Virtual Ventriloquism

Isabelle Viaud-Delmon[1], Ludivine Sarlat[1], Olivier Warusfel[2]

[1] *CNRS UMR 7593, Hôpital de la Salpêtrière, Pavillon Clérambault, 75013 Paris, France, Emails: ivd@ext.jussieu.fr and sarlat@ext.jussieu.fr*

[2] *IRCAM – CNRS UMR 9912, 1 place Igor Stravinsky, 75004 Paris, France, Email: olivier.warusfel@ircam.fr*

Virtual Reality in Psychopathology

Virtual reality (VR) represents a set of computer technologies, which allow users to interact with a 3-D computer-generated environment in real time. VR is now commonly used in psychophysics experiments as well as in psychological therapy. It provides a way to immerse a user in an environment in which the interaction between different sensory modalities can be controlled and is therefore an interesting tool to study the integration of space-related multisensory information in human and its disorders. The specific feature of VR compared to traditional displays is indeed that the environments it provides are places where as many as possible senses are meant to be active. *Multisensory* is a keyword for VR. The number of sensory modalities through which the user is coupled to the virtual environment (VE) is a main factor contributing to the feeling of *presence*. In spite of that, VR technologies rarely integrate the auditory modality, which is the only sense through which we are communicating instantly with the whole space around us.

Patients suffering from anxious disorders, depression or schizophrenia commonly report a hypersensitivity to auditory stimuli, while pure tone audiograms show generally normal hearing. Sound tolerance is influenced by stress and tiredness, and specific sounds can cause physical pain and nerve grating. There is a correlation between loudness tolerance and anxiety [1] and strong emotional reactions can easily be elicited through audition [2]. It is therefore all the more interesting to integrate the auditory modality in a realistic way when working with patients and to understand in which way it can be used for therapeutic purposes.

However, incorporating real-time updated 3-D sound to virtual reality technologies addresses several issues. If there seems to be a consensus on the fact that *presence* is improved by 3-D sound, little is known about how an auditory VE should be designed so that it does not interfere with the visual VE. It is well known that discrepancies in the location of synchronized auditory and visual events can lead to mislocalizations of the auditory source, so-called ventriloquism [3].

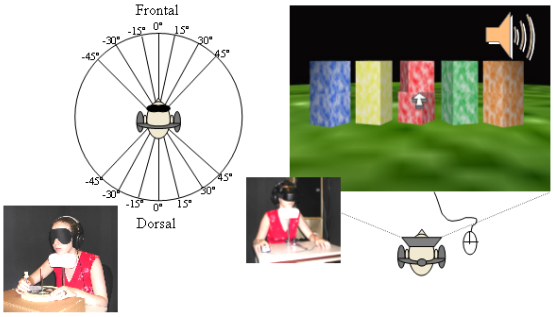
Aim of the study

Spatial attention has to be coordinated across several modalities. This is a non-trivial problem, given that each modality initially codes space in entirely different ways. The brain must continually recalibrate its inputs to optimise the correspondence between the external world and its internal visual representation. To study how different senses contribute to a coherent perception of a virtual space, we presented virtual auditory stimuli via headphones (HRTFs) in temporal synchrony with virtual visual events but with a +15° azimuth bias relative to them. We wanted to observe whether the association of virtual auditory and visual stimuli could lead to a "complete" remapping of auditory space, including stimulus locations not presented during the VR immersion.

Several hypotheses could be envisaged about the effect of immersion in this conflicting VE. Concerning sound sources located in the frontal hemi-space, we expected a rotation of subjective localization towards the left for sound sources located to the right. For sound sources located in the dorsal hemi-space of the subject, three concurrent hypotheses were proposed: (1) subjective localization could remain unchanged after exposition to visuo-auditory conflict; (2) rotation of subjective localization; (3) translation (lateralization) of subjective localization.

Procedure

For each subject (N=14, mean age=20 yrs), the experiment consisted of three main parts: auditory localization task – Phase 1, conflicting auditory-visual stimulation phase – Phase 2, auditory localization task –Phase 3.



Auditory localization task in Phase 1 & 3

Auditory-Visual stimulations in Phase 2
Task: reproduction of movement sequences

Figure 1: The 3 phases of the adaptation paradigm. In Phase 2, each visual object is associated with a sound file spatialized 15° to its right.

Phase 1: Auditory localization task

The subject was seated on a chair, blindfolded, wearing headphones. The head of the subject remained stabilized on a chin-rest. Subjects had to locate 112 stimuli in a pseudo random order, composed of white Gaussian noise (300 ms duration, 5ms raise/fall) pre-convolved with HRTFs measured at directions -45°, -30°, -15°, 0°, 15°, 30°, 45° in azimuth and 0° in elevation; both in the frontal and dorsal hemi-spaces. After each stimulus, the subject had to indicate the auditory event direction with an angular pointer (Figure

1). Although not using individual HRTFs, each subject had been invited during a preliminary session to select the best fitting HRTF set among a database of 50 measured heads [4].

Phase 2: Auditory-Visual Stimulation (AVS)

Subjects remained seated while they were immersed into a virtual environment by means of a stereoscopic head mounted display (Figure 1). We used interactive computer graphics (VR) and an electromagnetic sensor system to track the head in order to simulate the rotation of the virtual environment proportionally to the head angular motion.

Five objects were located in front of the subject (virtual distance from the subject of 2 meters, each object separated from the other by an angle of 7.5°). When activated, the objects rotated on their own axis for 300 ms, while a spatialized sound of the same duration was emitted. A spatial conflict was introduced between the visual stimuli and the auditory stimuli. Each visual object was associated with a sound spatialized 15° to its right. Thus, subjects were submitted to a conflict situation in which, for example, a 0° visual event was concurrent with a 15° to the right auditory input.

The task of the subject in the virtual environment was to reproduce the visual objects sequences of animation. Once the objects were stable, the subject had to reproduce the sequence of movements by clicking on the visual objects with a mouse. This task was applied in order to motivate the subjects to pay more attention to the virtual environment and to prevent the influence of declining alertness. About 1500 bimodal stimuli were released during 20 min. The subject was instructed to rest his head on the chin-rest during the task, and no bimodal stimulation was released during head rotations.

Phase 3: Auditory localization task

The procedure in this task was exactly the same as in Phase 1.

Results and Discussion

The comparison of auditory localization tasks in Phase 1 and 3 (Figure 2) indicated that subjective localization of auditory stimuli was modified after immersion in the conflicting VE. A three-way ANOVA (Phase x Hemi-space x Lateralization) showed a main effect of Phase (F(1,5374)=8.52; $p<0.004$) and an interaction between Phase and Lateralization (F(1,1342)=6,97; $p<0.001$). The auditory subjective localization was mainly modified for right lateralized stimuli in Phase 3, but was observed in both hemi-spaces (frontal and dorsal). The conflict phase exposition did not cause a simple re-association of auditory locations with altered visual locations for the specific locations tested in the training period.

Two control experiments tested whether the observed effect was specific of the bimodal conflict. In the first one, a -10° disparity was introduced between auditory and visual stimuli in Phase 2. This leaded to a subjective shift to the right for auditory sources in the left hemi-space in Phase 3. In the second one, we again used the same paradigm but no spatial conflict between auditory and visual stimuli was introduced in Phase 2. No modification of subjective localization of the auditory stimuli was observed in Phase 3.

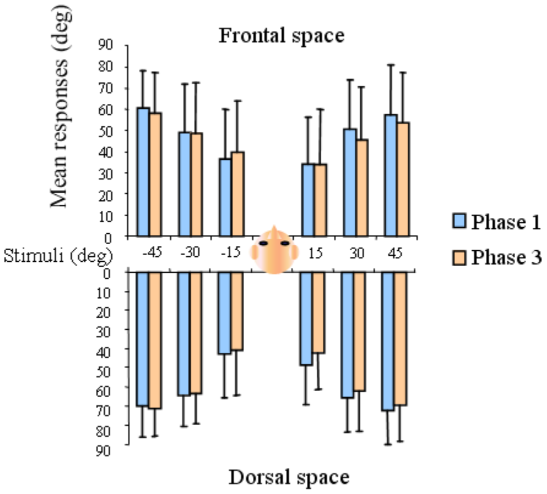


Figure 2: Mean subjective auditory localization in Phase 1 and 3, before and after AVS. Error bars indicate standard deviation.

It is therefore possible to induce a ventriloquist effect with VR, which can not be interpreted in terms of a simple visual biasing of auditory localization. However, our paradigm had several drawbacks: HRTF selection was not accurate enough since inversions of sources localization and overestimations of auditory eccentricities were common. An additional session to adapt to non-individual HRTFs proposed to the subjects prior to the experiment might be useful [5]. The role of the lack of relationship between the egocentric frame of reference of the subject and virtual events (both in VR and Phase 1&3) should as well be addressed. The same experiment presented in augmented reality, with real visual events but virtual auditory stimuli would avoid the partition between the egocentric and virtual reference frames in Phase 2, and allow better Perception / Action coupling, yielding to larger and more homogeneous adaptive changes.

References

[1] Stephens SD. Personality and the slope of loudness function. Q J Exp Psychol **22** (1970), 9-13

[2] Bremner JD, Staib LH, Kaloupek D, Southwick SM, Soufer R, Charney DS. Neural correlates of exposure to traumatic pictures and sound in Vietnam combat veterans with and without posttraumatic stress disorder. Biol Psychiatry **45** (1999), 806-16

[3] Lewald J. Rapid adaptation to auditory-visual spatial disparity. Learn Mem **9** (2002), 268-278

[4] Listen HRTF database. URL: `http://www.ircam.fr/equipes/salles/listen/`

[5] Blum A, Katz B, Warusfel O. Eliciting adaptation to non-individual HRTF spectral cues with multi-modal training. Proc. CFA/DAGA '04