Peripersonal space Task – Audio-tactile interaction paradigm, Set up Description

Rationale of the task.

In this paradigm, participants are requested to respond as fast as possible to a tactile stimulus administered on a part of their body, while task-irrelevant dynamic (approaching or receding) sounds (or visual stimuli) are presented (see Canzoneri et al., 2012; Teneggi et al., 2013, Noel et al., 2014). The key manipulation in this task is that, on each trial, the tactile target stimulus is presented at a different temporal delay from sound onset and thus touch is processed with sounds (or visual stimuli) perceived at a different distance from the participant's body. In this way, we sought to determine the critical distance from the body at which sounds (or visual stimuli) boost tactile processing. This point can be considered as the boundary of PPS representation, i.e. a behavioral proxy for the part of space covered by the RFs of multisensory integration systems mapping PPS in humans.

Tactile stimulation is given by means of a vibrotactile device consisting of a small vibrating motor (Precision MicroDrives shaftless vibration motors, model 312–101, 3V, 60mA, 9000 rpm, 150 Hz, 5 g). The motor had a surface area of 113 mm2 and reached maximal rotation speed in 50 ms. The placement and the number of vibrators used varied according to experimental needs.
The auditory stimulus is administered by means of a custom-made auditory set-up described in details below.

Hardware. The audio rendering system was made of two uniform linear arrays (ULA) of eight loudspeakers each (JBL Control 1 Pro WH Pair), a USB audio interface with eight synchronous analog outputs (M-Audio FastTrack Ultra 8R), and two four-channel amplifiers feeding the speakers (Ecler MPA 4 80R).
The two ULA's were placed on the right and left sides of the participants as depicted in Figure A1. The distance d between two consecutive loudspeakers was set to 27.5 cm, and the distance D between the two line arrays was set

to 1 meter. The loudspeakers were fixed on two independent metallic structures that allow horizontal positioning of the ULA's.

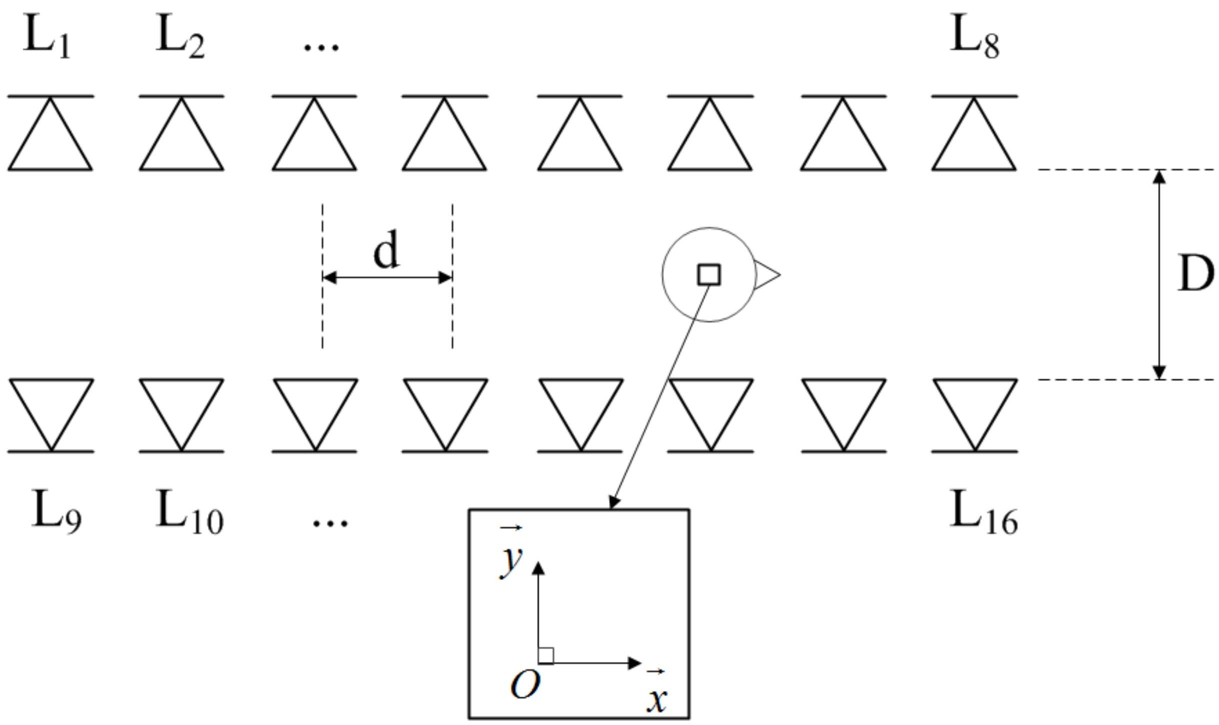

Figure A1: A) Schematization of the rendering system. B) Picture of the rendering system

Stimulus synthesis. The stimulus generated by the loudspeakers spatialized a moving broadband sound source moving through the participant from back to front, or front to back, at a constant speed. This section aims at describing the nature of the stimulus and the methodology for its synthesis.

The global strategy adopted consisted in synchronously sending the same signal to all the speakers, while modulating the amplitude of each channel separately in order to give the subject the impression that the sound source was approaching, receding, or approaching and then receding, from and to a particular direction. The modulation applied on each channel was a function of i) the relative position of the "Virtual Sound Source" (VSS) from the participant, and ii) the relative position of the considered loudspeaker to the Participant Head Center (PHC).

Now, let us consider the PHC as being the center of an orthonormal basis $\mathbf{R}(O,\vec{x},\vec{y})$ as shown in Figure A1 On this basis, let us denote the VSS position as (x^{vss}, y^{vss}), and the position of the q^{th} loudspeaker, $q \in \{1, 2, ..., 16\}$, as (x_q^{ls}, y_q^{ls}). Let us also assume that the VSS is forced to follow the x-axis direction only such that $y^{vss} = 0$ at any time of the experiment. After Figure 1A, it is clear that the loudspeakers can be grouped by pairs where a pair is composed of two loudspeakers with same abscissa. It follows, therefore;

$$\begin{aligned}
&\text{pair } p=1 \text{ (loudspeaker 1 and 9)} \quad : \left\{ x_1^{ls} = x_9^{ls}, y_1^{ls} = -y_9^{ls} \right\} \\
&\text{pair } p=2 \text{ (loudspeaker 2 and 10)}: \left\{ x_2^{ls} = x_{10}^{ls}, y_2^{ls} = -y_{10}^{ls} \right\} \\
&\ldots \\
&\text{pair } p=8 \text{ (loudspeaker 8 and 16)}: \left\{ x_8^{ls} = x_{16}^{ls}, y_8^{ls} = -y_{16}^{ls} \right\}
\end{aligned} \tag{1}$$

Since the objective was to give the impression of a sound source approaching toward or receding from the participant, the signal sent to a particular loudspeaker was the same as the signal sent to its facing loudspeaker. In practice this is achieved by connecting in series the two loudspeakers belonging to the same pair.

Synthesis Algorithm. The first step consisted in creating a pink noise of length N (in samples) and to duplicate it eight times so that each pair of loudspeaker could play its own signal. Let $s_p[n]$ be the signal vector of the p^{th} pair, $(p,n) \in \{1, 2, ..., 8\} \times \{1, 2, ..., N\}$. The integer N is defined by the VSS speed v^{vss} (in m/s). The VSS' start and stop coordinates, x_{start}^{vss} and x_{stop}^{vss} (in meters) and the sampling frequency f_s of the audio sound card (in Hz) are defined according to the following formula:

$$N = \left\lfloor \sqrt{\left(x_{stop}^{so} - x_{start}^{so}\right)^2} \times \frac{f_s}{v^{so}} \right\rfloor \tag{2}$$

where $\lfloor . \rfloor$ stands for the floor function. In practice, we set $f_s = 44.1\,\mathrm{kHz}$, while v^{vss}, x_{start}^{vss} and x_{stop}^{vss} varied according to the experiment (see below for details).

The second step consisted in dividing each signal vector into M frames of length N_f (in samples) with an overlap of N_o (in samples). This so-called framing procedure allowed updating the amplitude not every sample (i.e.

every 0.02 ms), which is a time consuming computation, but every few ms only. In practice, we set $N_f = 2048$ samples and $N_o = 1024$ samples allowing an update of the signal amplitude every 23.22 ms. The number of frames M contained in a signal of length N is given by the formula:

$$M = \left\lfloor \frac{N - N_f}{N_f - N_o} \right\rfloor + 1 \qquad (3)$$

Third, each frame was multiplied by a weighting factor a_{mp}, $m \in \{1, 2, ..., M\}$, whose role was to mimic the amplitude variation of a real moving sound source as a function of time and space. Here we set:

$$a_{mp} = a \times 10^{\frac{60 log_{10}(1/\lambda_{mp})}{20}} \qquad (4)$$

Where a is a scalar allowing the operator to adjust the maximal sound pressure level at the ear of the subject (and play the same role as the gain variator of the amplifier) and λ_{mp} is the distance (in meters) between the VSS and the q^{th} loudspeaker at the instant m. This distance is defined by:

$$\lambda_{mp} = \sqrt{\left(x_p^{ls} - x_m^{vss}\right)^2 + \left(\frac{D}{2}\right)^2} \qquad (5)$$

Note that, according to equation 4, the VSS pressure level decreases 60 dB per distance doubling. This value was chosen in order to mimic a realistic moving sound source. In theory, an omnidirectional sound source level decreases by only 6 dB (Piercy & Daigle, 1998), however, setting this value does not give a realistic perception of moving sound, since we are not used to hear real omnidirectional broad band moving sound sources in nature.

The fourth step consisted in windowing each frame by a suitable window in order to avoid clicks when passing from one frame to the other. We chose the Hann window (Salivahanan, Vallavaraj, & Gnanapriya, 2007) defined by:

$$w[\mathrm{k}] = 0.5\left[1 - \left(\cos\frac{2\pi k}{N_f - 1}\right)\right], \qquad \forall k \in \{1, 2, \ldots, N_f\} \tag{6}$$

Finally, the eight final signals $\tilde{s}_p[n]$, were reconstructed by concatenating all the scaled and windowed frames according to Algorithm 1 described below.

For $p = 1$ to 8 (loop on channels)
  For $m = 1$ to M (loop on frames)
    $j \leftarrow (m-1)(N_f - N_o)$

    For $k = 1$ to N_f (loop on samples)
      $\tilde{s}_p[j+k] \leftarrow \tilde{s}_p[\mathrm{j}+k] + a_{mp} w[\mathrm{k}] s_p[\mathrm{j}+\mathrm{k}]$
    endfor k

  endfor m
endfor p

Algorithm 1: concatenation of the VSS signal

A schematic summary of the steps explained above in order to realistically simulate a moving source of sound through space and in depth is provided in Figure A2.

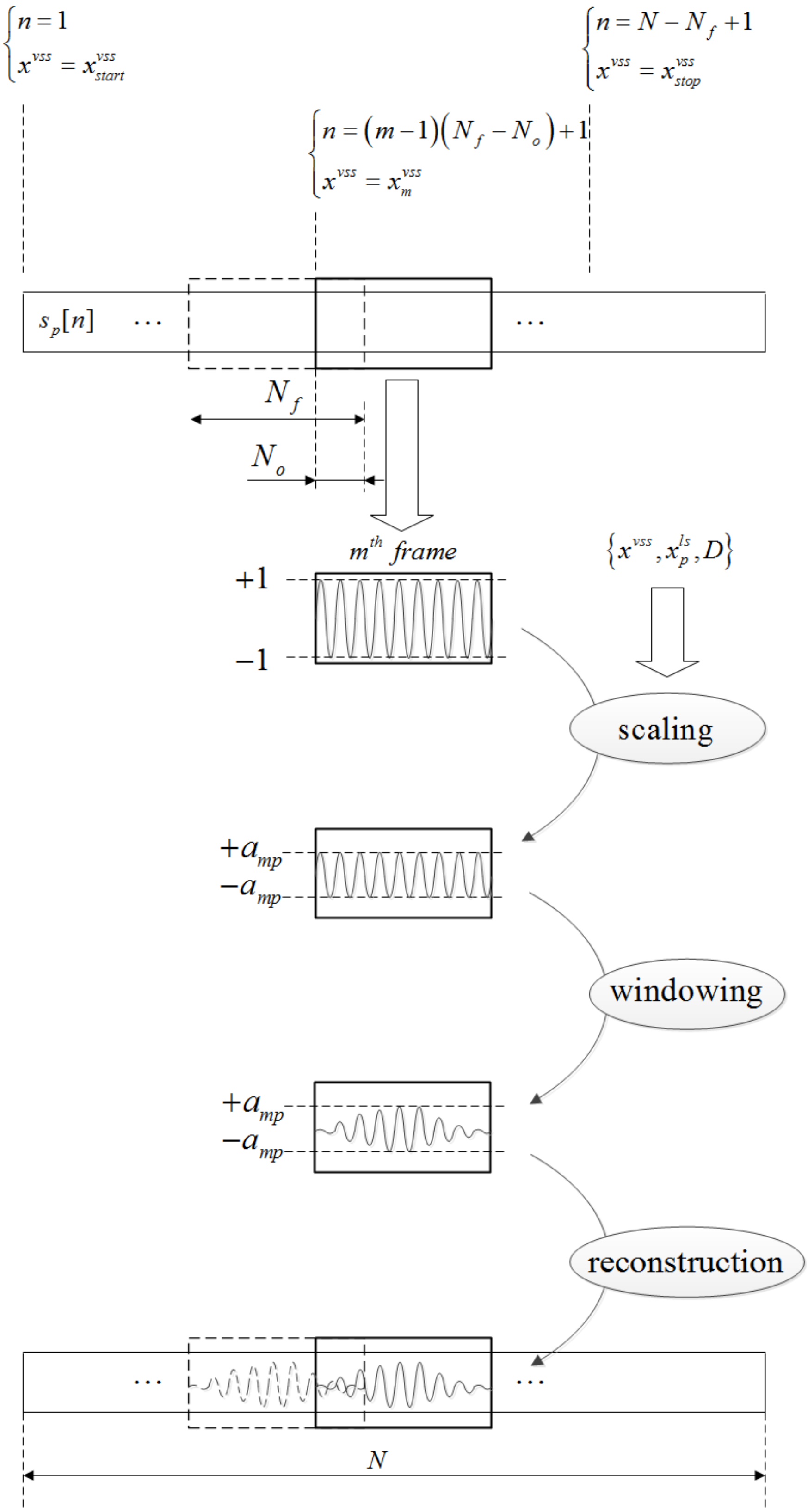

Figure A2: stimuli rendering procedure