedia computers) makes auditory data representation a viable option for large numbers of users. Thus, there exists today a synergism between the widespread need for new data comprehension methods and readily available technology that, with proper support and funding, can lead to a large number of users reaping the benefits conferred by the development of scientific sonification." (Kramer et al., 1999)

2.2.2 Movement sonification

{this section is just raw, copypasted quotes for now} "Approaches within the discipline of Sport Science reflect the whole range from fundamental research with high internal validity to applied research with high ecological validity. Applied research plays an important role for the development of new, more effective intervention methods. Assuming that more senses are more powerful in perceiving gross motor patterns it should be supportive to create and convey more acoustic movement information. For multisensory integration benefits, additional auditory movement information has to correspond to the structure of a perceptual feature stream of another modality (visual,

kinesthetic, tactile). For such an acoustic enhancement of motor perception Effenberg (1996, 2004, 2005) has established the concept of 'movement sonification', adapting the sonification approach of the early 1990s to the kinematics and dynamics of human motor behavior." One of the movement parameters that can be used is "kinematic parameters representing the spatiotemporal features of a pose or a movement pattern." "The question whether dynamic or kinematic movement parameters should be chosen for movement sonification should be answered under consideration of the sensory modality or modalities with which bi- or multimodal convergence should be achieved: If visual motion perception is the reference, movement sonification should be based on kinematic parameters." (Gerd Schmitz & Alfred O. Effenberg, 2017)

> "Subjects are able to perceive differences in swimming stroke frequency more accurately when visualizations of a swimmer are complemented with a kinematic sonification." (Gerd Schmitz & Alfred O. Effenberg, 2017)

2.2.3 Auditory perception

Research into auditory perception indicates two basic features of auditory perception that provide good arguments for representing data as sound. First, auditory perception is especially useful for detecting temporal characteristics, i.e. variations in sound over time. [find study that said we're better at detecting rhythm aurally vs visually?] 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 monitoring and alarm applications. (Kramer et al., 1999)

Other benefits of auditory perception that speak for sonification: parallel listening (ability to monitor and process multiple auditory data sets), rapid detection (especially in high-stress environments), affective response (ease of learning and high engagement qualities) and auditory gestalt formation (discerning relationships or trends in data streams) (Kramer et al., 1999).

Harmony in Motion: Real-time Sonification Strategies for Joint Action Research, (2023).

6

3 Hardware and Software Implementation

3.1 Motion Capture

Motion capture data were collected using a 9 camera (8 Qualisys Miqus M3 marker 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 version 2022.2 (build 7700) software was used to collect and process the data with real-time 3D tracking data output.

3.1.1 Markers

For the experimental set up, one passive marker was placed on each car, and two additional passive reference markers were placed on the front corners of the track (see Fig X). These additional markers provided reference points for 3D orientation of the track and the cars across trials in case of accidental track movement. QTM Automatic Identification of Marker (AIM) models were trained on variable speed sled movements along the track and were given labels for client-side identification.

3.1.2 Event Labels

Event labels were sent to QTM to mark the start and end of each condition and each trial within that condition.

3.2 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[1]. 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 Bela platform framework from 10 August 2022[2]. The experiment flow control was automated using the defined time periods for trials and breaks between them, and paused between conditions until a button on the device was pressed to allow sufficient time for subjects to rest and answer the IOS survey. Figure [xxx] describes the flow of data from motion capture to sonification.

---

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

[2]https://github.com/BelaPlatform/Bela/commit/42bbf18c3710ed82cdf24b09ab72ac2239bf148e

3.2.1 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 server and logging the elapsed time between resulting in a mean latency of x.xxxms (sd/min/max). 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.

4 Methods

4.1 Pilot 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 as possible.

4.2 Participants

An availability sample of ten subjects were recruited to participate in pairs, subjects were all aged between 20 and 30 years and were a mix of genders and handedness. Most subjects also reported no formal musicianship training. (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 and information sheet and were under the umbrella project...

4.3 Experimental Design

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 try to continuously move the sleds back and forth along the track 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. After the conclusion of subject briefing and they had indicated they were ready, they were blindfolded for the duration of all trials within a condition, with a pause between conditions where they could rest and remove the blindfold.
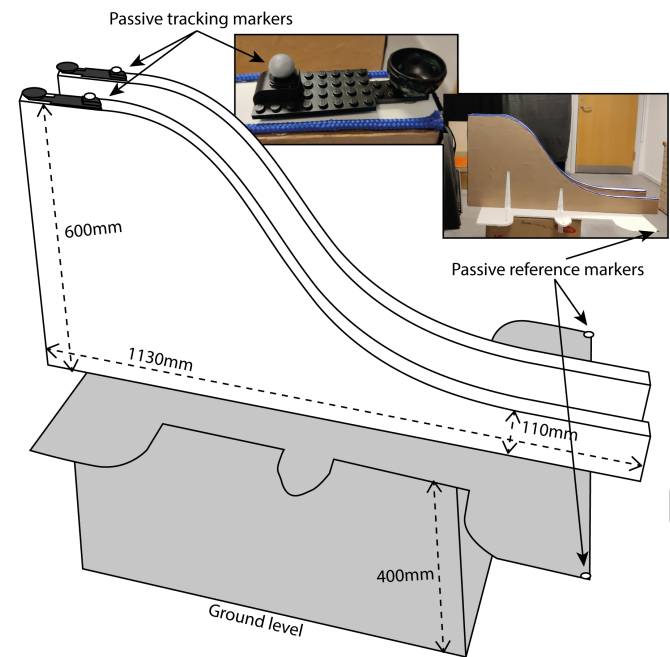
Figure 1: Experimental track setup showing track shape and dimensions

4.3.1 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 made from LEGO parts were constructed and three felt adhesive pads were attached to the underside of each to reduce resistance during movement. See Figure 1

4.3.2 Frequency Range Selection

Two distinct, continuous frequency ranges were selected for application in sonification conditions. These ranges are offset by a perfect fifth and span eight semitones, the overtone range was chosen based on a center frequency of 440Hz (A4) and the range was limited to avoid a large overlap with the undertone range during normal operation (Table X). Consideration was also given to creating ranges that were not sufficiently high to cause discomfort, nor sufficiently low that distinguishing slight differences becomes more difficult.

4.3.3 Sonification Strategies

The experiment consisted of three conditions that vary the sonification strategy employed, 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

Table 1: Frequency ranges for the two tones used in the sonification conditions. Note names are in International Pitch Notation, and are the closest approximation. Overtone frequencies were calculated to have a center frequency of 440Hz, and undertone frequencies are two-thirds of their overtone counterparts.

	Lower Bound		Center		Upper Bound	
	Freq	Note[a]	Freq	Note[a]	Freq	Note[a]
Overtone	349.230	F4	440.000	A4	554.365	C#5
Undertone	232.819	A#3	293.333	D4	369.577	F#4

[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 440Hz, and undertone frequencies are two-thirds of their overtone counterparts.

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 may use it to experiment with the sonification.

4.3.3.1 No sonification.

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

4.3.3.2 Task-oriented sonification strategy.

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 subjects were at the exact same x-coordinate, the two tones would be a perfect fifth apart, creating a harmonious chord, as the sleds drift further apart, the frequency difference would deviate from the perfect fifth and create a more discordant sound. This strategy was selected for sonifying the movement along the track, i.e. the task required of subjects.

4.3.3.3 Synchronization-oriented sonification strategy.

The sonification strategy oriented around synchronization represented the position of the sleds relative to each other,…

4.4 Procedure
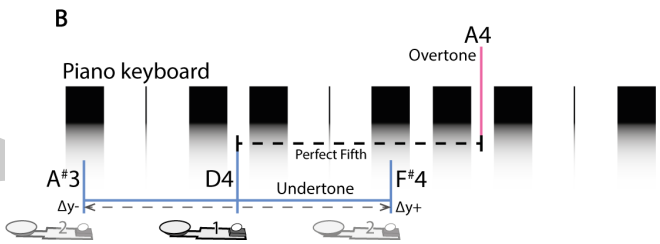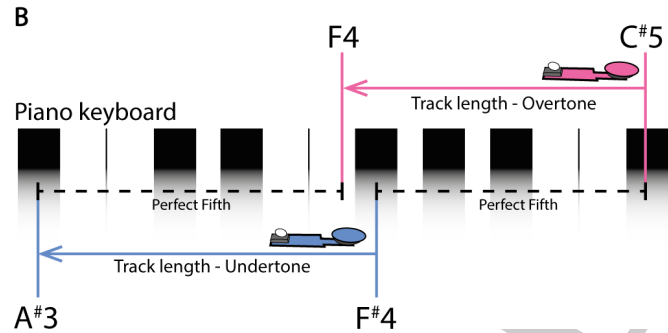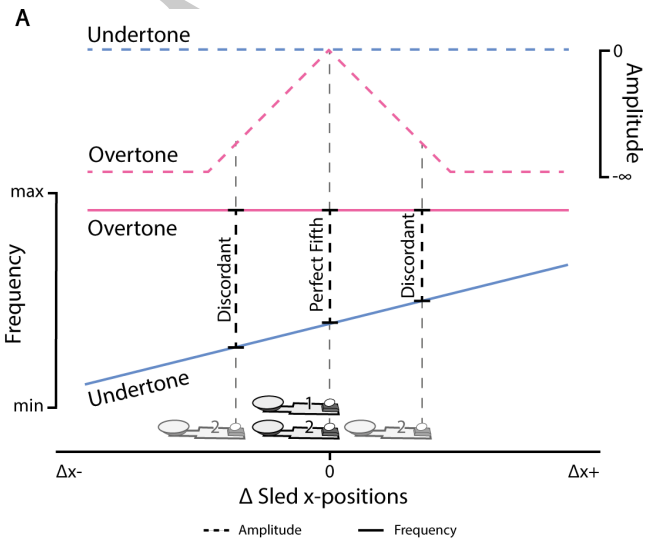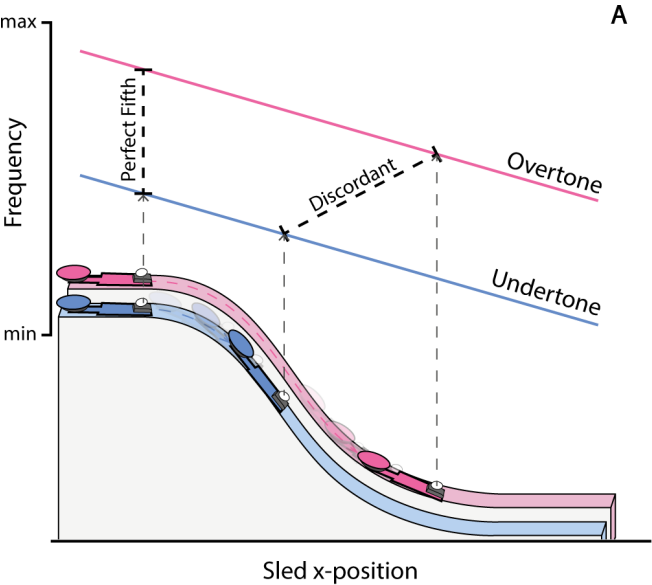
This is where the experiment flow will be described.

Harmony in Motion: Real-time Sonification Strategies for Joint Action Research, (2023).

8

Figure 2: Task

Figure 3: Sync

Harmony in Motion: Real-time Sonification Strategies for Joint Action Research, (2023).

9

5   Data Analysis

5.1   Data Preprocessing

5.1.1   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. Processing of every frame and 2D data preprocessing were disabled for real-time output to ensure minimal latency.

6   Results

[Describe]

7   Discussion

[Explain]

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)

Mention the marker placement as being problematic. Also the number of reference markers...also that we wanted to use headphones...

• 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

NULL

7.0.1   Testing different model / outputs / tables

just making sure that the output looks good and seeing which is the most useful for our case.

This has "add_glance_table"

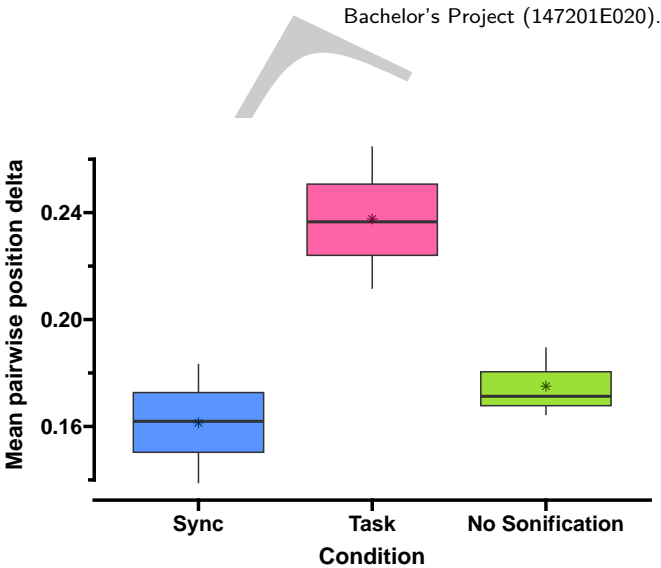Comparing the distance model with the null distance model

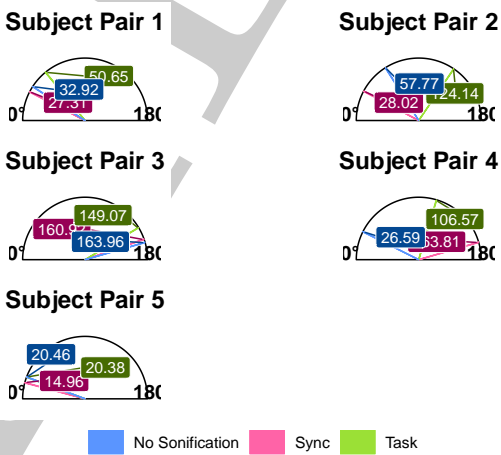**Figure 4:** Plot of pairwise position delta for each subject pair

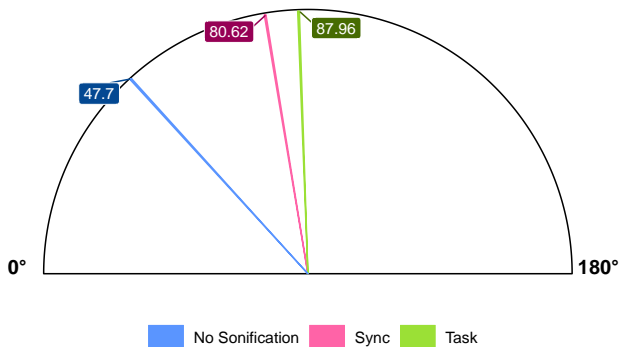**Figure 5:** Plot of mean pairwise phase angles for each subject pair

Figure 6: Plot of mean pairwise phase angles by condition

Characteristic	Beta	95% CI	p-value
condition			
No Sonification	—	—	
Task	0.06	-0.03, 0.14	0.2
Sync	-0.01	-0.09, 0.08	0.9

¹ CI = Confidence Interval

This has "Exponentiate = TRUE", and no "add_global_p"

This compares the angle model with the null angle model

testing another method this has the outputs from all four models in the order null distance, distance, null angle, angle

Simple lm, but grouped with trial, subject pair and condition

this is using the formula `delta_y ~ condition`

Characteristic	Table 1			Table 2		
	Beta	95% CI	p-value	Beta	95% CI	p-value
(Intercept)	0.18	0.10, 0.27	<0.001	0.17	0.08, 0.25	<0.001
condition						0.018
No Sonification				—	—	
Task				0.06	0.01, 0.11	
Sync				0.00	-0.06, 0.05	
No. Obs.	44			44		
Sigma	0.075			0.070		
Log-likelihood	43.0			41.2		
AIC	-79.9			-72.4		
BIC	-74.6			-63.4		
REMLcrit	-85.9			-82.4		
Residual df	41			39		

¹ CI = Confidence Interval

8 Results

term	key	1	2	3	4
(Intercept)	estimate	0.1849614	0.1652955	3.973976	3.862680
(Intercept)	std.error	0.0420439	0.0449666	2.761660	2.954416
conditionSync	estimate	NA	-0.0038204	NA	-1.192065
conditionSync	std.error	NA	0.0261331	NA	1.759635
conditionTask	estimate	NA	0.0613497	NA	1.517692
conditionTask	std.error	NA	0.0261331	NA	1.759635
sd__(Intercept)	estimate	0.0904966	0.0913137	5.963058	5.967414
sd__(Intercept)	std.error	NA	NA	NA	NA
sd__Observation	estimate	0.0754838	0.0702050	4.756061	4.727187
sd__Observation	std.error	NA	NA	NA	NA

Characteristic	Table 1		Table 2	
	exp(Beta)	95% CI	exp(Beta)	95% CI
(Intercept)	53.2	0.24, 11,929	47.6	0.15, 15,572
condition				
No Sonification			—	
Task			4.56	0.14, 144
Sync			0.30	0.01, 9.55

¹ CI = Confidence Interval

References

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

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.

Gerd Schmitz & Alfred O. Effenberg. (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.

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

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

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

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

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