SPATIO-TEMPORAL UNFOLDING OF SOUND SEQUENCES

Davide Rocchesso
IUAV - University of Venice
roc@iuav.it

Stefano Delle Monache
IUAV - University of Venice
stefano.dellemonache@gmail.com

ABSTRACT

Distributing short sequences of sounds in space as well as in time is important for many applications, including the signaling of hot spots. In a first experiment, we show that the accuracy in the localization of one such spot is not improved by the apparent motion induced by spatial sequencing. In a second experiment, we show that increasing the number of emission points does improve the smoothness of spatio-temporal trajectories, even for those rapidly-repeating pulses that may induce an auditory-saltation illusion. Other indications for auditory-display designers can also be drawn from the experiments.

1. INTRODUCTION

Everyday environments are populated of organisms and artefacts that constantly signal their presence and their state. Often, sound is exploited as the preferred channel to display the presence and the location of objects. For example, the Sonic Keyfinder [1] is a small key-ring that can be attached to any object, like bags, canes, remote controls, and reacts with bleeps to whistle or any loud noise, like shouts. Ordinary objects become sonically augmented and may exploit the human ability to locate a sound source in the physical space to communicate their location, and call attention. This scenario becomes intriguing for those objects that are provided with embedded computational affordances. For example, there are companies producing systems capable of charging mobile devices by proximity, without using any plugs. These systems rely on a charging hotspot that is usually embedded into furniture or dashboards and signaled by visual cues. In many circumstances, for aesthetic reasons or to avoid visual distraction, it would be preferable to use non-visual cues to signal a hotspot. However, auditory spatial resolution is poor [1] and, therefore, some degree of exploration is necessary. The apparent motion of a sound source may affect the instantaneous localization of sound events [2]. In this work, we are interested in measuring the quantity and quality of such apparent-motion effects in ecological conditions.

[1] http://www.youtube.com/watch?v=7FJVNOW7aOo&NR=1&feature=fvwp

The paper has the following structure: In Section 2 the literature on non-visual mis-localization of stimuli in space is reviewed. In Section 3 two experiments, the first on localization accuracy, and the second on spatio-temporal sonic gestures, are described and discussed, in terms of implications for design. In Section 4 we draw our conclusion.

2. THE TACTILE RABBIT AND AUDITORY SALTATION

There are some non-visual illusions that show how humans consistently mis-localize stimuli in space when these are presented under certain temporal constraints. In particular, the cutaneous rabbit effect occurs when stimulating the skin at different points in a temporal sequence, if the temporal interval between two stimuli is small and their actual displacement is large. In such case the perceptual system consistently underestimates inter-stimulus distance and over-estimates inter-stimulus time. This illusion is correctly predicted by a Bayesian model that incorporates prior expectation for speed [3]. These effects have been recently exploited in product design, to actuate a jacket with vibration motors that are sparsely located on a large area of the body [4]. Thanks to the "rabbit", a small number of "actuators can create the sensation that the arm is being tapped in several spots between the motors" [5]. It has also been shown how the duration of vibration bursts and the inter-onset-interval affect the experienced continuity and pleasantness of tactile stimuli [6].

In the auditory domain, an illusion similar to the cutaneous rabbit was reported in the seventies and called the auditory saltation [7]. A sequence of clicks was emitted by means of three loudspeakers only. However, for a certain range of inter-stimulus intervals, the subjects consistently reported sound events occurring between the actual emission points, with a phenomenal experience described as "a stick being run along a picket fence". This particular even distribution of apparent locations was reported for very short inter-stimulus intervals (less than 50 ms), but a spread of apparent locations was reported up to about 200 ms of inter-stimulus interval. Experiments that measured the strength of auditory saltation with presentation of clicks via headphones were performed twenty years later [8]. For monoaural stimuli the effect never occur, and localization is discrete. For localization to be continuous in space, the stimuli must be dichotic with Interaural Time Difference (ITD) in a certain range (less than 1 ms) and inter-stimulus interval shorter than 100 ms. This kind of stimuli are perceived similarly as a a variable ITD click train, representing a source that is actually hopping between the two ex-

treme positions. Further experiments with dichotic clicks measured the strength of the saltation effect under different degrees of lateralization [9]. A temporal window for stationarity was also measured to be about 350 ms long, thus meaning that if the initial stationary clicks (before the actual change in ITD occurs) stay within this window, the whole progression through space is reported to begin immediately. A procedure for the psychophysical assessment of individual auditory saltation, useful for the diagnosis of dyslexia, was also proposed [10]. With the method of constant stimuli, subjects had to discriminate between "actual" motion and saltation, with sequences played via headphones. The mean saltation threshold was found to be around 100 ms. In another experiment, subjects had to adjust eight sliders to report on the apparent position of individual clicks. It was found that some individual responses can be non monotonic. The reduced rabbit paradigm (three clicks) was used to check the effect of spectral content on saltation [11]. When the second click has a different content from the third click the effect is much weakened. It is argued that a form of perceptual masking occurs, where the localization of the target is impaired by the subsequent click. Displacements associated with saltation are stronger when the temporal, spatial, and spectral proximity of the stimuli is higher.

3. EXPERIMENTS ON SPATIO-TEMPORAL SONIC GESTURES

When sound is associated with movement we can talk about sonic gestures, with or without human agency [12]. Schemata of action-sound types can be summarized in: (i) *iterative*, when quick successions of small movements, and therefore corresponding sounds, are fused in a single gesture, or sound event, such as a drum roll; (ii) *impulsive*, namely gestures that imply discontinuous effort and are aimed at discrete events, such as hitting or knocking; (iii) *sustained*, when the action type requires a prolonged and continuous effort, such as bowing a string [13]. In music, a gesture is a coherent unit that develops in time, a trajectory in a space where a musical parameter (typically pitch) unfolds. Music theory, and especially music rhetoric, is in a large extent about how to design, concatenate, and overlap gestures [14]. Only occasionally the musical gestures inhabit a physical space, when there is an explicit displacement of a sound source, or when a gesture spans a spatial arrangement of sound sources (as in an orchestra). As long as energetic coherence unfolding in time and space is preserved, the action-sound types may be applied to new sounds, and be physically distributed in space to afford some gestural configurations.

3.1 Experiment 1: Is localization accuracy gesture-dependent?

The first experiment is aimed at studying if accuracy on spatial localization of a sound may be affected by sound motion, namely if arranging a point-like sound in a sequence of pulses distributed in space and time leads to a localization improvement.

In our experimental environment four piezo speakers are taped on a line along the middle line of a cardboard panel ($1000 \times 700\mathrm{mm}$), and arranged at equal distance from each other. The panel is hanging on a wall, with the speakers hidden from view, and the long side parallel to the floor. A projector beams a horizontal strip on the middle line of the panel. The strip is the clicking area for the user who will be using a mouse to manually input the estimated location of the sound event.

Rapid sequences of one to four impact sounds are played in various positions and direction, each impact assigned to a speaker. Various basic gestures are performed, in the classes: (i) point-like, (ii) linear monotonic sequence, (iii) linear sequence with one inversion of direction (at second or third impact). Subjects are asked to point to the position of the last impact in the sequence. Dispersion of answers gives an indication of accuracy in localization. The hypothesis is that accuracy increases with apparent sound motion, and with expectation on the final point in the sequence. We expect that accuracy on static localization integrates with other information coming from motion (essentially temporal information and expectation), thus increasing the final accuracy. The psychoacoustic literature has previously faced the problem of measuring accuracy in localization tasks [15].

3.1.1 Setup

Let δ be the distance between two contiguous piezo speakers, and d be the distance between the panel and the listener. If the listener has the head facing the panel and symmetrically located in the middle of the two piezo speakers, the angle between the listener and such two speakers is

$$\alpha = 2 \arctan \frac{\delta}{2d}. \tag{1}$$

If $d = 1700\mathrm{mm}$ and $\delta = 300\mathrm{mm}$, we get $\alpha = 10.085°$. This is well beyond the minimum audible angle, which amounts to a couple of degrees for frontal sources [1]. Let T be the inter-onset interval between the sounds being emitted by two adjacent speakers. To be sure that no offset displacement due to auditory saltation occurs, T should be larger than about $300\mathrm{ms}$ [9]. Although we are interested in the final position, a displacement in the intermediate positions may affect the regularity of the pattern and, therefore, the expected final position.

If a continuous sound source would move continuously between two adjacent points, the velocity limen would be $9.1°/\mathrm{s}$ for a source moving at $30°/\mathrm{s}$ [16]. Given the above constraint on the time to jump from a location to the adjacent one, we would have velocities lower than this, thus ensuring that α is larger than the minimum audible movement angle.

The subject seats in front of the cardboard panel, approximately at a distance of $1.7\mathrm{m}$. The test is run in ecological conditions, that is in an ordinary, everyday life environment, and in our case in a small room of $3 \times 5\mathrm{m}$ with a wooden floor and a glass door in a glass wall, equipped with regular office furniture (a large bookshelf, a desk and chairs). The background noise includes the fan of the video

projector, the hum of the heating system, various noises coming from the corridor, church bells in the distance from time to time. The measured average background noise under experimental conditions was $40.7\mathrm{dbA}$ (rms value with frequency weighting A and fast exponential averaging of $125\mathrm{ms}$).

3.1.2 Stimuli

The sound stimuli are synthetic impact sounds, generated with the Sound Design Toolkit [17], and designed to convey the impression of short impacts on a wooden surface. Figure 1 shows the four pulses, each emitted by a piezo speaker, as captured at the position of the listener's head, with a Zoom H2 Handy Recorder with X/Y internal microphones configuration at $120°$. To have a clearer view, the waveforms of the impact sound have been juxtaposed with an inter-onset interval much larger than the $300\mathrm{ms}$ used in the experiment. The two channels have been displaced $200\mathrm{ms}$ for a better comparison. The synthetic sound for the experiment has been prepared by interactive listening through the actual setup, with the whole transmission chain, from real-time synthesis to pressure waves at the ear. It has to be noted that differences in the piezo speakers and in their mechanical coupling with the board give rise to different waveforms at the listening position. As seen from figure 1, the stimuli peaked about $20\mathrm{dB}$ higher than background rms noise level.

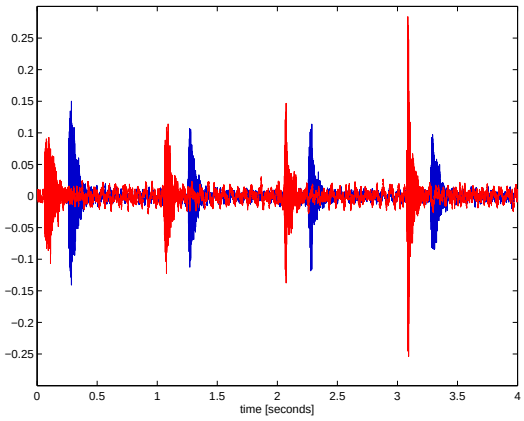

Figure 1. Waveforms of the synthetic, wooden impact sound stimulus as coming from the four emission points and reaching the listener's head. Left (blue colour) and right (red colour) channels have been chopped and displaced for better visibility.

3.1.3 Procedure

The trial is automated and formed of 5 cycles of 26 randomly played sound stimuli, for a total amount of 130 stimuli. Each stimulus is either a single impact sound or a sequence of two to four impacts emitted from different piezo loudspeakers. The list of stimuli is described in table 1. The subject is asked to locate the last heard sound along a red line projected on the panel. At the end of each stimulus the colour of the line switches to green, and the subject can indicate the final landing point, by pointing and clicking with a mouse. Between the subject selection and the following stimulus there is a pause of $3\mathrm{s}$. Collected responses are saved in a text file. For each stimulus the text file includes: the corresponding sequence, the inter onset interval, the final location along the line in a range between 0 and 1, the subject's response time in ms.

n.	sequence of actuated speakers			
1	1			
2	2			
3	3			
4	4			
5	1	2		
6	2	3		
7	3	4		
8	1	2	3	
9	2	3	4	
10	1	2	3	4
11	4	3		
12	3	2		
13	2	1		
14	4	3	2	
15	3	2	1	
16	4	3	2	1
17	1	2	1	
18	2	3	2	
19	3	4	3	
20	4	3	4	
21	3	2	3	
22	2	1	2	
23	1	2	3	2
24	2	3	4	3
25	4	3	2	3
26	3	2	1	2

Table 1. The 26 stimuli used in the experiment

The subjects are briefly informed on the objective of the experiment, namely measuring the accuracy in the spatial localization of moving sound sources, and asked to give their informed consent. Afterwards the task is explained through a metaphor: the subject will hear a pet pattering behind the red strip. When the pet stops, the color of the strip turns green and the position of the pet on the strip has to be located by pointing and clicking with the mouse. After the experiment, each subject is debriefed and his or her comments recorded for further analysis.

3.1.4 Results

Eleven subjects, seven males and four females, ranged in age between 27 and 43, performed the experiment. Only subject n. 5 reported a partial hearing loss at one ear. A posteriori we verified that he could easily discriminate between left and right stimuli, and we decided to keep him in the pool of subjects. The boxplots of figure 2 show how the individual responses to the five cycles of the single stimulus are distributed. It is clear that some subjects (5, 7) preferred to collapse their responses toward the extremes, while the others used most of the availabe space.

To test the hypothesis that the sequencing of impact sounds affects the localization accuracy for the last impact, the results from all 11 subjects and all 26 stimuli have been aggregated in the boxplots of figure 3. The nine boxplots correspond to the categories of impact sound sequences listed in table 2.

The visual inspection of the boxplots induces some observations: (i) emission points are quite well localized; (ii)

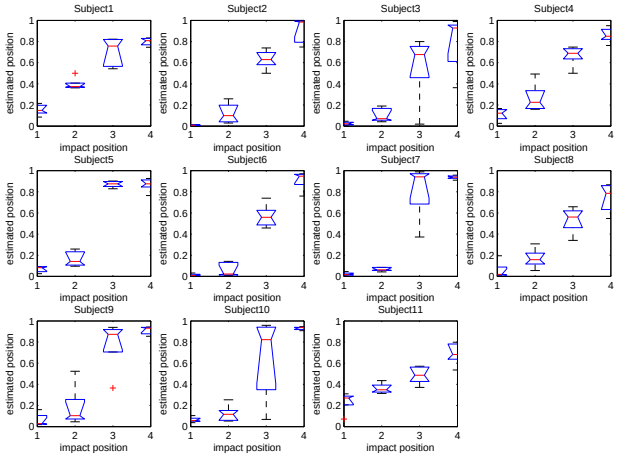

Figure 2. Boxplots of individual responses sorted by emission point (1 to 4)

sequence description	sequence n. (see table 1)
single impact	1, 2, 3, 4
double impact - left to right	5, 6, 7
double impact - right to left	11, 12, 13
triple impact - LR (3,4) and RL (1,2)	8, 9, 14, 15
quadriple impact - LR (4) and RL (1)	10, 16
triple impact back and forth - LRL	17, 18, 19
triple impact back and forth - RLR	20, 21, 22
quadriple impact back and forth - LLR	23, 24
quadriple impact back and forth - RRL	25, 26

Table 2. Categories of impact sequences

localization is more accurate near the board rim; (iii) accuracy is not affected by the sequence. Overall, the subjects localize the final impact in each sequence around the actual emission point with a standard deviation that is always less then 20% of the whole strip length, which is significantly larger than the minimum audible angle.

From visual inspection of the single-impact boxplot of figure 3 it seems that standard deviation is smaller at the rim. That there is a significant variation in variance of localization among the different emission points is confirmed by a Levene's test on equality of variances ($F(3, 216) = 8.4994$, $p = 2.3E\text{-}5$). Standard deviation at positions 1 and 4 is around 10% of the whole strip length. That localization is more accurate at the rim is confirmed by the results of localization of arrival points for the other sequences. In order to eliminate the variability among repeated measures of the same subject in the same condition, we took the median of each set of five cycles and checked if such median estimate would change its variance with the different categories described in table 2. A Levene's test did not allow to confute the null hypothesis of equality of variances. For example, for final emitting point at position 1 (five possible sequences in table 1): $F(4, 50) = 0.484$, and $p = 0.747$. Similarly, for final emitting point at position 2 (seven possible sequences in table 1): $F(6, 70) = 0.709$, and $p = 0.6444$.

3.1.5 Discussion

It seems that the hypothesis of better accuracy for sequences of three or more impacts, as induced by some expectation

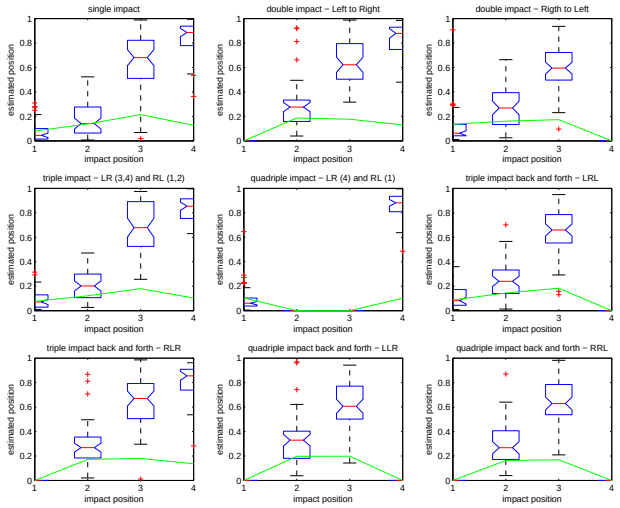

Figure 3. Boxplots of all collected responses sorted by emission point (1 to 4). Each boxplot represents a subset of stimulation sequences (see table 2). The green line connects the values of standard deviation, for each emission point.

on the space-time localization of the final impact, is not confirmed by the experiment. In fact, auditory saltation effects typically occur for intermediate emissions in a sequence. Stretching and compression of time and space do not occur for the extreme stimulation points, also in the cutaneous rabbit illusion. However, the experiment shows that the points close to the boundary are located quite accurately.

In the comments, all the subjects pointed out a major perceived immediacy in locating point-like, stationary sounds. The majority exploited the sounds occurring on the rims of the panel as anchors and reference points to construct the reach of the designed auditory space. In addition, several subjects showed a preference to an eyes-free interaction approach in executing the task, due to a relative uselessness of sight (subject 1), a clearer identification of the landing point at the extreme left, right or the center of the strip (subject 6), and a certain misleading and interfering effect of the cursor on screen with the localization task (subject 9).

Indeed, the primary aim of audition with respect to space is to orient the gaze toward the source [18, 19]. In this respect, using a sequence that traverses the available surface space, as in stimuli 10 and 16 of table 1 gives some advantages, as compared to a in-place emission. First, the early impacts of the sequence work as *attensons*, to drive the users' attention toward the emission point, since "they emphasize the sound motion" (subject 1). In our experimental conditions the subjects were attending the audio-visual display, but in ecological and practical settings, the attention of potential users will need to be driven toward the interaction point. Second, sequencing acoustic emissions in time and space opens a wide design space: While point-like stimuli were reported simply as "sound", sound

motion in sequences was described in terms of pairs of opposites near/far, back and forth, and through metaphors, therefore stressing a major, inner expressiveness. For instance, subject 6 visualized the perceived jumps of the virtual pet, while subject 9 figured the task as a sort of Duck Hunt game [2]. Ultimately, subject 7 imagined something similar to the conjuring trick of the cups and the balls.

Several subjects reported to be misled by pitch during the experiment. Although the stimulus is a synthetic sound, perceived timbral differences at the various emission points are due to the non-linear characteristics of the four piezo-speakers, and to the different excitation of the normal modes of the board. Several subjects reported the metaphor of a piano keyboard as early strategy in executing the task: "Following the sounds as if they were moving on a musical scale", "with lower pitches on my left and higher ones on my right" (subjects 2, 3). In fact, pitch height (frequency) is known to have an associative spatial stereotype effect, with the apparent movement of a sound source in the orthogonal plane. Higher-frequency pitches tend to be associated to right/up locations, while lower-frequency pitches to left/down locations [20].

In summary, spatio-temporally distributed pulses do not affect significantly the accuracy in the localization of the final landing point. Nonetheless, the experiment shows some implications for the design of auditory interfaces: Even simple arrangements of point-like sounds in basic, linear, monotonic sequences allow to construct expressive gestures and give rise to meaningful, interpretive processes.

3.2 Experiment 2: Auditory saltation and perceived gesture

In the second experiment we explored the gestural dimension of spatio-temporally distributed pulses. In particular, we investigated how short sequences of point-like sounds that originate on the rim and traverse the available surface space, are perceived and represented in terms of trajectories and gestures. Indeed, short sequences of pulses displaced in space and rapidly repeated in time (inter-onset-interval – IOI below 100ms) are perceived as continual. The hypothesis is that actual perception of the time-space distribution of events may be affected by illusory saltation (see section 2).

It is interesting to see how subjects represent the perceived gesture, and how this relates to the physical spatio-temporal distribution of pulses. For this purpose, participants were asked to reproduce, by tapping or tracing on a graphic tablet, the sequence of pulses. The precise timing, position and displacement of the pen tip on the tablet was acquired.

3.2.1 Setup and stimuli

This second experiment was run with the same setup and impact sounds, as in the first experiment, except that no visuals are projected on the panel. As an input device, a Wacom Intuous 2 USB tablet is used. The trial consists of 3 groups of 8 sound stimuli, for a total of 24 randomly played stimuli. Each stimulus is a sequence of twelve impact sounds evenly distributed in time and spatially arranged as (i) a traversing sequence from one side to the other, (ii) or a traversing sequence with one inversion of direction (at the anchor point), (iii) or a sequence presenting two blocks of six impacts at the opposite rims, (iv) or a sequence of two blocks of six impacts at adjacent positions on the left or right half of the panel. The first group of sequences is played with an IOI of 300ms, the second group with an IOI of 150ms, and the third group with an IOI of 75ms. The complete list of stimuli is given in table 3.

[2] `http://en.wikipedia.org/wiki/Duck_Hunt`.

n.	sequence of actuated speakers						IOI (ms)
1	111111			444444			300 ms
2	444444			111111			”
3	111	222	333	444			”
4	444	333	222	111			”
5	111111	222222					”
6	444444	333333					”
7	11	22	33	44	33	22	”
8	44	33	22	11	22	33	”
9	111111			444444			150 ms
10	444444			111111			”
11	111	222	333	444			”
12	444	333	222	111			”
13	111111	222222					”
14	444444	333333					”
15	11	22	33	44	33	22	”
16	44	33	22	11	22	33	”
17	111111			444444			75 ms
18	444444			111111			”
19	111	222	333	444			”
20	444	333	222	111			”
21	111111	222222					”
22	444444	333333					”
23	11	22	33	44	33	22	”
24	44	33	22	11	22	33	”

Table 3. The 24 stimuli used in the experiment. Numbers 1 to 4 represent the active speaker, each number repeated according to the number of pulses per actuated speaker. Reading a line left to right gives the spatial and temporal unfolding of the sequence.

The Wacom tablet is used as scaled analogue of the cardboard panel, where the subjects can represent the stimuli. The collected data for the subject's response to each stimulus includes an indexical flag of the randomly played sequence, the corresponding IOI, the points marked with the relative temporal distance in ms between each pair, the XY coordinates of the pen strokes per instant of time.

The subject is instructed about the objective of the experiment, namely observing if and how sequences of short pulses, spatio-temporally distributed along the horizontal axis located in the middle part of a surface (the cardboard panel) may be perceived as gestural strokes. Hence, the subject is asked to reproduce on the tablet the last heard sequence, with freedom to use both pointillistic and/or continuous strokes, according to his/her ease and confidence. The subjects are explicitly acquainted of the unidimensional nature of the experiment that takes in account only the perceived movement of the sound sequences in the horizontal plane, and does not consider the perceived displacement in the vertical plane. A short training session is dedicated to listening to a couple of sequences per group of stimuli, in order to raise any misunderstanding about the task. Two sequences with long IOI, one traversing the panel and one presenting the blocks of impacts at the opposite rims, are always played as initial training elements,

in order to highlight the difference between the stimuli tha are coming from left and right extremities of the panel anc the stimuli that traverse the board. The long IOI acts a control condition, being the auditory saltation effect tipy cally occurring for shorter IOI. The sequences are man ually triggered by the experimenter, after the experimen each subject is debriefed and his or her comments recorde for further analysis.

3.2.2 *Precision of the input device*

Graphic tablets have been extensively used as input device in sound and music computing [21]. Some measurement of total latency have been performed [22] by capturing th contact sound of the pen on the tablet with a microphon and by measuring the time lag between the detected soun impulse and the contact information received via the tablet We did a similar measurement using the Alesis soundcard iO26 firewire, a MacBook Pro 2.33 GHz Intel Core Duo running Max/MSP 5 on Mac OS X 10.6.6. The external `wacom` by Jean-Michel Couturier [3] was used to capture the tablet data. The two stimuli (audio and tablet data) were displaced in time of a few (positive or negative) milliseconds. On a sequence of 128 stimuli, the measured mean temporal displacement was -1.44ms, with a standard deviation of 13.77ms. The time interval between detected XY events for continuous strokes is on average between 8ms and 12ms.

3.2.3 *Results*

Twelve volunteer subjects, four males and eight females in an age between 29 and 70, participated to the experiment.

For each subject and each stimulus, the X coordinates of the Wacom events were plotted versus time and the plots for all stimuli were arranged in a table, as reported in figure 4 for one subject. For ease of comparison, time was normalized to overall gesture duration (represented by number 1000). In addition, we measured the density of pen events sent by the tablet in dots/ms. It is clear from the zero overlap between boxes in figure 5, that the density for 300ms of IOI (1 in the figure) is significantly smaller than the density for 75ms of IOI (3 in the figure). The overall duration of gestures has also been measured and is represented in the boxplot of figure 6. It shows that gestures corresponding to 300ms of IOI take significantly longer than gestures corresponding to 75ms of IOI.

3.2.4 *Discussion*

On the single subject of figure 4, several observations may be made: (i) plots in the leftmost column are more step-like than plots in the rightmost column; (ii) plots in the leftmost column are more dot-like, while plots in the rightmost column are more continuous; (iii) local inconsistencies are found, as for example in the central plot of the bottom row. The first observation seems to support the emergence of a saltation effect. To see how this generalizes across subjects, the gestures of all subjects were made continuous by resampling with a zero-order hold, and the mean gestures plus/minus standard variance were plotted (see figure 7).

[3] `http://www.jmc.blueyeti.fr/download.html`

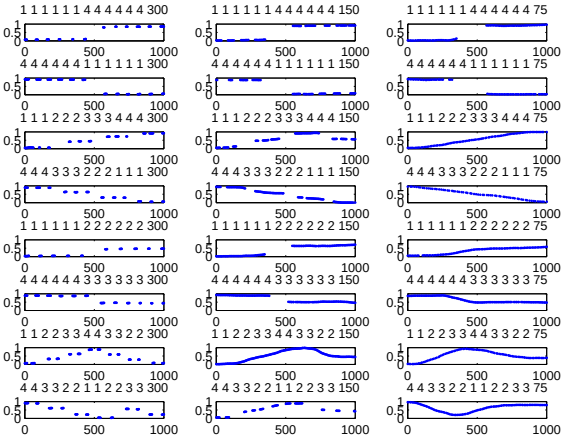

Figure 4. Reproduced gestures for one subject. The three columns, left to right, correspond to IOI of 300ms, 150ms, and 75ms, respectively. In each plot, vertical axis is the normalized horizontal position in the tablet, and horizontal axis is time, normalized to the whole gesture duration.

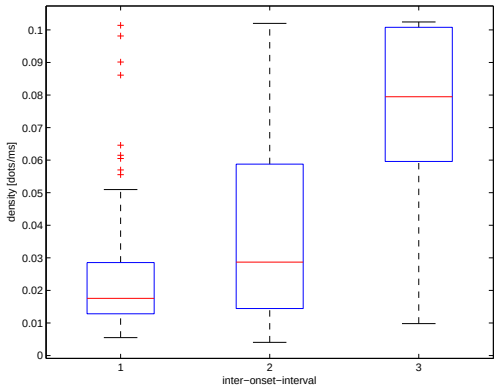

Figure 5. Boxplots of density of pen events for all subjects and all sequences, for the three different IOIs.

Here, the difference between the leftmost and rightmost plots is much harder to appreciate. On the contrary, it is clear that the sequences that traverse the board by using all the four speakers (second and third row) are smoother than the sequences that use only the extremal speakers. This shows that the internal speakers are not useless in defining the movement and that illusory saltation between the extremal position is not very relevant. We should be careful before saying that there is no or little saltation when going from 300ms to 75ms, just because the leftmost and rightmost plots in rows 3 and 4 of figure 7 look similar. Indeed, the steps that are clearly visible in the corresponding leftmost plots of figure 4 may disappear from the mean trajectory just because different subjects locate them at different positions in time.

The second observation made for the single subject is supported by the fact that density of pen events for 75ms of IOI (3 in figure 5) is significantly larger than density

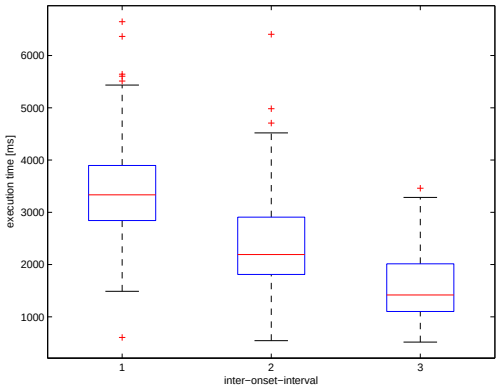

Figure 6. Boxplots of duration of reproduced gestures for all subjects and all sequences, for the three different IOIs.

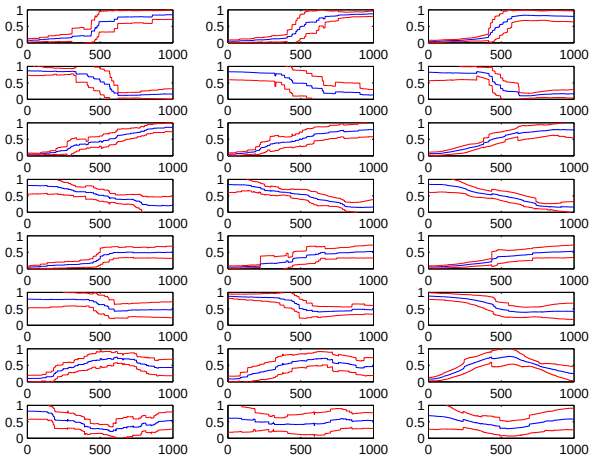

Figure 7. Mean and standard deviation of the collected responses of all the subjects. Duration of responses has been normalized to 1000ms. The three columns are sorted by IOI (300ms, 150ms, and 75ms), from left to right.

for 300ms of IOI (1 in the figure), thus meaning that gestures were definitely denser, or more continuous, in the latter case. The perceived major smoothness and continuity of sequences with shorter IOI (75ms) is corroborated by participants' comments. This fact is only partially emerging from figure 7. Finally, with respect to gesture duration (figure 6), it should be noted that gestures corresponding to stimuli with 75ms of IOI take one quarter of the time taken by the stimuli using a 300ms. Subjects were left free to use their own time scale in the reproducing gesture and, indeed, they did not scale with the duration of stimuli, as the ratio between median execution times in the first and third columns of figure 6 is about 2.35. Conversely, the medians of densities reported in figure 5 scale almost perfectly with the density of stimuli.

Several subjects reported to perceive the sequences with shorter IOI and presenting pulses at the extremities (n. 17 and 18 in table 3), as two *blocks* of events very close to each other and almost tied. Nonetheless only three subjects represented them with a tied and continuous stroke, which highlights the limited incidence of auditory saltation effect in ecological conditions. Subjects 1 and 9 stressed a sort of wake effect, subject 7 recalled the sound of rolling on a snare, while subject 9 reported the metaphor of two separate blocks of *domino* pieces falling. In traversing sequences, the *domino* effect sensation was in fact complete, running from one edge to the other, the movement more fluent and pleasurable (subject 7). Traversing sequences, in particular those with shorter IOI, give rise to cognitive representations emanating from the sound characteristics [23], and the spatio-temporal distribution. The depicted gestures represent a sort of signature, or perceived morphology of the sound contours in space and time. An additional consideration concerns how timing of pulses, and total duration of sequences affect the adopted strategy. In the comments, the control condition sequences (IOI 300ms) are described with terms, often inappropriate or naïve, like *rhythm*, *beats*, *jumps*, *hits*, therefore implying blocks of discrete events. This group of sequences acted as reference point for the execution of the task, even when they were randomly presented later in the trial. It is a natural and immediate strategy to try to count the pulses and tap them accordingly, when possible. On the contrary, the comparison against the other two groups of sequences is done in terms of a scale of speed. Pulses with short IOI (75ms) are almost perceived as tied and induce a change of strategy, showing the difficulty of tapping at the same tempo in the attempt of reproducing the sequence. Pulses with intermediate IOI (150ms) seem to require some preparatory gestures in order to reproduce the taps. It was observed that several subjects that preferred to tap these sequences prepared them by tapping in the air in order to keep the tempo. These sequences reveal the subjects' attitude toward a pointillistic or stroke-like approach to the task.

4. CONCLUSION

Sound is increasingly being used as intentional design element in artefacts and environments, as specifier of brands and privileged channel of interaction. Space is an obviously ineluctable dimension of the experience of the world and the moderate accuracy of the human ear in sound localization is a matter of fact. Research in sound design has to look closely at the element of space. For instance, the effectiveness of a well designed sound logo may be reduced if badly presented at the touch points between companies and their customers. For this purpose, we investigated the quantity and quality of apparent sound motion effects, such as auditory saltation, in ecological conditions. In the first experiment we found that arranging a point-like sound in spatio-temporally distributed sequences does not improve noticeably the localization of the final, landing point. Yet, design can exploit the emerging anchor effects of the extremal elements to display interaction reaches, while taking advantage of the initial elements to call attention toward the emission point. The second experiment showed that auditory saltation effect in ecological conditions is reduced compared to headphones listening. Nonetheless we

could observe how it affects the perceived representations of the sound motion, in terms of gestural strokes. This opens a wide design space, since acoustic brand units, in the form of sound logos, jingles, display or product sound, can be developed in space and time, thus introducing a shape aspect that is normally not explicit.

5. REFERENCES

[1] A. W. Mills, "On the minimum audible angle," *The Journal of the Acoustical Society of America*, vol. 30, no. 4, pp. 237–246, 1958.

[2] J. Neuhoff, "Auditory motion and localization," in *Ecological psychoacoustics*, J. Neuhoff, Ed. New York: Academic Press, 2004, pp. 87–111.

[3] D. Goldreich, "A Bayesian Perceptual Model Replicates the Cutaneous Rabbit and Other Tactile Spatiotemporal Illusions," *PLoS ONE*, vol. 2, no. 3, pp. e333+, March 2007.

[4] P. Lemmens, F. Crompvoets, D. Brokken, J. van den Eerenbeemd, and G.-J. de Vries, "A body-conforming tactile jacket to enrich movie viewing," in *Proceedings of the World Haptics 2009 - Third Joint EuroHaptics conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*. Washington, DC, USA: IEEE Computer Society, 2009, pp. 7–12.

[5] W. D. Jones, "Jacket lets you feel the movies," *IEEE Spectrum*, 2009, http://spectrum.ieee.org/biomedical/devices/jacket-lets-you-feel-the-movies.

[6] J. Raisamo, R. Raisamo, and V. Surakka, "Evaluating the effect of temporal parameters for vibrotactile saltatory patterns," in *Proceedings of the 2009 international conference on Multimodal interfaces*, ser. ICMI-MLMI '09. New York, NY, USA: ACM, 2009, pp. 319–326.

[7] C. D. Bremer, J. B. Pittenger, R. Warren, and J. J. Jenkins, "An Illusion of Auditory Saltation Similar to the Cutaneous "Rabbit"," *The American Journal of Psychology*, vol. 90, no. 4, 1977.

[8] D. I. Shore, S. E. Hall, and R. M. Klein, "Auditory saltation: A new measure for an old illusion," *The Journal of the Acoustical Society of America*, vol. 103, no. 6, pp. 3730–3733, 1998.

[9] D. P. Phillips and S. E. Hall, "Spatial and temporal factors in auditory saltation," *The Journal of the Acoustical Society of America*, vol. 110, no. 3, pp. 1539–1547, 2001.

[10] J. C. Kidd and J. H. Hogben, "Quantifying the auditory saltation illusion: An objective psychophysical methodology," *The Journal of the Acoustical Society of America*, vol. 116, no. 2, pp. 1116–1125, 2004.

[11] S. Getzmann, "The Effect of Spectral Difference on Auditory Saltation," *Experimental Psychology*, vol. 55, no. 1, pp. 64–71, 2008.

[12] A. R. Jensenius, M. M. Wanderley, R. I. Godøy, and M. Leman, "Musical gestures: concepts and methods in research," in *Musical Gestures - Sound, Movement, and Meaning*, R. I. Godøy and M. Leman, Eds. New York: Routledge, 2010, pp. 12–35.

[13] R. I. Godøy, "Gestural affordances of musical sound," in *Musical Gestures - Sound, Movement, and Meaning*, R. I. Godøy and M. Leman, Eds. New York: Routledge, 2010, pp. 103–125.

[14] A. Schneider, "Music and Gestures: A Historical Introduction and Survey of Earlier Research," in *Musical Gestures: sound, movement, and meaning*, R. I. Godøy and M. Leman, Eds. New York, NY, USA: Routledge, 2010, pp. 69–100.

[15] S. Carlile, P. Leong, and S. Hyams, "The nature and distribution of errors in sound localization by human listeners," *Hearing Research*, vol. 114, no. 1-2, pp. 179 – 196, 1997.

[16] S. Carlile and V. Best, "Discrimination of sound source velocity in human listeners," *The Journal of the Acoustical Society of America*, vol. 111, no. 2, pp. 1026–1035, 2002.

[17] S. Delle Monache, P. Polotti, and D. Rocchesso, "A toolkit for explorations in sonic interaction design," in *Proceedings of the 5th Audio Mostly Conference: A Conference on Interaction with Sound*, ser. AM '10. New York, NY, USA: ACM, 2010, pp. 1:1–1:7.

[18] D. V. Valkenburg and M. Kubovy, "In defense of the theory of indispensable attributes," *Cognition*, vol. 87, pp. 225–233, 2003.

[19] M. Kubovy and M. Schutz, "Audio-visual objects," *Review of Philosophy and Psychology*, vol. 1, pp. 41–61, 2010.

[20] E. Rusconi, B. Kwan, B. L. Giordano, C. Umilt, and B. Butterworth, "Spatial representation of pitch height: the SMARC effect," *Cognition*, vol. 99, no. 2, pp. 113 – 129, 2006.

[21] M. Zbyszynski, M. Wright, A. Momeni, and D. Cullen, "Ten years of tablet musical interfaces at CNMAT," in *Proceedings of the 7th international conference on New interfaces for Musical Expression*. ACM, 2007, pp. 100–105.

[22] M. Wright, R. J. R. Cassidy, and M. F. Zbyszynski, "Audio and gesture latency measurements on Linux and OSX," in *Proceedings of the International Computer Music Conference*, 2004, pp. 423–429.

[23] B. Caramiaux, P. Susini, T. Bianco, F. Bevilacqua, O. Houix, N. Schnell, and N. Misdariis, "Gestural embodiment of environmental sounds: an experimental study," in *Accepted for publication in Proc. NIME 2011 - New Interface for Musical Expression*, Oslo, Norway, 2011.