Temporal detection threshold of audio-tactile delays under conditions of active touch with and without a visual cue

Detjon Brahimaj, Giulia Esposito, Arthur Courtin, Frédéric Giraud, Betty Semail, Olivier Collignon, Andre Mouraux

Temporal detection threshold of audio-tactile delays under conditions of active touch with and without a visual cue

1st Detjon Brahimaj
L2EP
Université de Lille
Lille, France
dbrahimaj@univ-lille.fr

2nd Giulia Esposito
IONS
Université catholique de Louvain
Woluwe-Saint-Lambert, Belgium
giulia.esposito@uclouvain.be

3rd Arthur Courtin
IONS
Université catholique de Louvain
Woluwe-Saint-Lambert, Belgium
arthur.courtin@uclouvain.be

4th Frederic GIRAUD
L2EP
Université de Lille
Lille, France
frederic.giraud@univ-lille.fr

5th Betty SEMAIL
L2EP
Université de Lille
Lille, France
betty.semail@univ-lille.fr

6th Olivier Collignon
IONS
Université catholique de Louvain
Woluwe-Saint-Lambert, Belgium
olivier.collignon@uclouvain.be

7th Andre Mouraux
IONS
Université catholique de Louvain
Woluwe-Saint-Lambert, Belgium
andre.mouraux@uclouvain.be

***Abstract*—While much research has been conducted on multisensory interactions in passive touch, research on how active touch influences how the senses interact remains scarce. Using a haptic surface based on ultrasonic vibrations, we investigated the perception of synchronization of audio-tactile stimuli in active touch. Tactile stimuli were delivered upon sliding the finger, and auditory stimuli followed with a delay ranging from $0-700\,\mathrm{ms}$. In this simultaneity judgment task, two visual conditions were employed: (i) a visual cue showing the location of the active zone on the screen; (ii) a black picture on the screen. We also consider two sliding directions: (i) right-to-left (RTL); (ii) left-to-right (LTR).**

We estimated the psychometric function (threshold and slope) of the ability to judge whether the auditory and tactile stimuli were temporally synchronous. We found a threshold of 201.26 and $211.73\,\mathrm{ms}$ for LTR and 233.3 and $207.23\,\mathrm{ms}$ for RTL with and without visual cues, respectively. We translated temporal delays into distances (mm) using the finger sliding velocity measured for each trial. The results indicate that the simultaneity judgment was independent of sliding velocity.

Our results provide insights into participants' sensitivity in perceiving simultaneous audio-tactile feedback generated during active touch exploration.

***Index Terms*—active touch, surface haptic, audio-tactile temporal perception, sensory synchronization**

I. Introduction

Sound and touch are natural features of the physical environment that surrounds us. These modalities provide essential information on mechanical impacts and vibrations we experience in everyday life, and all this data is combined by the brain. This process, by which humans merge the available information into a unique perceptual event, is known as multisensory integration [1] and has been extensively studied, especially for audio-visual interactions. However, how haptic feedback in the condition of active touch, which implies voluntary, self-generated movement [2], integrates with the other senses is more limited [3] since much of the previous research explored responses to passive touch [4]. Therefore, it is essential to design experiments that allow investigation in active touch. In particular, how the temporal synchronization between tactile and auditory stimuli is perceived in this condition remains unclear.

In the context of multisensory synchronization, it has been

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement No 860114.

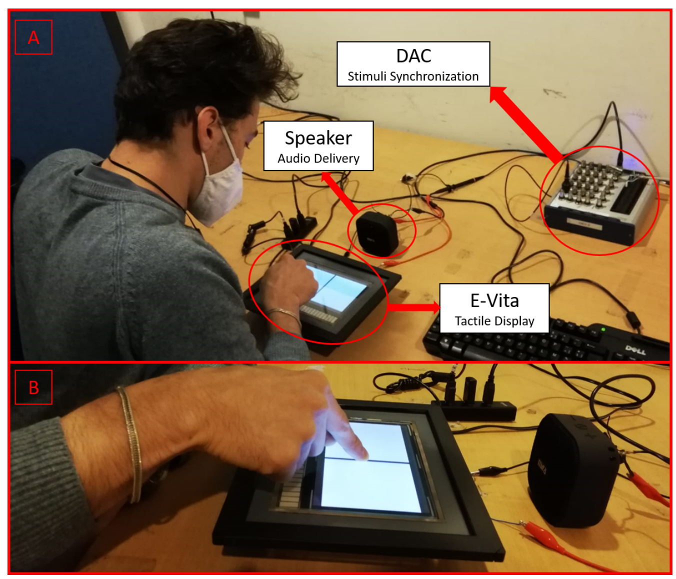

Fig. 1. A) a participant during the experimental session with the Line condition. The main components (E-vita, speaker, and data acquisition card) are highlighted, and the experimental setup is illustrated. B) A zoom on the participant's finger performing the required task.

proposed that the brain monitors temporal coherence among multimodal signals, accounting for possible asynchronies between senses [5]. The advantage of multisensory integration seems closely associated with these temporal relations between the multimodal information and leads to better performances [6].

For example, one of the first investigations that attempted to access temporal perception within auditory and tactile modalities compared people's ability to judge temporal features. The authors [7] presented stimuli either within or across different pairs of sensory modalities and with different stimulus onset asynchronies (SOAs). They used a temporal order judgment (TOJ) task where participants had to judge which stimulus appeared first. Hirsh et al. found that the interval to correctly judge the temporal order was approximately 20 ms for unimodal (touch, auditory, or vision) and multimodal conditions (e.g., audio-tactile).

More recently, Fujisaki et al. [8] investigated the temporal resolution of various modality pairings (audiovisual, audio-tactile, and visual-tactile). Participants were asked to determine if the presented stimuli (single stimulus or repetitive trains) were synchronous or asynchronous. The authors found that audio-tactile pairing was the most reliable for asynchrony detection ($15.6\,\mathrm{ms}$) as compared to other sensory pairings, consistent with the study of Hirsh [7].

Other studies have also examined the role of attention in audio-tactile synchronization. For example, in a study by Spence et al. [9], participants were presented with pairs of auditory and tactile stimuli while performing a visual task. The authors found that when the visual task was easy, participants were able to accurately perceive the temporal order of the auditory and tactile stimuli. However, when the visual task was more difficult, their ability to accurately perceive the temporal order of the stimuli was impaired. This result suggests that attentional resources play a role in the perception of audio-tactile synchronization.

The relative spatial position of stimuli can also modulate the temporal perception of the combined audio-tactile modalities. On this aspect, Ocelli et al. [10] investigated audio-tactile temporal perception with blind and sighted individuals. They compared the co-location of the stimuli against the spatial separation of the stimuli (e.g., stimuli delivered at different locations). Their results demonstrated that the performance of sighted individuals was not affected by the space separation or co-location of the stimuli. At the same time, blind participants were significantly more accurate when the two stimuli were presented from different spatial positions rather than co-located. This result sustains the hypothesis that the lack of a visual cue or visual information, as stimuli presented outside the participant's range of vision [11], is related to more dominant audio-tactile spatial interactions than those arising in the presence of visual information [12]. However, we must consider that blind individuals are not equivalent to sighted individuals temporarily deprived of visual information. Caution is therefore needed when generalizing these results to sighted individuals.

Hence, several characteristics of temporal-perceptual asynchrony have been proposed for some modality pairings. However, the process that governs cross-modal temporal perception is far from being fully understood [13]. At the same time, it represents a critical element of any multimodal feedback system. As the asynchrony between modalities increases, the sense of realism and presence decreases, producing sensory conflicts between modalities and discomfort for the user. Indeed, an appropriate timing relation between modalities is crucial for a congruent perception of multimodal stimuli. Therefore, it is essential to investigate and understand the perceived simultaneity of multimodal stimuli, such as audio-tactile stimuli, especially for active touch.

This paper presents the results of an experiment assessing the ability of individuals to detect temporal delays between tactile and auditory stimuli, in conditions of active touch. To this aim, we performed a psychophysical experiment in which auditory stimuli were delivered with varying delays relative to tactile stimuli to determine the threshold and slope of the psychometric function (PF) describing the relationship between audio-tactile delay and the probability of detecting the stimuli as asynchronous. Furthermore, we investigated whether a visual cue co-localized with the tactile stimuli had an effect on the ability to discriminate audio-tactile synchronicity.

This investigation is of interest in various fields, including psychology, neuroscience, and engineering, as it can provide insight into how the brain processes sensory information and how different sensory modalities interact. Our results can also help multimodal designers better understand the temporal aspects of audio-tactile interaction and therefore design multimodal experiences or devices with a proper timing relation between these modality pairs.

II. Method

To assess the sensitivity of individuals and evaluate how well participants can detect spatiotemporal delays between pairs of auditory and tactile stimuli, we employed a forced choice method where participants were asked to perform a Simultaneity Judgment (SJ) task. Stimulus placement was continuously optimized using the psi method [14].

The Analysis was performed using R version 4.1.1 and the state-of-the-art platform Stan for Hierarchical model fitting [15]. Stan is a C++ library for Bayesian inference using the No-U-Turn sampler (a variant of Hamiltonian Monte Carlo) or frequentist inference via optimization.

A. Experimental setup

Stimuli were delivered using a visual-tactile display named E-vita. Developed by Vezzoli et al. [16], E-vita (Enhanced Visual-Tactile) is a flat Haptic Surface based on low-frequency friction modulation through ultrasonic actuation. With this technology, it is possible to modulate the friction between the fingertip and the vibrating plate using acoustic waves thanks to a phenomenon called active lubrication [17] [18]. Transient changes in friction produce a naturalistic tactile input generated by the mechanical interactions between the display and sliding fingertip [19] [20].

E-vita is built around a Banana Pi (Shenzen LeMaker Technology Co. Ltd, China) single-board computer and presents a $1\,\text{GHz}$ ARM Cortex-A7 dual-core CPU with $1\,\text{GB}$ of RAM. The Banana Pi works in parallel with a microcontroller (STM32F4) responsible for the control of the tactile plate. The tactile stimulator comprises 23 piezoceramic components that actuate a $123 * 165 * 2\ mm^3$ fixed glass plate with a resonating sinusoidal mode-shape at $60750\ Hz$, where the half wavelength is $8\ mm$. 20 piezoceramic components are used as actuators. In contrast, three actuators measure the vibration of the plate provided to the microcontroller. Moreover, an IR (infrared) frame tracks the finger position and velocity, and this information is provided to the microcontroller, which synthesizes a Pulse Width Modulation signal to pilot a voltage inverter that actuates the piezoceramics. The ultrasonic signal provided to the piezoceramics is modulated at a lower frequency ($250\,\text{Hz}$) in order to be perceivable by the mechanoreceptors [21]. Furthermore, E-Vita connects to a 7-inch touchscreen, which can display images according to the experiment. Also, it can detect the finger position and, thus, calculate its velocity and produce tactile feedback only on an area previously selected, denoted 'active zone' in the remaining paper.

To ensure a low latency in the system's actuation, we coupled E-vita with a Data Acquisition Card (DAC, National Instrument - USB 6363). The tactile stimuli delivered by E-Vita are used as a trigger for the DAC. Once triggered, the DAC can deliver auditory feedback with a maximum latency between the two modalities of around $3.9\,\mu\text{s}$, which is negligible in our experiment.

B. Stimuli

The stimuli selected for tactile and auditory feedback and the two visual conditions are described hereafter.

Tactile:

The tactile stimulus used in this experiment consists of a sinusoidal signal with a spatial period of $5000\,\mu\text{m}$ and an amplitude of 40 % (Relative Voltage). The ultrasonic signal (square wave) is modulated with the above sinusoidal waveform at about $200\,\text{Hz}$. This is has been done as ultrasonic frequencies are not perceived by the mechanoreceptors.

The tactile active zone is a rectangle with a height of $100\,\text{mm}$ and a width of $2.5\,\text{mm}$ centered in the middle of the haptic surface. Hence, tactile feedback was delivered while sliding across the plate when the participant's finger reached the middle of the screen. The co-localized visual-tactile screen is illustrated in Fig. 1, and the active zone is represented by the black line centered on the screen.

Auditory:

Auditory stimulus is delivered through a speaker placed behind the tactile stimulator and in front of the participants. The speaker's location was chosen to maximize the impression that the audio signal came from the interaction of the finger with the device.

The selected signal for auditory feedback was a short burst of white noise with a duration of $50\ ms$. The white noise signal was selected to avoid performance biases due to the choice of a specific frequency. Indeed, a participant may be better than others on the SJ task due to a more sensitive hearing at specific frequencies rather than a higher ability for synchronicity discrimination.

In order to ensure a similar perceptual magnitude for the auditory and tactile stimuli, thus avoiding biases due to one stimulus being perceived much stronger than the other, the intensity of both stimuli were matched before the actual data collection started. To do so, we used a classical staircase with constant tactile and varying auditory intensity. Participants were asked to report which of the two stimuli (delivered simultaneously) was stronger. The staircase procedure ended after 10 reversals occurred and the auditory stimuli intensity was set to the average of the last 9 reversals for the remainder of the experiment. Finally, once triggered, the audio is delivered with or without a selected delay depending on the trial. More information on the selected delays is provided in the next section.

Visual condition:

In order to investigate the influence of the presence of a visual cue on audio-tactile stimuli perceived synchronization, two visual conditions (V) were selected for our experiment:

- Black Screen (B): This first visual condition represents our control condition. In this case, no visual cue was provided to the participants, and the screen was a full-black picture.
- Line (L): This second visual condition was a white picture with a black line in the middle. This display was located just below the haptic surface to provide a visual cue that overlapped the active zone (see 1.B)

C. Experimental Protocol

Participants:

We recruited 14 participants (six males and eight females, Age M=28.43 STD=6.72) for this experimental study. Ten participants were right-handed, and four were left-handed.

All participants confirmed they had no current history of neurological or psychiatric disorders and no loss of skin sensitivity in their hands. All participants confirmed not to have any hearing loss or disorders.

The study received ethical approval from the Comité d'Ethique Hospitalo-Facultaire Saint-Luc UCLouvain, and all participants provided written informed consent and were remunerated for their participation in the study.

We excluded the data of one participant due to an issue with the velocity recording system.

Procedure:

Participants sat comfortably in front of the setup in a quiet, dimly lit room. We then allowed participants to test the device to familiarize themselves with the system. They explored the tactile plate using the index of their dominant hand from one side to the other of the haptic surface. To investigate whether the direction of the movement led to a bias in the perception of synchronicity, participants alternated between a left-to-right (LTR) and a right-to-left (RTL) swipes. For every sliding gesture, a tactile stimulus was delivered in the middle

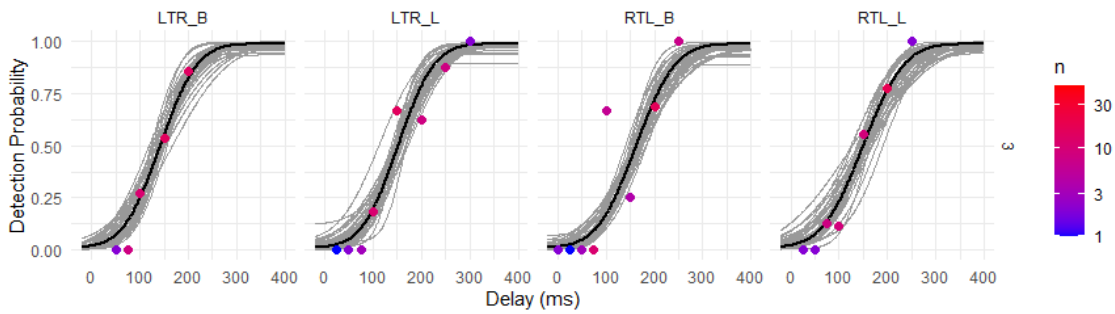

Fig. 2. Individual participant fit of the model (participants 3). For each condition, the detection probability for the various delays are depicted as dots, color-coded based on the number of trials used to compute the plotted probability. The PF constructed using the most likely parameter (median) is plotted in black. 100 PFs constructed using random draws from the posterior distribution of the parameters are plotted in grey (uncertainty around the true parameter values).

of the tactile plate (independently of the sliding direction). When the participant's finger reached the active zone (middle of the tactile plate), an auditory stimulus was delivered with different SOAs, chosen during preliminary tests (0, 25, 50, 75, 100, 150, 200, 250, 300, 350, 400, 500, 600, 700 ms). The SOA was varied on a trial-by-trial basis using the psi method, a Bayesian adaptive algorithm that optimizes stimulus placement to maximize information gain about the PF parameters [14].

Before performing the sliding task and starting the experiment, participants were asked to use talc powder on the index of their dominant hand. The talc powder was used as it allows a homogeneous friction reduction independent of the finger's velocity, as highlighted by the results of Weal et al. [22]. Participants performed a SJ task where, at the end of each slide (50 trials for each condition), they had to state whether they perceived the tactile and auditory stimuli to be synchronous or not.

The order of the two visual conditions was counterbalanced across participants to avoid biases.

In total, the experimental session was 30 min long for each participant with each trial aproximately 10 s. The entire experiment comprised four independent psi procedures, two without a visual cue (LTR_B and RTL_B) and two with a visual cue (LTR_L and RTL_L). Thus, a value of threshold and slope was estimated for each movement direction and visual condition separately.

D. Measures and Analyses

Since, within participants, separate psi algorithms were run for the four experimental conditions. The resulting estimates of threshold and slope parameter values obtained with this method assume independence between conditions. This assumption does not seem reasonable, as participants can be expected to have a baseline threshold/slope which is moderately modulated by conditions rather than a threshold/slope that is completely driven by each condition. As a result, it is difficult to distinguish the parts of the estimates obtained with the psi method associated with the participants and those associated with the conditions. To address this issue, we fitted a new model that takes the entire structure of the data into account (conditions within participants, participants within the population). Hierarchical models are well-suited for such structured data [23].

In this case, we used a hierarchical multiple probit (i.e., cumulative normal) regression model. This model was preferred to its logistic counterpart for theoretical reasons as it is "perhaps the most justifiable form" of psychometric function as, if we assume sensory noise to be the sum of many independent noise sources, it should be approximately normally distributed per the Central Limit Theorem [24].

The threshold of the PF was modeled as a linear combination of subjects random intercepts and fixed and random slopes for the factors *visual condition* (V), *movement direction* (D) and the interaction *visual condition* x *movement direction* (VxD). The slope of the PF was modeled as the base 10 exponential of a similar linear combination (constraining it to be positive). The guess and lapse rates were treated as nuisance parameters, varying between participants but not between conditions.

Model fitting and diagnosis:
A total of 438 stimulus-response pairs, coming from 13 participants, and four conditions were used to fit our model with the Rstan toolbox [15] [25].

Five chains of 4000 iterations (including a warm-up period of 2000 iterations, which was discarded) were used, leading to a total of 10000 draws from the posterior distribution.

Generic non-informative gaussian hyper-priors were used for the threshold ($\mu = 0$ and $\sigma = 100$, except for the threshold intercept for which $\sigma = 200$) and slope ($\mu = 0$ and $\sigma = 0.5$) coefficients. Priors for the guess and lapse rates were beta distributions selected to keep most of the probability mass close to 0 but allow for non-zero values ($\alpha = 1$, $\beta = 50$).

Appropriate sampling of the posterior distributions was assessed using the ShinyStan (Stan Development Team, 2017) package based on the following diagnostic criteria: absence of divergent transitions; no reaching of maximum tree depth; good alignment of energy diagnostic plots, E-BFMI larger than 0.2; adequate sample size larger than 10% of the total sample size; Monte Carlo standard error smaller than 10% of the posterior standard deviation; and $\hat{R}$ equal to or smaller than 1.01. Individual participants' data-fits were inspected visually as a posterior predictive check (an example of one participant

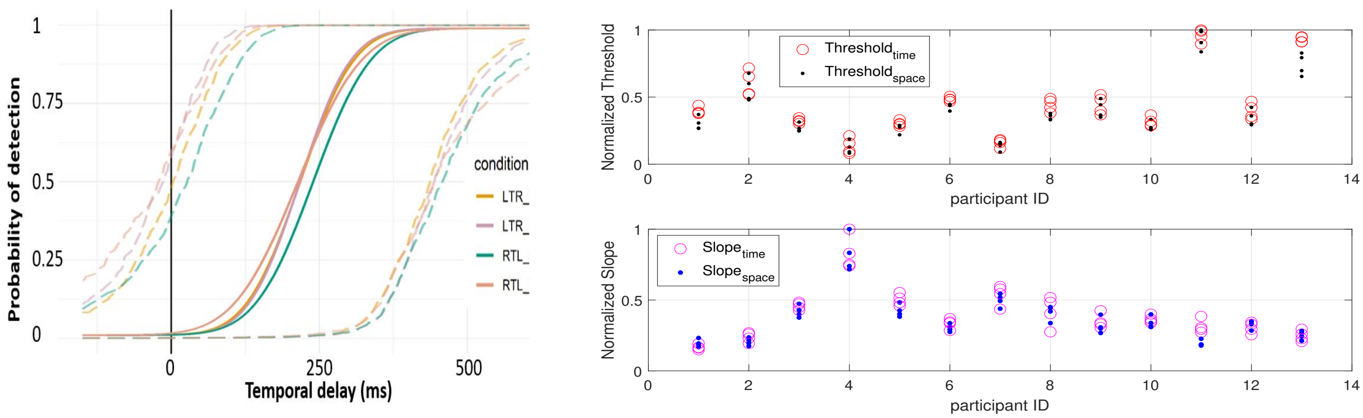

Fig. 3. **Left)** Posterior expected psychometric function. Bold lines represent the PF constructed with the most likely parameter values (i.e. the median of 10000 random draws from the posterior distribution of the model parameters). Dotted lines represent the uncertainty around the posterior expected PF (95% highest probability density intervals).
Right) Normalized threshold (top) and slope (bottom) values for each participant and condition. This Comparison is between values calculated considering time delay or space (i.e., time delay and sliding velocity). Participants that have threshold/slope values close to the maximum (i.e., 1) means that they have the higher threshold/slope among all the participants.

is depicted in Fig. 2).

Between conditions differences:
The presence of significant effects of factors V, D or VxD was determined using bootstrapping of the posterior coefficient estimates (10^4 samples per test, two-sided tests). Pairwise comparisons were conducted using the same bootstrapping approach on posterior marginal means estimates.

Lastly, to better illustrate the PF fitted from the posterior distribution of the model parameters, an expected PF was constructed for each condition. This was achieved by constructing 10^4 PFs using random draws from the posterior distribution of the parameters and taking, for each temporal delay, the 50^{th} percentile of the stimulus detection probability. To get a sense of the uncertainty around these values, the 2.5^{th} and 97.5^{th} percentiles (95% highest probability density intervals) are also plotted. These PFs can be interpreted as the expected values for a new unobserved participant coming from the same population.

Two hierarchical models were fitted to the data, one using the temporal delays (ms) between tactile and auditory stimuli and one using the distance (mm) between the center of the screen (tactile stimulus location) and the position of the finger when the auditory stimulus was delivered.

A repeated measure ANOVA was used to assess whether the finger velocity differed between conditions. We also perform a statistical analysis (repeated measure ANOVA) on the normalized threshold (slope) to see if the finger velocity has an influence on the estimated parameters. The normalization was performed to have threshold (slope) values in the range from 0 to 1 and therefore compare delays with distances.

III. Results

The model diagnostics revealed appropriate sampling, and the model appeared to fit well with the participants' raw data. We show an example of a single participant in Fig. 2.

A. Threshold and slope

Posterior probabilities of effects of *visual condition* and *movement direction* on threshold revealed a main effect of *movement direction* ($p = 0.003$), as well a significant *visual condition* x *movement direction* interaction ($p = 0.014$). No main effect of the *visual condition* was identified.

During pairwise comparison, the thresholds for the black screen conditions with movements from left-to-right or right-to-left appeared to be significantly different ($p = 0.003$). All other comparisons were not significant.

A summary of posterior threshold parameter values for the different conditions is shown in Table I.

TABLE I
Posterior estimates for group-level threshold in the different experimental conditions.

Visual Condition	Movement direction	Threshold (ms)	[95% HPDI] (ms)
B	LTR	211.73	[140.08; 283.65]
L	LTR	210.26	[136.48; 296.71]
B	RTL	233.30	[173.13; 315.60]
L	RTL	207.23	[141.80; 294.43]

Posterior probabilities of effects of *visual condition* and *movement direction* on slope revealed no main effect of *movement direction*, no main effect of *visual condition*, and no *visual condition* x *movement direction* interaction.

A summary of posterior slope parameter values for the different conditions is shown in Table II.

We illustrate in Fig. 3 a visual representation of the model fitting to the expected PF from a new participant from the same population, with both median values and uncertainties.

B. Finger Velocity

Overall, participants' finger velocities appeared independent of the *visual condition* or *movement direction* with a relatively small standard deviation, as depicted in Table III.

TABLE II
POSTERIOR ESTIMATES FOR GROUP-LEVEL SLOPE IN THE DIFFERENT EXPERIMENTAL CONDITIONS.

Visual Condition	Movement direction	Slope	[95% HPDI]
B	LTR	0.014	[0.010; 0.020]
L	LTR	0.015	[0.010; 0.022]
B	RTL	0.013	[0.009; 0.019]
L	RTL	0.012	[0.008; 0.018]

By performing a repeated measure one-way ANOVA, our analysis did not reveal any statistical difference for finger velocity among the four conditions ($p = 0.0567$, $p = 0.5184$, $p = 0.1920$).

TABLE III
FINGER VELOCITIES: MEAN AND STANDARD DEVIATION FOR ALL OUR CONDITIONS

	LTR_B	RTL_B	LTR_L	RTL_L
Mean (mm/s)	27.9148	29.7789	28.5927	29.4776
STD (mm/s)	2.4440	2.3008	3.0481	3.2110

This result suggests that participants used a finger velocity that was not dependent on the condition. Therefore, the perceived delays are not influenced by the exploration velocity of participants but depend only on the temporal aspect between the two stimuli.

To confirm the independency of the sliding velocity, we compare the normalized values of threshold and slope estimated with both delay (ms) and distance (mm), as shown in Fig. 3.

A visual inspection and a statistical analysis (repeated measure one-way ANOVA) on the normalized threshold and slope calculated in time and space confirmed the independence of the finger velocity. Hence, analyzing the performance of participants in relation to the spatial distance between stimuli does not provide additional information compared to the temporal delay analysis.

IV. DISCUSSION

Using psychophysics, we explored audio-tactile temporal interactions under conditions of active touch. To our knowledge, this is the first experiment investigating the maximum delay necessary for auditory and tactile stimuli (delivered during free exploration of a tactile display) to be perceived as synchronous in conditions of active touch.

Our results show that movement direction influences the ability of participants to perceive these delays, right-to-left swipes leading to significantly larger thresholds than left-to-right swipes. This difference could be related to the fact that participants reported swiping their finger in that direction to be easier, possibly because, for all participants, left-to-right was the direction of writing. Interestingly, the presence of a visual cue, indexing the position of the tactile stimulus, did not seem to systematically alter the ability of participants to detect temporal delays between the stimuli. However, the effect of the direction of movement seemed to be mediated by the presence or absence of the visual cue, the difference between RTL and LTR conditions being much more pronounced in the black screen condition than in the line condition. This could indicate that in familiar situations (LTR), the individuals don't use the additional visual information but that they do in situations for which they have less prior experience/skills.

The acceptable delay we found in our study ($200\,\mathrm{ms}$) is much greater than what was found, for example, in [7] ($20\,\mathrm{ms}$) with a factor of 10. This large difference could be partially but not exclusively explained by differences between the tasks used by previous authors and ours. Moreover, another important difference is that the tactile condition we investigated was active touch, which may involve processes that are not activated when passive touch is under investigation. For example, the voluntary movement could affect temporal perception. Indeed, movements could result in a bias or a more precise temporal estimation, as reported by [26]. Nevertheless, the processes that govern how we perceive time across different senses still need to be better understood [13] and requires further investigation.

Regarding the effect of finger sliding velocity, our empirical data show that participants had similar velocities regardless of the visual condition or movement direction. This result suggests that the finger exploration speed, chosen freely by the participants in our active touch experiment, may be related to the SJ task rather than the movement direction or the visual condition. We believe that participants chose the speed that allowed them to obtain the maximum information related to audio-tactile synchronization independently of the condition. Similarly, in rough macro-textures, the finger velocity does not affect perception [27]. However, in our experiment, no textures were explored, and the tactile information was informative of the event itself (I felt something on the fingertip). The finger sliding velocity's independence may be due to the nature of the tactile stimuli we chose. Moreover, by translating time delays into spatial distances (mm) and considering sliding velocity at each trial, pairwise comparisons did not reveal statistically significant differences for both visual conditions and direction of movement (for both slope and threshold), confirming the idea of finger speed independence. Nevertheless, our result is important as it shows a dissociation between the finger speed and the audio-tactile synchronization.

V. CONCLUSION

We show that sensitivity to spatiotemporal audio-tactile delays can be investigated using a novel paradigm.

Overall, for developing audio-tactile haptic technologies, a delay of about $200\,\mathrm{ms}$ between tactile and auditory signals should be acceptable for users to experience synchronicity between the two modalities. These results suggest that a visual cue overlapping with the tactile feedback does not always influence the participant's detection of asynchronies.

These results will help the community and multisensory designers shed light on auditory-tactile synchronization in active touch.

REFERENCES

[1] N. P. Holmes and C. Spence, "Multisensory integration: Space, time and superadditivity," *Current Biology*, vol. 15, no. 18, pp. R762–R764, 2005.

[2] C. E. Chapman, *Active Touch*, pp. 35–41. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.

[3] S. Soto-Faraco and G. Deco, "Multisensory contributions to the perception of vibrotactile events," *Behavioural Brain Research*, vol. 196, no. 2, pp. 145–154, 2009.

[4] J. R. Timora and T. W. Budd, "Steady-State EEG and Psychophysical Measures of Multisensory Integration to Cross-Modally Synchronous and Asynchronous Acoustic and Vibrotactile Amplitude Modulation Rate," *Multisensory Research*, vol. 31, no. 5, pp. 391–418, 2018.

[5] A. Kopinska and L. R. Harris, "Simultaneity constancy," *Perception*, vol. 33, no. 9, pp. 1049–1060, 2004.

[6] M. O. Ernst and M. S. Banks, "ErnstBanks-Nature2002," *Nature*, vol. 415, no. 6870, pp. 429–433, 2002.

[7] I. J. Hirsh and C. E. Sherrick, "Perceived order in different sense modalities," *Journal of Experimental Psychology*, vol. 62, no. 5, pp. 423–432, 1961.

[8] W. Fujisaki and S. Nishida, "Audio-tactile superiority over visuo-tactile and audio-visual combinations in the temporal resolution of synchrony perception," *Experimental Brain Research*, vol. 198, no. 2-3, pp. 245–259, 2009.

[9] C. Spence, J. Driver, and J. C. Driver, *Crossmodal space and crossmodal attention*. Oxford University Press, 2004.

[10] V. Occelli, C. Spence, and M. Zampini, "Audiotactile temporal order judgments in sighted and blind individuals," *Neuropsychologia*, vol. 46, no. 11, pp. 2845–2850, 2008.

[11] N. Kitagawa, M. Zampini, and C. Spence, "Audiotactile interactions in near and far space," *Experimental brain research*, vol. 166, no. 3, pp. 528–537, 2005.

[12] M. Zampini, T. Brown, D. I. Shore, A. Maravita, B. Röder, and C. Spence, "Audiotactile temporal order judgments," *Acta psychologica*, vol. 118, no. 3, pp. 277–291, 2005.

[13] V. Occelli, C. Spence, and M. Zampini, "Audiotactile interactions in temporal perception," *Psychonomic Bulletin and Review*, vol. 18, no. 3, pp. 429–454, 2011.

[14] L. L. Kontsevich and C. W. Tyler, "Bayesian adaptive estimation of psychometric slope and threshold," *Vision Research*, vol. 39, no. 16, pp. 2729–2737, 1999.

[15] Stan Development Team, "Stan modeling language users guide and reference manual, version 2.17.0," 2017.

[16] Y. Rekik, E. Vezzoli, L. Grisoni, and F. Giraud, "Localized haptic texture: A rendering technique based on taxels for high density tactile feedback," *Conference on Human Factors in Computing Systems - Proceedings*, vol. 2017-May, pp. 5006–5015, 2017.

[17] T. Watanabe and S. Fukui, "Method for controlling tactile sensation of surface roughness using ultrasonic vibration," *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 1, pp. 1134–1139, 1995.

[18] E. Vezzoli, Z. Vidrih, V. Giamundo, B. Lemaire-Semail, F. Giraud, T. Rodic, D. Peric, and M. Adams, "Friction Reduction through Ultrasonic Vibration Part 1: Modelling Intermittent Contact," *IEEE Transactions on Haptics*, vol. 10, no. 2, pp. 196–207, 2017.

[19] M. Biet, F. Giraud, and B. Lemaire-Semail, "Squeeze film effect for the design of an ultrasonic tactile plate," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 54, no. 12, pp. 2678–2688, 2007.

[20] T. Sednaoui, E. Vezzoli, B. Dzidek, B. Lemaire-Semail, C. Chappaz, and M. Adams, "Friction Reduction through Ultrasonic Vibration Part 2: Experimental Evaluation of Intermittent Contact and Squeeze Film Levitation," *IEEE Transactions on Haptics*, vol. 10, no. 2, pp. 208–216, 2017.

[21] G. Gescheider, *Psychophysics: The Fundamentals*. Taylor & Francis, 2013.

[22] W. Ben Messaoud, "Design and Control of a Tactile Stimulator for Real Texture Simulation : Application to Textile Fabrics Thèse Design and Control of a Tactile Stimulator for Real Texture Simulation : Application to Textile Fabrics," no. June, p. 147, 2016.

[23] A. Gelman and J. Hill, "Data analysis using regression and multilevel/hierarchical models," 2007. Includes bibliographical references (pages 575-600) and indexes.

[24] F. Kingdom and N. Prins, *Psychophysics: a practical introduction, Second Edition*. Academic Press, 2016.

[25] Stan Development Team, "shinystan: Interactive visual and numerical diagnostics and posterior analysis for bayesian models.," 2017. R package version 2.4.0.

[26] R. Kock, K. Gladhill, M. Ali, W. Joiner, and M. Wiener, "How movements shape the perception of time," *Trends in Cognitive Sciences*, vol. 25, 09 2021.

[27] M. K. Saleem, C. Yilmaz, and C. Basdogan, "Tactile Perception of Virtual Edges and Gratings Displayed by Friction Modulation via Ultrasonic Actuation," *IEEE Transactions on Haptics*, vol. 13, no. 2, pp. 368–379, 2020.