Unheard Potential: Exploring Haptic-Auditory Feedback in Joint Action Tasks

Mitchell McDermott
School of Electronic Engineering and Computer Science
Queen Mary University of London
London, UK

***Abstract*—This study explores the vital role of auditory feedback in joint action tasks and its interplay with haptic cues. Central to human collaboration, joint action tasks involve multiple participants synchronizing their efforts to achieve common goals, such as moving a table, necessitating precise coordination bolstered by neurocognitive mechanisms. While extensive research investigates the role of visual and haptic cues in these tasks, the domain of haptic-auditory cues, particularly in collaborative contexts, still warrants further exploration. Other collaborative tasks, such as rowing, marching, and dancing, demonstrate the effectiveness of a shared perception of pace in fortifying temporal synchronization. Through our controlled experimental setup, dyadic participants engaged via H-Man haptic devices. Our pilot study explored various auditory modalities, translating participants' spatial dynamics into distinctive auditory feedback. The pilot study results particularly emphasized the effectiveness of binaurally presented spatiotemporally discrete auditory cues. In our main study, which combined this modality with mechanical coupling, the findings showed a pronounced enhancement in temporal synchronization in the Haptic-Auditory condition compared to the Haptic-Only setting. Moreover, spatial closeness results strengthened the significance of haptic cues and demonstrated the supplementary role of auditory feedback. This study also observed auditory feedback's potential to modulate the intensity and nature of interactions in joint tasks. Crucially, these findings have practical implications, especially in collaborative systems design, haptic-interactive platforms, and potential applications in rehabilitation settings, underscoring the need for synchronization and mutual understanding between users. While presenting valuable insights, the study also pointed to some limitations, offering directions for future exploration in the field.**

***Index Terms*—Auditory Feedback, Collaborative Joint Action Tasks, Haptic Interface, Multimodal Sensory Perception, Synchronization**

I. INTRODUCTION

Joint action tasks form a quintessential component of human collaboration. It is imperative to understand how these tasks, especially those requiring a strong sense of shared pace, such as dancing, marching, and rowing, result in more temporally precise outcomes. This concept of shared rhythmic synchronization—whether in rowing to a common beat, marching in unison, or dancing to a tune—is more than mere temporal coordination; it reflects the intricate and profound neurocognitive mechanisms underpinning our collective actions.

Human cognition and interaction are vast domains intricately woven with sensory modalities, each contributing profoundly to our motor and perceptual experiences. Joint action tasks are central to understanding these interactions, wherein multiple participants synchronize their endeavors to achieve shared objectives. Such tasks embody the essence of human collaboration, demanding meticulous coordination. These coordinated activities often require synchronicity and are supported by several neurocognitive mechanisms that cognitive scientists propose, such as shared representations, predictive processing, and real-time adaptation. In cognitive science, "shared representation" suggests that one can internally simulate or represent the actions, goals, and perceptions of others within their cognitive system. Predictive processing relates to the brain's ability to predict future events, and real-time adaptation refers to how these predictions are updated based on actual sensory feedback [19], [31].

Recent advancements in motor studies, such as the novel via-point task, challenge traditional beliefs in consistent motor plans. These findings suggest a multifaceted approach where choices are made based on prior experiences and body constraints [43]. Additionally, the advent of artificial partners in research offers controlled environments, facilitating in-depth exploration into human coordination dynamics [42].

Among the sensory modalities, auditory feedback stands out for its widespread influence across several disciplines, from ensemble music, which relies heavily on auditory synchronization [17], [36], to cognitive science's auditory-motor entrainment [6]. Additionally, the confluence of auditory and somatosensory systems demonstrates auditory feedback's profound impact on enhancing somatosensory encoding and tactile actions [11]. Similarly, in the context of haptic collaborative interfaces, auditory feedback has proven influential [13], and it has demonstrated its potency in affecting the behavior of physical entities [16], [27].

Despite the wealth of research emphasizing visual and haptic cues in joint actions [3], the domain of auditory cues, especially in collaborative scenarios, still needs to be explored. While its role in individual tasks is recognized [8], [18], the broader ramifications in cooperative tasks remain an open question. Beyond task dynamics, the judicious utilization of non-speech sounds offers a medium that is natural, ubiquitous, and information-rich, yet unintrusive [10].

Music theory offers additional insights, demonstrating how manipulating musical elements—such as tempo, rhythm, pitch, and harmony—can create various types of auditory feedback.

Each feedback type carries distinct implications for cognitive processing and motor coordination [14], [26]. These insights elucidate the diverse possibilities of auditory feedback, each harboring distinct cognitive and motor implications.

One particularly salient application that stands to benefit from a deeper understanding of auditory feedback in joint tasks is stroke rehabilitation. Music and robot-aided therapy significantly benefit motor training within the rehabilitation context [30]. By delving into the symbiotic relationship between auditory and haptic cues in joint action tasks, we may uncover novel therapeutic strategies that harness this synergy to facilitate improved motor outcomes for stroke patients. This knowledge is a powerful motivation for our inquiry, emphasizing the theoretical significance of our exploration and its profound real-world implications.

This study aims to bridge the research gap by investigating the impact of haptic-auditory feedback in joint action tasks. Our preliminary findings delve into the potential synergies between auditory and haptic cues and examine their combined effects on joint task synchronization. Ultimately, our insights aspire to inform the development of intuitive, collaborative systems and advance therapeutic interventions in stroke rehabilitation.

II. METHODOLOGY: PILOT STUDY

A. Participant Selection and Demographic Data

For the pilot study, we recruited a cohort of 10 participants (5 dyads) from the postgraduate student and Ph.D. scholar population within the Human Robotics Group at Imperial College London. The participants' ages varied from 21 to 34 years, with an average age of 25.

The participants completed a pre-experiment questionnaire, which collected detailed information about participants' demographics, their familiarity with haptic devices and computer games, physical health with a particular focus on auditory and visual health and upper limb pain, musical training, and handedness. We designed the musical training portion of the questionnaire by referencing The Goldsmiths Musical Sophistication Index [23]. One participant reported a slight hearing impairment in the right ear but could still effectively participate in the experiment.

While all but one participant had previous experience with similar experiments involving the H-Man device, it is essential to consider the potential influence of this familiarity on the study's results. Although this prior experience might introduce a certain level of bias, it was deemed negligible for this pilot study. However, in a full-scale study, measures would be taken to ensure a more naive participant pool or to account for this potential bias in the analysis.

All procedures involving human participants were approved by the Imperial College Research Ethics Committee (ICREC). Informed consent was obtained from all individual participants included in the study. All aspects of their participation were explicitly communicated prior to commencement, and they were informed of their right to stop at any time.

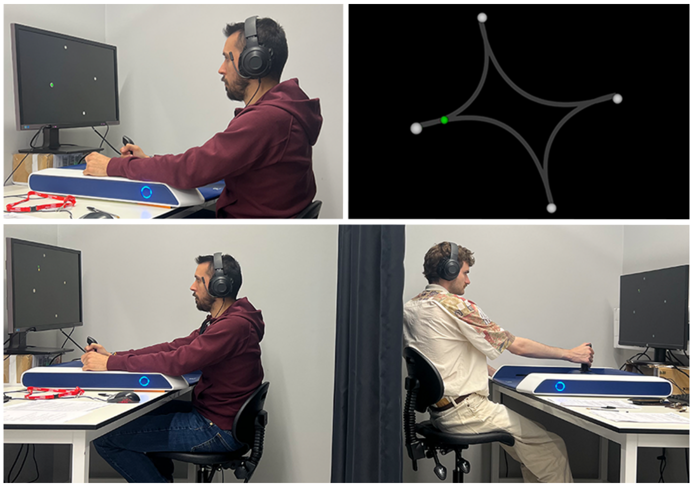

Fig. 1. System Configuration

B. Experimental Design

Both studies adopted a controlled experimental within-subjects design. As shown in **Figure 1** The dyadic participants interacted through H-Man haptic devices, though they were not explicitly informed that they were collaborating with another agent. A 2D virtual environment was presented on screen. A curtain separated the two participants, ensuring physical separation to exclude visual cues and maintain an initial ambiguity about collaboration. Auditory stimuli were delivered via headphones to control the auditory environment.

C. Protocol

The study protocol began with participants positioning themselves at a designated starting point, marked by a red arrow. Upon receiving a signal to start, they were instructed to perform planar point-to-point movements through different via-points clockwise and then return to the starting position. There was no explicit instruction to follow the visual trajectories. Participants can only view their current position, starting, and target points. Players cannot observe one another or converse aloud.

The prior condition consisted of twelve trials, presenting distinct trajectories to each participant. These trials employed an ambient pink noise as the auditory stimulus. The pink noise was ambisonic, where the cursor's visual position correlated to the auditory stimulus's spatial position along a 2D plane.

In the training phase of the experiment, we examined four distinct auditory modalities. The aim was to investigate how varying types of auditory feedback could influence the participants' behavior and perceptions. These auditory conditions were either congruent with the haptic feedback used in a previously unpublished study by colleagues at Imperial College London, providing the same information through a different sensory channel, or introducing new elements into the experimental framework. Each condition was thus a unique exploration of the impact of specific auditory feedback mechanisms on participant behavior.

The order of the auditory modalities was specified, with similar modalities grouped and temporal and pitch-dependent tasks spaced out. We acknowledge the potential for order effects.

After each auditory modality trial, participants completed a feedback questionnaire. This questionnaire captured their subjective experiences with the auditory stimuli and task execution, assessing factors such as predictability, pleasantness, perceived assistance, interactivity, entertainment, clarity, likability, sense of agency, comprehension of the task and the auditory feedback, task demands (mental, physical, and temporal), self-assessment of performance, exerted effort, and experienced frustration.

Intermediary trials between the sets of auditory modality trials served as a control. These trials, which included three trials each of trajectory with ambisonic pink noise and trajectory without noise, were designed to provide a baseline for comparison with the auditory modality trials.

Finally, after completing all trials, participants ranked the four forms of auditory feedback, measuring their preferences and further insight into their subjective experiences with each modality.

D. Auditory Modalities

The auditory modalities used the Pure Data and the HOA library for ambisonics. Each modality presented unique auditory feedback mechanisms to observe their effects on participant behavior. These included Distance Pitch Dissonance, Distance Rhythmic Synchrony, Target Impact Chime, and Speed-Adaptive Tempo. The specifics of each modality were chosen based on existing studies and theories about the relationship between auditory stimuli and human behavior.

1) Distance Pitch Dissonance: In this modality, the difference in the subjects' positions translates into an increased dissonance between two pitches, achieved using three sine wave oscillators. When the position difference is zero, the three pitches are identical, resulting in the perception of a single, stable pitch. As the position difference increases, the pitches are detuned to increase dissonance.

This design choice relies on the understanding that adults typically associate consonance with pleasantness and dissonance with unpleasantness [20], [28], [29], [37], [38]. Therefore, the increasing dissonance is an intuitive signal for the subjects to adjust their positions. The auditory feedback in this modality is congruent with the haptic feedback from the previously referenced experiment.

2) Distance Rhythmic Synchrony: In this modality, the variation in the subjects' positions correlates with changes in rhythmic synchronization, achieved using a stereo percussive drum loop. When the position difference is minimal, the drum rhythms synchronize, resulting in a coordinated and musical pattern. As the difference in positions expands, the signal in the right ear delays progressively, leading to an increasingly chaotic and unmusical rhythm.

This design is grounded in the understanding that humans inherently perceive a regular pulse or beat in an auditory signal, a phenomenon referred to as ”tactus” in music theory, which aids in synchronizing movements with each other [39]. Therefore, the rhythmic synchronization is an intuitive signal for the subjects to adjust their positions. Similar to the previous modality, the auditory feedback aligns with the haptic feedback used in the previous experiment. Furthermore, the salience of a pulse sensation, which is most prominent at moderate tempos, plays a crucial role in distinguishing musical rhythm from non-rhythm [24]. As such, manipulating the salience of the pulse sensation through the synchronization or desynchronization of the drum loop could significantly impact the subjects' perception of rhythm and their ability to coordinate their actions.

3) Target Impact Chime: This modality employs the subtle yet informative auditory feedback mechanism by employing a distinct bell sound when participants surpass a visual target. The central hypothesis is that introducing discrete time visually linked auditory stimuli can accentuate spatial consciousness and rhythm in a joint task environment. Through the punctual chiming sound, participants gain a heightened understanding of their relative pace, catalyzing their ability to synchronize actions and coordinate seamlessly.

Upon surpassing a visual target, participants encounter a distinct bell sound tuned to a specific pitch (e.g., C3 for one participant, E3 for the other). The temporal sequence of these sounds equips the participants with a sense of their speed relative to others. To enhance the localization of these chimes, they are spatially presented through the application of ambisonics, thereby assisting participants in distinguishing the pace-based positioning of one another.

The efficacy of such a mechanism relies heavily on the concept of crossmodal interactions and the consequential influence that auditory stimuli can impose on spatial perception. Increasing attention is being given to exploring multisensory integration and the co-influence of auditory and tactile senses in spatial and non-spatial domains [9], [15], [22], [32], [33].

The utilization of ambisonics to spatially distribute the target chimes permits the subjects to discern each other's relative positioning in correspondence to their pacing. Such an approach, therefore, does not just fortify the spatial consciousness of the participants but also amplifies their ability to coordinate and synchronize their actions.

4) Speed-Adaptive Tempo: In this paradigm, each participant's progress along a predefined angular path dynamically modulates a rhythmic drum beat's tempo. Those advancing faster along the path experience a slowing drum tempo, gently signaling the need to decelerate and maintain synchrony with their partner. Conversely, those lagging hear a progressively quickened drum beat, nudging them to hasten their pace. The change in tempo is scaled to the degree of spatial separation between the participants, creating a sonically intuitive guide for movement coordination.

The foundation for including this modality in our study lies in our understanding of the profound effects auditory feedback, specifically musical tempo, can exert on the speed and timing of human movement. Such feedback can be a game-changer in joint action scenarios, where movement synchronization and coordination are paramount.

Pioneering studies have demonstrated tempo's effect on physical performance and behavior. As Becker et al. [1] observed, faster music could coax subjects into covering longer distances on an exercise bike. Conversely, slower beats led to diminished walking distances [2], illustrating the direct impact tempo can have on the velocity of physical motion. This discovery describes a principle we sought to harness within our experimental design.

Additionally, research suggests that background music's speed can dictate the pace of motor behaviors [7], [21], [34], lending support to our conjecture that the Speed-Adaptive Tempo modality, which adapts the drum beat tempo based on participants' progress, can effectively orchestrate their pace.

III. RESULTS: PILOT STUDY

In our pilot study's findings, the Target-Impact Chime modality emerged as the most effective synchronization, demonstrated by the lowest average position difference among all tested modalities. Additionally, participant feedback resonated with these quantitative results: almost unanimously, participants preferred the Target-Impact Chime modality, with only one exception. This convergence of objective metrics and subjective preferences underscores the potential of the Target-Impact Chime modality in enhancing joint action task performance. Given its evident superiority, we chose the Target-Impact Chime modality for further exploration in the main study.

IV. METHODOLOGY: MAIN STUDY

A. *Participant Selection and Demographic Data*

For the main study, we recruited a cohort of 16 participants (8 dyads). Similar to the pilot study, the selection criteria ensured diverse nationalities and linguistic backgrounds, and included a near-even gender ratio with 9 female and 7 male participants. The participants' ages varied from 20 to 33 years, with an average age of 25.

The participants completed the same pre-experiment questionnaire implemented in the pilot study. No participants had previous experience with similar experiments involving the H-Man device.

B. *Protocol*

The participants were given the same task and setup as in the pilot study. In this study, subjects experience haptic feedback during training trials and auditory feedback in the Target-Impact Chime modality.

First, during the biasing phase (twelve trials), mechanical coupling was disabled, and we presented distinct trajectories to each participant, and each participant acted alone. During the subsequent baseline phase, they performed three trials without trajectories. During the training phase, subjects performed six epochs where they had twenty trials where they were mechanically connected and received auditory feedback. After every twenty trials, participants performed two post-trials with the coupling removed. Similar to the pilot study, ambient pink noise was the auditory stimuli during unconnected trials. The ambient pink noise was not presented binaurally during control trials in this study.

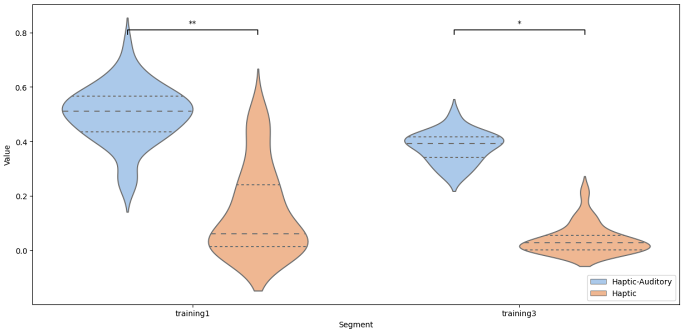

Fig. 2. Distribution of Speed Correlation (Training)

V. RESULTS: MAIN STUDY

Our comprehensive main study used various metrics to assess participants' performances across the Haptic-Auditory and Haptic-Only conditions. For context, the Haptic-Only data originates from a prior experiment that utilized a similar protocol, focusing exclusively on the impact of haptic coupling in joint action tasks.

A. *Training*

1) Speed Correlation (Temporal Alignment): Temporal alignment offers a lens into how participants synchronize their actions over time in joint action tasks. Effective coordination, anticipation of partner's movements, and adjustments in their behaviors to achieve synchronization are hallmarks of improved temporal alignment, pointing to a strong mutual understanding and shared temporal goal. In the Haptic-Auditory condition, there was a notable progression in performance. From Training1 to Training3, the results in **Figure 2** show there was a marked enhancement in speed synchronization (p=4.87e-06, r=-0.7238). In contrast, the Haptic-Only condition displayed a more consistent performance regarding speed synchronization (p=0.5652, r=0.086).

2) Fréchet Distance (Spatial Closeness): Spatial closeness gauges how participants' spatial paths mirror each other during joint action tasks. Participants moving in unison, gravitating towards similar or complementary trajectories, and effectively navigating the task's spatial elements signify enhanced spatial closeness. It reflects mutual spatial understanding and joint spatial goal-setting. Conversely, diverging trajectories suggest participants are pursuing distinct spatial paths, potentially due to competitive tendencies or differing spatial priorities. For spatial closeness, as determined by the Fréchet Distance metric, both conditions—Haptic-Auditory and Haptic-Only—recorded significant improvements **(Figure 3)**. The Haptic-Auditory condition evidenced trajectories becoming

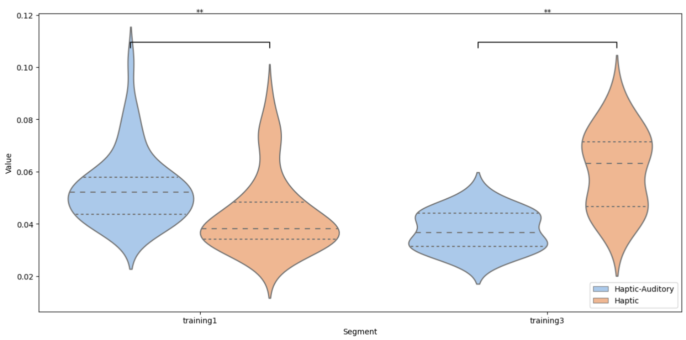

Fig. 3. Distribution of Fréchet Distance (Training)

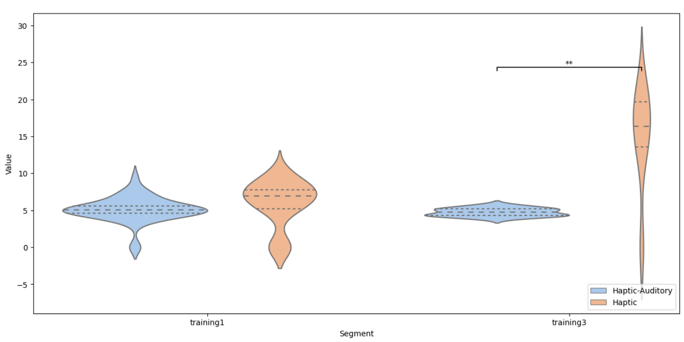

Fig. 4. Distribution of Mean Interaction Force (Training)

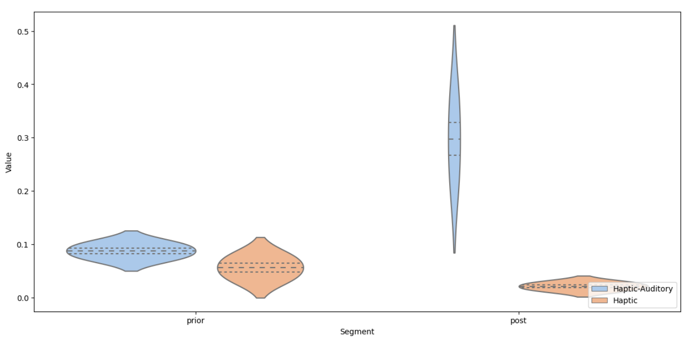

Fig. 5. Distribution of Speed Correlation (Prior and Post)

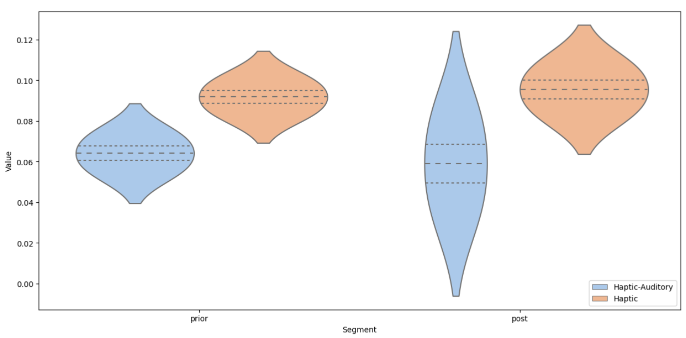

Fig. 6. Distribution of Fréchet Distance (Prior and Post)

more congruent over the training blocks (p=1.74e-06, r=-0.7571). A parallel trend was observed in the Haptic-Only condition, though with a slightly diminished effect size (p=2.12e-05, r=-0.6300). These observations indicate that participants exhibited increasing spatial alignment in their actions over time regardless of auditory feedback.

3) Mean Interaction Force (Interaction Intensity): Mean Interaction Force provides a window into the cooperative or combative nature of participants' interactions during joint action tasks. Lower forces suggest a harmonious collaboration, while higher forces can point to competitive inclinations. When evaluating interaction force, the two conditions showed distinct trends **(Figure 4)**. Within the Haptic-Auditory framework, there was a trend toward reduced interaction force (r=-0.2595). However, it lacked statistical significance (p=0.1023), implying potential directions for a more extensive dataset or additional research. In contrast, the Haptic-Only condition showcased a clear and significant increase in interaction force between Training1 and Training3 (p=2.62e-07, r=0.7613).

B. Prior and Post

Post-training assessments provide a unique perspective to gauge the lasting effects of training by comparing performances before and after the intervention. Our examination focused on temporal alignment and spatial closeness across the Haptic-Auditory and Haptic-Only conditions. Mean interaction force was not compared in this context, as the methodology differed; subjects were not mechanically linked during the prior and post conditions, disqualifying the measurement.

1) Speed Correlation (Temporal Alignment): For the Haptic-Auditory condition, distinctions between the prior and post-training stages were not evident **(Figure 5)**. While the effect size (r=1.0) hinted at a positive shift in the post-condition, this non-significant p-value negates any evidence of impact. In the Haptic-Only scenario, although the data suggested a bias towards the prior condition (r=-1.0), the p-value's lack of significance dampened definitive conclusions.

2) Fréchet Distance (Spatial Closeness): In the context of spatial closeness in the Haptic-Auditory condition, prior and post phases were closely matched, with minimal discrepancies (r=0.0). The Haptic-Only scenario showed a mild inclination toward the post-condition (r=0.5), but this trend did not achieve statistical significance **(Figure 6)**.

3) Subjective Measures: The participant questionnaires revealed no significant change in their perceived sense of mental, physical, or temporal demand, confidence in performance, effort, or frustration. Also, to our surprise, informal post-experiment discussions revealed that most participants did not know they were collaborating with another human being.

In summary, the findings from the main study highlight the multifaceted effects of auditory feedback on joint action tasks. Differences in temporal alignment and interaction forces between the conditions underscore the importance of auditory feedback when crafting cooperative systems, especially where haptic interactions are concerned. These results underscore the nuanced impacts of training and highlight areas warranting further investigation, emphasizing the potential benefits of varied or prolonged training sessions.

VI. DISCUSSION

This investigation aimed to elucidate the role of auditory feedback in joint action tasks, specifically its interaction with haptic cues. The objective was to address a recognized gap in understanding haptic-auditory cues in collaborative scenarios, considering their documented importance in individual tasks.

Our findings on temporal alignment suggest a significant influence of auditory feedback in joint action tasks. In the Haptic-Auditory condition, there were notable improvements in synchronization, which were not as pronounced in the Haptic-Only condition. This data suggests that auditory cues can enhance synchronization in joint tasks, especially when temporally-informing. This finding aligns with prior research that has established the importance of auditory feedback but extends its application to the collaborative domain.

While not statistically significant, the trend observed in the interaction force within the Haptic-Auditory condition remains intriguing. It presents an avenue for further research, especially considering the stark difference juxtaposed against the Haptic-Only condition. The results indicate that auditory cues might influence how forcefully participants interact in a joint task.

The spatial closeness results suggest that while haptic cues play a significant role in joint action tasks, including auditory feedback augments this effect. The difference in interaction force between the two conditions also hints at the potential of auditory feedback to influence interactions, promoting more cooperative behavior.

From an application perspective, the findings have implications for designing collaborative systems or haptic-interactive platforms that involve joint tasks. As demonstrated in this study, the value of temporally significant auditory feedback presented spatially through binaural means could influence designs that prioritize synchronization and mutual understanding between users. This discovery could be especially relevant in scenarios like haptic collaborative interfaces and systems where synchronization between physical entities is crucial.

There are limitations to this study that warrant mention. The participant pool was limited and might not be representative of broader demographics. The potential order effects related to the modality in the pilot study might also play a role in the observed results. While constraining the direct generalization of the findings, these limitations also highlight potential areas for future research.

The chosen auditory feedback modality, the Target-Impact Chime, represents one of many possible auditory cues. Given music theory's breadth and the varied elements of auditory feedback, future studies can explore a broader range of auditory feedback types and understand their implications for cognitive processing and motor coordination. For example, researchers may explore the Speed-Adaptive Tempo modality as our pilot study data suggests it may improve temporal alignment.

VII. CONCLUSION

This investigation delved into the role of auditory feedback in joint action tasks, spotlighting its relationship with haptic cues. An apparent gap in the research landscape drives our motivation, where the importance of haptic-auditory cues is underrepresented in collaborative settings, even though their significance in individual tasks is well-recognized.

The study's pivotal findings reveal:

1. Temporal Alignment: The Haptic-Auditory condition demonstrated superior synchronization compared to the Haptic-Only condition. This result brings to the fore the ability of temporally informative auditory cues to bolster synchronization in joint tasks.

2. Interaction Force: Although not statistically conclusive, the trends indicate a potential role of auditory feedback in shaping force dynamics in joint tasks.

3. Spatial Closeness: The synthesis of auditory feedback amplifies the efficacy of haptic cues in joint tasks, suggesting a propensity for more cooperative behavior.

In practical terms, these discoveries have significant ramifications for the design of joint task systems. The emphasis on temporally apt auditory feedback, especially when spatially presented, promises to enhance synchronization and mutual comprehension amongst users, particularly pertinent in haptic collaborative interfaces.

However, the broader impact of this research transcends its immediate theoretical revelations. A salient application is in the domain of stroke rehabilitation. As the synergy between auditory and haptic cues becomes more evident, it unlocks potential pathways to innovatively leverage this understanding to create more effective rehabilitation interventions for stroke patients. Given the established benefits of music and robot-aided therapy in stroke rehabilitation, our findings offer a promising avenue for integrating these modalities to yield better therapeutic outcomes.

In essence, this study addresses an existing research void concerning auditory feedback in joint tasks and pioneers insights with far-reaching implications, especially in stroke rehabilitation. Exploring auditory and haptic feedback's synergistic relationship holds promise to enrich our theoretical knowledge base and profoundly impact real-world therapeutic strategies.

ACKNOWLEDGMENT

I would like to express my profound gratitude to my supervisor, Dr. Ekaterina Ivanova, at Queen Mary University of London, for her passionate guidance and invaluable feedback throughout this research.

I also wish to acknowledge Dr. Etienne Burdet and the Human Robotics Group at Imperial College London for providing access to their lab and equipment, making this experiment possible. Their support was critical to the completion of my research.

Lastly, I extend my deepest thanks to my dear friends and collaborators from the University of Genoa, namely Dr. Cecilia de Vicariis, Giada Parodi, and Ludovica Viola. Their inspiration, expertise, and contributions were instrumental to the development of this study.

REFERENCES

[1] N. Becker *et al.*, "Mellow and Frenetic Antecedent Music during Athletic Performance of Children, Adults, and Seniors," *Percept Mot Skills*, vol. 79, no. 2, pp. 1043–1046, Oct. 1994, doi: 10.2466/pms.1994.79.2.1043.

[2] N. Becker, C. Chambliss, C. Marsh, and R. Montemayor, "Effects of Mellow and Frenetic Music and Stimulating and Relaxing Scents on Walking by Seniors," *Percept Mot Skills*, vol. 80, no. 2, pp. 411–415, Apr. 1995, doi: 10.2466/pms.1995.80.2.411.

[3] H. Bekkering, E. R. A. De Bruijn, R. H. Cuijpers, R. Newman-Norlund, H. T. Van Schie, and R. Meulenbroek, "Joint Action: Neurocognitive Mechanisms Supporting Human Interaction," *Topics in Cognitive Science*, vol. 1, no. 2, pp. 340–352, 2009, doi: 10.1111/j.1756-8765.2009.01023.x.

[4] S. A. Brewster, P. C. Wright, and A. D. N. Edwards, "An evaluation of earcons for use in auditory human-computer interfaces," in *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, in CHI '93. New York, NY, USA: Association for Computing Machinery, May 1993, pp. 222–227. doi: 10.1145/169059.169179.

[5] A. Caclin, S. Soto-Faraco, A. Kingstone, and C. Spence, "Tactile 'capture' of audition," *Perception & Psychophysics*, vol. 64, no. 4, pp. 616–630, May 2002, doi: 10.3758/BF03194730.

[6] J. E. Crasta, M. H. Thaut, C. W. Anderson, P. L. Davies, and W. J. Gavin, "Auditory priming improves neural synchronization in auditory-motor entrainment," *Neuropsychologia*, vol. 117, pp. 102–112, Aug. 2018, doi: 10.1016/j.neuropsychologia.2018.05.017.

[7] J. Edworthy and H. Waring, "The effects of music tempo and loudness level on treadmill exercise," *Ergonomics*, vol. 49, no. 15, pp. 1597–1610, Dec. 2006, doi: 10.1080/00140130600899104.

[8] N. Ehrlé and S. Samson, "Auditory discrimination of anisochrony: Influence of the tempo and musical backgrounds of listeners," *Brain and Cognition*, vol. 58, no. 1, pp. 133–147, Jun. 2005, doi: 10.1016/j.bandc.2004.09.014.

[9] M. Eimer, D. Cockburn, B. Smedley, and J. Driver, "Cross-modal links in endogenous spatial attention are mediated by common external locations: evidence from event-related brain potentials," *Exp Brain Res*, vol. 139, no. 4, pp. 398–411, Aug. 2001, doi: 10.1007/s002210100773.

[10] W. W. Gaver, "The SonicFinder: An Interface That Uses Auditory Icons," *Human–Computer Interaction*, vol. 4, no. 1, pp. 67–94, Mar. 1989, doi: 10.1207/s15327051hci0401_3.

[11] L. Godenzini, D. Alwis, R. Guzulaitis, S. Honnuraiah, G. J. Stuart, and L. M. Palmer, "Auditory input enhances somatosensory encoding and tactile goal-directed behavior," *Nat Commun*, vol. 12, no. 1, Art. no. 1, Jul. 2021, doi: 10.1038/s41467-021-24754-w.

[12] J. Guo, T. Liu, and J. Wang, "Effects of auditory feedback on fine motor output and corticomuscular coherence during a unilateral finger pinch task," *Frontiers in Neuroscience*, vol. 16, 2022, doi: 10.3389/fnins.2022.896933.

[13] Y. Y. Huang, J. Moll, E.-L. Sallnäs, and Y. Sundblad, "Auditory feedback in haptic collaborative interfaces," *International Journal of Human-Computer Studies*, vol. 70, no. 4, pp. 257–270, Apr. 2012, doi: 10.1016/j.ijhcs.2011.11.006.

[14] M. R. Jones, "Time, our lost dimension: Toward a new theory of perception, attention, and memory," *Psychological Review*, vol. 83, no. 5, pp. 323–355, 1976, doi: 10.1037/0033-295X.83.5.323.

[15] V. Jousmäki and R. Hari, "Parchment-skin illusion: sound-biased touch," *Curr Biol*, vol. 8, no. 6, p. R190, Mar. 1998, doi: 10.1016/s0960-9822(98)70120-4.

[16] C. I. KARAGEORGHIS, P. C. TERRY, and A. M. LANE, "Development and initial validation of an instrument to assess the motivational qualities of music in exercise and sport: The Brunel Music Rating Inventory," *Journal of Sports Sciences*, vol. 17, no. 9, pp. 713–724, Jan. 1999, doi: 10.1080/026404199365579.

[17] P. E. Keller, "Joint action in music performance," in *Enacting intersubjectivity: A cognitive and social perspective on the study of interactions*, in Emerging communication: Studies on new technologies and practices in communication. Amsterdam, Netherlands: IOS Press, 2008, pp. 205–221.

[18] P. E. Keller, "Mental imagery in music performance: underlying mechanisms and potential benefits," *Annals of the New York Academy of Sciences*, vol. 1252, no. 1, pp. 206–213, 2012, doi: 10.1111/j.1749-6632.2011.06439.x.

[19] P. E. Keller, G. Novembre, and M. J. Hove, "Rhythm in joint action: psychological and neurophysiological mechanisms for real-time interpersonal coordination," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 369, no. 1658, p. 20130394, Dec. 2014, doi: 10.1098/rstb.2013.0394.

[20] S. Koelsch, S. Jentschke, D. Sammler, and D. Mietchen, "Untangling syntactic and sensory processing: An ERP study of music perception," *Psychophysiology*, vol. 44, no. 3, pp. 476–490, 2007, doi: 10.1111/j.1469-8986.2007.00517.x.

[21] R. Kuribayashi and H. Nittono, "Speeding up the tempo of background sounds accelerates the pace of behavior," *Psychology of Music*, vol. 43, no. 6, pp. 808–817, Nov. 2015, doi: 10.1177/0305735614543216.

[22] D. M. Lloyd, N. Merat, F. Mcglone, and C. Spence, "Crossmodal links between audition and touch in covert endogenous spatial attention," *Perception & Psychophysics*, vol. 65, no. 6, pp. 901–924, Aug. 2003, doi: 10.3758/BF03194823.

[23] D. Müllensiefen, B. Gingras, J. Musil, and L. Stewart, "The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population," *PLOS ONE*, vol. 9, no. 2, p. e89642, Feb. 2014, doi: 10.1371/journal.pone.0089642.

[24] R. Parncutt, "A Perceptual Model of Pulse Salience and Metrical Accent in Musical Rhythms," *Music Perception*, vol. 11, no. 4, pp. 409–464, Jul. 1994, doi: 10.2307/40285633.

[25] H. L. Pick, D. H. Warren, and J. C. Hay, "Sensory conflict in judgments of spatial direction," *Perception & Psychophysics*, vol. 6, no. 4, pp. 203–205, Jul. 1969, doi: 10.3758/BF03207017.

[26] S. Proksch, D. C. Comstock, B. Médé, A. Pabst, and R. Balasubramaniam, "Motor and Predictive Processes in Auditory Beat and Rhythm Perception," *Front Hum Neurosci*, vol. 14, p. 578546, Sep. 2020, doi: 10.3389/fnhum.2020.578546.

[27] M. Rath and D. Rocchesso, "Continuous sonic feedback from a rolling ball," *IEEE MultiMedia*, vol. 12, no. 2, pp. 60–69, Apr. 2005, doi: 10.1109/MMUL.2005.24.

[28] M. Rigg, "Musical Expression: An Investigation of the Theories of Erich Sorantin," *Journal of Experimental Psychology*, vol. 21, no. 4, p. 442, 1937, doi: 10.1037/h0056388.

[29] D. Sammler, M. Grigutsch, T. Fritz, and S. Koelsch, "Music and emotion: Electrophysiological correlates of the processing of pleasant and unpleasant music," *Psychophysiology*, vol. 44, no. 2, pp. 293–304, Mar. 2007, doi: 10.1111/j.1469-8986.2007.00497.x.

[30] T. Särkämö *et al.*, "Music listening enhances cognitive recovery and mood after middle cerebral artery stroke," *Brain*, vol. 131, no. 3, pp. 866–876, Mar. 2008, doi: 10.1093/brain/awn013.

[31] N. Sebanz, H. Bekkering, and G. Knoblich, "Joint Action: bodies and minds moving together," *Trends Cogn Sci*, vol. 10, no. 2, pp. 70–76, Feb. 2006, doi: 10.1016/j.tics.2005.12.009.

[32] C. E. Sherrick, "Basic and applied research on tactile aids for deaf people: Progress and prospects," *The Journal of the Acoustical Society of America*, vol. 75, no. 5, pp. 1325–1342, May 1984, doi: 10.1121/1.390853.

[33] C. Spence, M. E. R. Nicholls, N. Gillespie, and J. Driver, "Cross-modal links in exogenous covert spatial orienting between touch, audition, and vision," *Perception & Psychophysics*, vol. 60, no. 4, pp. 544–557, Jun. 1998, doi: 10.3758/BF03206045.

[34] A. Szabo, A. Small, and M. Leigh, "The effects of slow- and fast-rhythm classical music on progressive cycling to voluntary physical exhaustion," *The Journal of sports medicine and physical fitness*, vol. 39, pp. 220–5, Oct. 1999.

[35] M. H. Thaut, G. P. Kenyon, M. L. Schauer, and G. C. McIntosh, "The connection between rhythmicity and brain function," *IEEE Engineering in Medicine and Biology Magazine*, vol. 18, no. 2, pp. 101–108, Mar. 1999, doi: 10.1109/51.752991.

[36] R. Timmers, S. Endo, A. Bradbury, and A. M. Wing, "Synchronization and leadership in string quartet performance: a case study of auditory and

visual cues," *Frontiers in Psychology*, vol. 5, 2014, doi: 10.3389/fpsyg.2014.00645.

[37] J. P. Van De Geer, W. J. M. Levelt, and R. Plomp, "The connotation of musical consonance," *Acta Psychologica*, vol. 20, pp. 308–319, Jan. 1962, doi: 10.1016/0001-6918(62)90028-8.

[38] L. Wedin, "A MULTIDIMENSIONAL STUDY OF PERCEPTUAL-EMOTIONAL QUALITIES IN MUSIC," *Scand J Psychol*, vol. 13, no. 1, pp. 241–257, Sep. 1972, doi: 10.1111/j.1467-9450.1972.tb00072.x.

[39] I. Winkler, G. P. Háden, O. Ladinig, I. Sziller, and H. Honing, "Newborn infants detect the beat in music," *Proceedings of the National Academy of Sciences*, vol. 106, no. 7, pp. 2468–2471, Feb. 2009, doi: 10.1073/pnas.0809035106.

[40] D. M. Wolpert, J. Diedrichsen, and J. R. Flanagan, "Principles of sensorimotor learning," *Nat Rev Neurosci*, vol. 12, no. 12, Art. no. 12, Dec. 2011, doi: 10.1038/nrn3112.

[41] J. Yang, P. Sasikumar, H. Bai, A. Barde, G. Sörös, and M. Billinghurst, "The effects of spatial auditory and visual cues on mixed reality remote collaboration," *J Multimodal User Interfaces*, vol. 14, no. 4, pp. 337–352, Dec. 2020, doi: 10.1007/s12193-020-00331-1 n

[42] C. De Vicariis, G. Pusceddu, V. T. Chackochan and V. Sanguineti, "Artificial Partners to Understand Joint Action: Representing Others to Develop Effective Coordination," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 30, pp. 1473-1482, 2022, doi: 10.1109/TNSRE.2022.3176378.

[43] Kodl J, Ganesh G, Burdet E (2011) The CNS Stochastically Selects Motor Plan Utilizing Extrinsic and Intrinsic Representations. PLoS ONE 6(9): e24229. https://doi.org/10.1371/journal.pone.0024229