EEG measurement of binaural sound immersion

Rozenn Nicol, Olivier Dufor, Laetitia Gros, Pascal Rueff, Nicolas Farrugia

EEG MEASUREMENT OF BINAURAL SOUND IMMERSION

Rozenn Nicol[1] **Olivier Dufor**[2] **Laetitia Gros**[1]
Pascal Rueff[3] **Nicolas Farrugia**[2]
[1] Orange Labs, 2 avenue Pierre Marzin, 22300 Lannion, France
[2] CNRS Lab-STICC, Technopole Brest-Iroise, 29238 Brest, France
[3] Agence du Verbe, 2 Keraudren, 22260 Plouec Du Trieux, France
rozenn.nicol@orange.com

ABSTRACT

The purpose of this study is to examine whether neuroimaging could provide new insights into the assessment of the listening experience of spatial audio content. Our objective is to explore cognitive processes involved when listening to spatial audio content by electroencephalography (EEG) measurement. The experiment is based on a reversed implementation of the oddball paradigm. Audio stimuli are used to interfere with a detection task of visual stimuli. It is expected that the distracting effect of binaural reproduction is stronger than that of stereophony. Evidence of this is sought in both behavioral data (i.e. the performances in detecting deviant stimuli) and EEG-HR measurements (e.g. P300 response). The differential response between standard and deviant stimuli was examined as a function of the audio distractor. It was observed that the P300 amplitude was significantly higher in presence of the binaural distractor than in the stereophonic condition, suggesting a stronger effect of surprise possibly caused by a more immersive reproduction. This was confirmed by behavioral results which showed a longer response time for binaural distractors (566 ms) than for stereophonic ones (550 ms).

1. INTRODUCTION

The purpose of this study is to examine whether neuroimaging could provide new insights into the assessment of the listening experience of spatial audio content. Conventional methods of assessing sound reproduction are based on either Quality of Experience (QoE) scores or localization judgments [1]. Their main disadvantage is that the subject is aware of his (her) rating task. Besides, the assessment is restricted to specific dimensions (i.e. perception of degradation or spatial attributes). Dimensions such as emotions or cognitive load are generally not taken into account. Therefore, our objective is to explore cognitive processes involved when listening to spatial audio content by EEG (Electroencephalography) measurement. Descriptors of the listening experience (e.g. immersion, realism) are sought in the electrical activity of the brain.

Neurophysiological methods are already identified as a promising way of investigating the field of QoE assessment [2, 3]. A study showed that some EEG features are correlated with emotion primitives (i.e. valence and arousal) [4]. They were successfully used to predict the influence of human factors (i.e. users' perception, emotional and mental state) on a QoE score for a comparison of text-to-speech and natural speech. In a similar study, preference judgments were related to functional near-infrared spectroscopy (fNRIS) features [5]. However, so far there have only been a few experiments that implemented these methods to assess spatial-audio technologies. A first one compared spatial-response fields of the primary auditory cortex with virtual sound sources synthesized with individual and non-individual HRTFs in ferrets [6]. It was shown that the responses obtained with an animal's own ears differed significantly in shape and position from those obtained with another ferret morphology. More recent work has confirmed a positive correlation between various levels of accuracy of spatial sound reproduction (e.g., individual HRTF vs. generic HRTF vs impoverished localization cues [7,8], natural vs. artificial cues of auditory motion [9], individual binaural vs. stereo recordings for sound externalization [10]) and the activity of the auditory cortex measured either by magnetoencephalography (MEG), by EEG or by fMRI. Most of the results reported above were obtained by using binaural synthesis.

In the present study, the quality of spatial sound reproduction will be assessed in terms of its capacity to distract the subject from a task of visual detection. Binaural and stereophonic reproduction will be compared. It is expected that the distracting effect of binaural reproduction is stronger than that of stereophony. Evidence of this will be sought in both behavioral data (i.e. the performances in detecting deviant stimuli) and EEG-HR measurements (256 sensors). First, the experiment is described. Then, all the preprocessing of EEG data is detailed, before presenting results.

doi:10.25836/sasp.2019.21

Figure 1. EGI sensor net composed of 256 electrodes.

2. METHOD

2.1 Participants

Nineteen healthy adult participants reporting normal hearing and vision participated in the experiment (9 males and 10 females, mean age = 34.3 years; min/max = 18/59 years). All participants were correctly informed of the experiment and signed a consent form before being included following the international Helsinki recommendations on human research.

2.2 Experimental setup

Subjects were seated in a comfortable chair in an electrically shielded room at the Pontchaillou Hospital in Rennes. The chair faced a computer screen (1 m distant) and was surrounded with opaque curtains, so that the participants could not see what was happening in the room.

A EGI sensor net of 256 electrodes (Fig. 1) was placed on the participant's head to record his (her) brain activity. In addition, a headphone (Sennheiser HD650) was arranged over it to reproduce sound scenes that were previously recorded in the same experimental room. The pretext for justifying the use of headphones was a listening session as a further step in the experiment. The participant was thus not aware that virtual sounds were going to be played by headphones, and that his (her) listening experience was under study. Besides, at no time was he (she) asked direct questions about the quality of sound reproduction. Interviewed after the experiment, most participants reported that they believed that the sounds really emanated from the experimenter, not from the headphone reproduction.

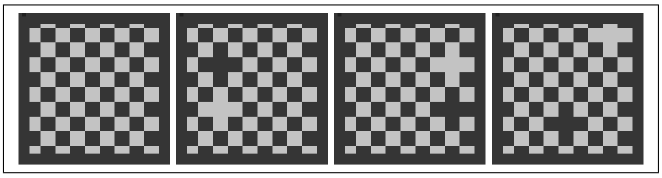

Figure 2. Visual stimuli used in the oddball paradigm: standard stimulus (first on the right) and 3 examples of deviant stimuli.

2.3 Oddball paradigm

The experiment consisted in a visual oddball paradigm programmed with E-prime software. Both accuracy and reaction time were recorded. The participant was presented a sequence of images representing small (1 cm sided) grey checkerboards randomly distributed (n=320). Standard stimuli (n=265) represented perfect checkerboards, while deviant stimuli (n=55) were representing altered checkerboards (see Fig. 2). The rare stimuli (i.e. altered checkerboards), which represented 17% of the whole set of stimuli, were introduced to generate an event-related potential (ERP), known as the P300 in the brain. Four sequences of pseudo-randomly distributed stimuli were created. The participant was asked to click whenever he (she) detects a deviant stimulus. Audio stimuli were used to interfere with the visual oddball task. The objective was to observe the effect of such distractors on the P300 response as a function of the type of sound spatialization. Every participant had to perform the task twice; with stereo and binaural sounds separately (see section 2.4). Sounds and visual stimuli were not time locked, but both stereo and binaural conditions were associated with perfectly counterbalanced sequences of visual stimuli across subjects.

A trial was composed of the following sequence: first, a grey screen (i.e. no visual stimulus) for a variable period of time chosen to be either 150 or 650 ms; second, the visual stimulus for a period of 250 ms; third, a grey screen for a period of 750 ms. Thus, the participant had 1 s (250 + 750 ms) from the moment the stimulus was displayed, to click if he (she) detected a deviant stimulus. At the beginning of the experiment, the participant carried out a training phase for the visual detection task. At the end, he (she) had to complete a questionnaire about the task difficulty. It was also checked whether he (she) realized that sounds were coming from headphones. Triggers corresponding to the onsets of visual stimuli were directly sent to the EEG recording system.

2.4 Audio stimuli

The sound scenes contained ambient noise in combination with isolated sounds potentially emanating from the experimenter behind the opaque curtains (e.g. falling keys, running water from a tap, leafing through a book, quietly walking around), but they were totally free from speech and music signals. Sounds were delivered thanks to a Focusrite Scarlett 6i6 external sound interface connected to a Mac mini computer. The listening level was adjusted to

doi:10.25836/sasp.2019.21

Figure 3. Recording setup.

60 dBA. The oddball experiment was repeated twice: once with the binaural recording (by a dummy head Neumann KU 100, see Fig. 3) and the other time with the stereophonic recording (by a XY pair, see Fig. 3). The presentation order was randomized. Intensity stereophony (i.e. XY recording) was chosen to provide the highest contrast with binaural spatialization which uses both interaural differences of time and level, as well as spectral cues. The binaural and stereophonic scenes were similar but not identical to prevent any habituation effect.

2.5 Preprocessing of EEG data

EEG signals were recorded with NetStation software and saved as raw files with triggers. They were then preprocessed using custom scripts based on Fieldtrip [11] (http://fieldtriptoolbox.org) and MNE-Python [12]. Preprocessing followed a predefined pipeline consisting of the following steps.

i) Channel removal (channels for which experimenters noticed high impedance values during subjects recordings, mostly in the neck and on the face of participants) and raw signal inspection, annotation and filtering (0.5Hz-45Hz).

ii) Isolating signals of interest around trials onsets (from -200ms to 800ms) and semi-automatic artefact annotations (jumps and muscle activity).

iii) Visual inspection of annotated artifacts, channels to be interpolated and trials to eliminate.

iv) Independent component analysis (ICA) on raw trials with annotated time segments and channels from steps i, ii and iii.

v) Visual inspection of components (the sixty first ones) and removal of components sharing blinks, saccades and cardiac activity mostly.

vi) Back-projection of left components in the channel space for signal reconstruction, removal of annotated artefacts or trials and interpolation of annotated channels.

vii) Calculation of event related potentials (ERP) and statistical analysis.

At the end of the aforementioned preprocessing steps, we obtain 124 electrodes left across all participants. The data set consisted of 6080 trials, 1045 of which were deviants. ERPs were averaged per electrode and per subject for each condition. The differential response between deviant and standard stimuli was examined as a function of the audio distractor (binaural or stereophonic).

2.6 Analysis of EEG data

We first ran a cluster analysis (number of permutations = 1000) using Fieldtrip [11], to identify groups of adjacent electrodes showing a P300 effect in both stereo and binaural conditions. We then used these channels to build regions of interest (ROI) wherein we looked for differences in the P300 amplitude for stereophonic vs. binaural sounds. More exactly, the differential response between deviant and standard stimuli is compared between the stereophonic and binaural conditions (i.e. $[\text{Deviant} - \text{Standard}]_{\textbf{binaural}} - [\text{Deviant} - \text{Standard}]_{\textbf{stereophonic}}$). ROI were made of equal numbers of channels (n=17). The P300 amplitude difference was analyzed on a 20 ms sliding window from 280 ms to 440 ms with a 10 ms overlap.

3. RESULTS

3.1 Behavioral results

Our paradigm integrates sounds as interference disturbing the main task. The literature has already reported that binaural sounds are more immersive than stereophonic sounds [13]. Consequently, our hypothesis was that accuracy scores and reaction times would be respectively decreased and increased in the binaural condition compare to the stereophonic condition. Accordingly, we used unilateral Student paired t-test to report behavioral results. In terms of accuracy, we did not found any difference between the two listening conditions (binaural vs. stereophonic) with p=0.46 for missed deviants and p=0.064 for

doi:10.25836/sasp.2019.21

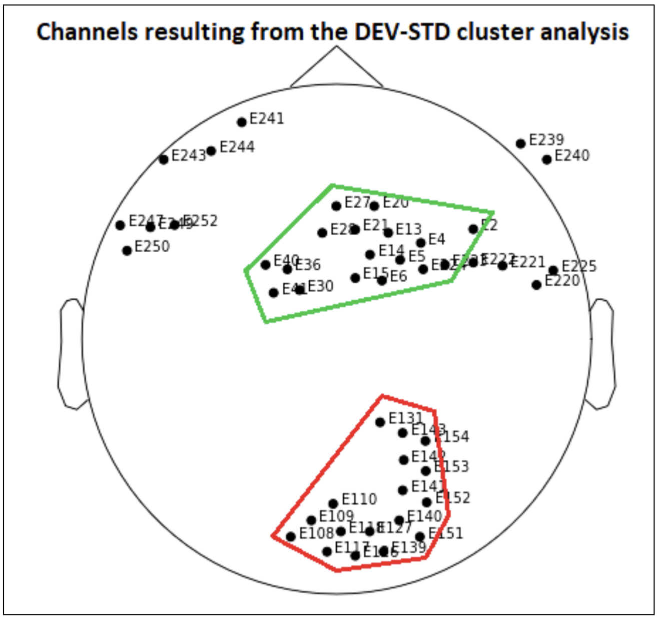

Figure 4. Channels exhibiting a P300 response resulting from the cluster analysis of the contrast DEVIANT-STANDARDS for both conditions (stereophonic and binaural sounds).

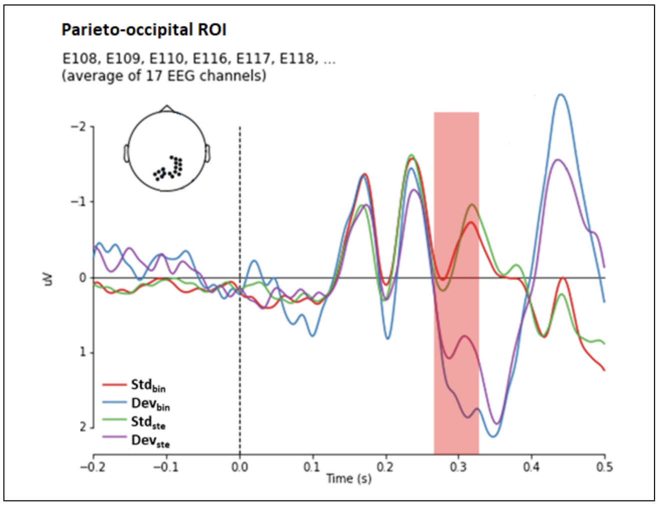

Figure 5. ERPs averaged over the 17 channels of the parieto-occipital region. Blue and purple curves correspond to $\text{Deviant}_{\textbf{binaural}}$ and $\text{Deviant}_{\textbf{stereo}}$ respectively, while red and green curves correspond to $\text{Standard}_{\textbf{binaural}}$ and $\text{Standard}_{\textbf{stereo}}$ respectively. The blue and purple curves differ significantly over the time period highlighted in pink (from 280 ms to 319 ms).

false recognitions. However, reaction times (RT) were significantly longer for the binaural condition than for the stereophonic condition with p=0.027 (binaural mean RT = 566 ms and stereophonic mean RT = 550 ms). The fastest participant used to press the response button on average 440 ms after the image onset. His average RT was chosen to stop the signal analysis and prevent any interpretation of signals after this delay.

3.2 EEG results

3.2.1 Cluster analysis

Inspection of individual ERPs revealed that the P300 was not starting before 280 ms in our dataset. Testing for a P300 effect in the latency range from 280 to 440 ms post-stimulus, the cluster-based permutation test showed a significant difference between the deviant and standard stimuli, either with binaural or stereophonic interferences ($p<0.05$). In this latency range, the difference was most pronounced over two central groups of channels (frontal and parieto-occipital). For the binaural condition, the analysis indicated the presence of 31 significant clusters within this time period while only 18 significant clusters were found for the stereophonic condition. The channels that form clusters in both conditions are presented on Fig. 4. We created two ROI from this channel set to run the amplitude comparison analysis. The first ROI groups 17 channels in the frontal part of the brain (area delimited by a green line on Fig. 4) and the second ROI groups 17 channels in the parieto-occipital region (area delimited by a red line on Fig. 4).

3.2.2 Compared analysis of P300 amplitude

The comparison of P300 amplitude consists in investigating whether deviant visual stimuli generate larger amplitude of the P300, or part of it, in a given condition compared to the other (binaural or stereophonic). We observed that only the binaural condition was associated with larger P300 in both ROI. The effect was precocious for the parieto-occipital ROI and was immediately followed by the frontal ROI. After 350 ms, deviants did not produce different P300 signals. Looking at ERP curves in each ROI (Fig. 5 and Fig. 6) and statistical data of Tab. 1, we identified that only the early phase of the P300 is concerned. This phase corresponds to the P3a. The second phase of the P300 - clearly observable on Fig. 5 (second peak) - seems to correspond to the P3b. This phase was not affected by sound interferences.

4. DISCUSSION

Taken together our results are in favor of a better surprise effect mediated by binaural sounds when compared to stereophonic sounds. We observed that the P300 amplitude was significantly higher in presence of the binaural distractor especially during its first phase known as the P3a [14]. This part of the P300 is described as reflecting surprise or novelty among sensory inputs or at least transitory losses of attentional focus when competing tasks interfere. Associated with longer reaction times, an increase of the P3a has already been described accounting for attentional switching even using multimodality [15,16]. In these studies, authors aimed at quantifying the physical similitudes or differences between relevant and non-relevant stimuli (Go-NoGo task and oddball tasks) to see which amount of them is needed to affect the electrophysi-

doi:10.25836/sasp.2019.21

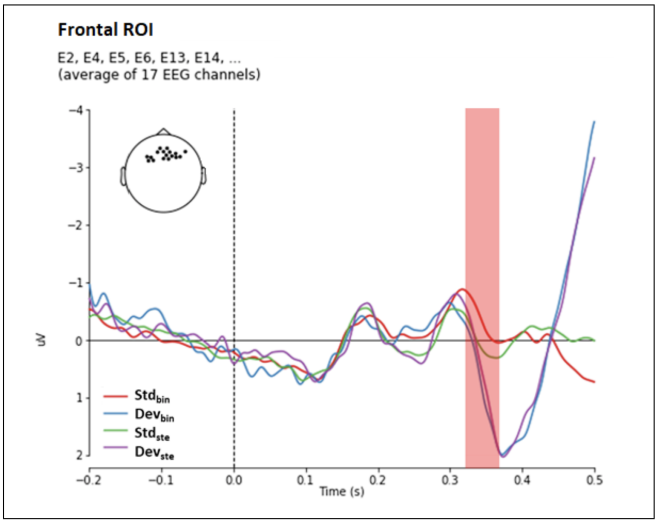

Figure 6. ERPs averaged over the 17 channels of the frontal region. Blue and purple curves correspond to Deviant$_{\mathbf{binaural}}$ and Deviant$_{\mathbf{stereo}}$ respectively while red and green curves correspond to Standard$_{\mathbf{binaural}}$ and Standard$_{\mathbf{stereo}}$ respectively. In the time period where the difference between the ERP of deviant stimuli is significant and which is highlighted in pink (from 320 ms to 359 ms), the difference between the red and blue curves is slightly larger than the difference between the purple and green curves, suggesting that deviant stimuli were affected in the binaural condition and not in the stereophonic one.

Time period (ms)	Parieto-occipital ROI	Fronto-central ROI
[280-299]	T= 2.52, p=0.021*	T=-0.21, p=0.833
[290-309]	T=2.64, p=0.017*	T=-0.50, p=0.621
[300-319]	T=2.74, p=0.013*	T=-1.27, p=0.222
[310-329]	T=2.22, p=0.039*	T=-2.07, p=0.053
[320-339]	T=1.43, p=0.170	T=-3.04, p=0.007*
[330-349]	T=0.77, p=0.453	T=-3.41, p=0.003*
[340-359]	T=0.34, p=0.740	T=-2.74, p=0.014*
[350-369]	T=0.13, p=0.901	T=-1.72, p=0.103
[360-379]	T=0.13, p=0.898	T=-0.81, p=0.429
[370-389]	T=-0.02, p=0.985	T=-0.39, p=0.697
[380-399]	T=-0.57, p=0.578	T=-0.32, p=0.755
[390-409]	T=-1.02, p=0.323	T=-0.63, p=0.536
[400-419]	T=-1.08, p=0.296	T=-0.40, p=0.690
[410-429]	T=-1.13, p=0.272	T=0.21, p=0.840
[420-439]	T=-1.32, p=0.204	T=0.24, p=0.812

Table 1. Detailed analysis of the contrast [Deviant − Standard]$_{\mathbf{binaural}}$ − [Deviant − Standard]$_{\mathbf{stereophonic}}$. The P300 period is divided into 20 ms time windows overlapping by 10 ms.

ological response. Indeed, Comerchero and Polich explain that the effect of target and non-target stimuli considering task difficulty should be considered [17]. Especially, it is shown that P3a is increased if the target/standard discrimination is difficult while the non-target/standard stimulus difference is large. For example, Schröger and Wolff, who used only audio stimuli in a three stimulus oddball paradigm, showed that task-irrelevant frequency deviants elicited MMN, N2b, and P3a components, and caused impoverished behavioral performance to targets [18]. Distractors of our experiment, and more generally in interference task, play the exact same role as the irrelevant targets in the three stimulus oddball paradigm introduced by [19]. When using multimodality, the distance between deviants and distractors is such that the only possible interpretation would be a loss of attentional focus (see [18] for a discussion between memory and attentional theory involvement given distractor and deviant similarities). In our task, the only difference between sounds lies in their recording qualities. Our results suggest that observed electrophysiological and behavioral differences are a consequence of recording qualities. We deliberately chose to use sounds recorded from the experimental room placing ourselves in the context of hyper-realism as one of the best inducer for immersion.

5. CONCLUSION

Binaural and stereophonic reproduction were compared by neuroimagery according to an oddball paradigm. It was shown that the P300 amplitude was significantly affected when the audio distractor is binaural. A similar tendency was observed on the reaction time of visual detection, which was longer in presence of binaural sound. Our study introduced a promising paradigm to objectively assess immersive properties in audio media.

6. ACKNOWLEDGMENTS

This study was co-funded by the LabEx Comin-Labs under the Neural Communications project, and by Orange. We also wish to thank all the participants that undertook the study, as well as the staff from the epileptology unit at the University Hospital.

7. REFERENCES

[1] R. Nicol, L. Gros, C. Colomes, M. Noisternig, O. Warusfel, H. Bahu, B. Katz, and L. Simon, “A roadmap for assessing the quality of experience of 3d audio binaural rendering,” 2014.

[2] Z. Akhtar and T. H. Falk, “Audio-visual multimedia quality assessment: a comprehensive survey,” *IEEE Access*, vol. 5, pp. 21090–21117, 2017.

[3] K. u. R. Laghari, R. Gupta, S. Arndt, J. N. Antons, R. Schleicher, S. Møller, and T. H. Falk, “Neurophysiological experimental facility for Quality of Experience (QoE) assessment,” in *2013 IFIP/IEEE International*

Symposium on Integrated Network Management (IM 2013), (Ghent, Belgium), pp. 1300–1305, May 2013.

[4] R. Gupta, K. Laghari, H. Banville, and T. H. Falk, "Using affective brain-computer interfaces to characterize human influential factors for speech quality-of-experience perception modelling," *Human-centric Computing and Information Sciences*, vol. 6, Dec. 2016.

[5] K. u. R. Laghari, R. Gupta, S. Arndt, J. N. Antons, S. Møllery, and T. H. Falk, "Characterization of human emotions and preferences for text-to-speech systems using multimodal neuroimaging methods," in *2014 IEEE 27th Canadian Conference on Electrical and Computer Engineering (CCECE)*, (Toronto, ON, Canada), pp. 1–5, May 2014.

[6] T. D. Mrsic-Flogel, A. J. King, R. L. Jenison, and J. W. Schnupp, "Listening through different ears alters spatial response fields in ferret primary auditory cortex," *Journal of Neurophysiology*, vol. 86, pp. 1043–1046, Aug. 2001.

[7] K. J. Palomäki, H. Tiitinen, V. Mäkinen, P. J. C. May, and P. Alku, "Spatial processing in human auditory cortex: the effects of 3D, ITD, and ILD stimulation techniques," *Brain Research. Cognitive Brain Research*, vol. 24, pp. 364–379, Aug. 2005.

[8] M. G. Wisniewski, G. D. Romigh, S. M. Kenzig, N. Iyer, B. D. Simpson, E. R. Thompson, and C. D. Rothwell, "Enhanced auditory spatial performance using individualized head-related transfer functions: An event-related potential study," *The Journal of the Acoustical Society of America*, vol. 140, pp. EL539–EL544, Dec. 2016.

[9] S. Getzmann and J. Lewald, "Effects of natural versus artificial spatial cues on electrophysiological correlates of auditory motion," *Hearing Research*, vol. 259, no. 1, pp. 44 – 54, 2010.

[10] A. Callan, D. E. Callan, and H. Ando, "Neural correlates of sound externalization," *NeuroImage*, vol. 66, pp. 22–27, Feb. 2013.

[11] R. Oostenveld, P. Fries, E. Maris, and J. M. Schoffelen, "Fieldtrip: open source software for advanced analysis of meg, eeg, and invasive electrophysiological data," *Computational intelligence and neuroscience*, vol. 1, 2011.

[12] A. Gramfort, M. Luessi, E. Larson, D. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, and M. Hämäläinen, "Meg and eeg data analysis with mne-python," *Frontiers in Neuroscience*, vol. 7, 2013.

[13] R. Nicol, L. Gros, C. Colomes, E. Roncière, and J.-C. Messonnier, "Etude comparative du rendu de différentes techniques de prise de son spatialisée aprs binauralisation," in *Congrès Français d'Acoustique*, SFA, 2016.

[14] J. Polich, "Updating p300: an integrative theory of p3a and p3b," *Clinical neurophysiology*, vol. 118, no. 10, pp. 2128–2148, 2007.

[15] K. Ahlo, C. Escera, R. Díaz, E. Yago, and J. Serra, "Effects of involuntary auditory attention on visual task performance and brain activity," *NeuroReport*, vol. 8, p. 32333237, 1997.

[16] J. I. Katayama and J. Polich, "Auditory and visual p300 topography from a 3 stimulus paradigm," *Clinical Neurophysiology*, vol. 110, no. 3, pp. 463–468, 1999.

[17] M. D Comerchero and J. Polich, "P3a, perceptual distinctiveness, and stimulus modality," *Cognitive Brain Research*, vol. 7, no. 1, pp. 41–48, 1998.

[18] E. Schröger and C. Wolff, "Behavioral and electrophysiological effects of task-irrelevant sound change: A new distraction paradigm," *Cognitive brain research*, vol. 7, no. 1, pp. 71–87, 1998.

[19] J. I. Katayama and J. Polich, "P300 from one-, two-, and three-stimulus auditory paradigms," *International Journal of Psychophysiology*, vol. 23, no. 1-2, pp. 33–40, 1996.

doi:10.25836/sasp.2019.21