Perceptual assessment of binaural sound

Orange Labs

Rozenn Nicol, Laetitia Gros, Cathy Colomes, Pierre-Yohan Michaud

Research & Development
24/11/2015, Sense Camp

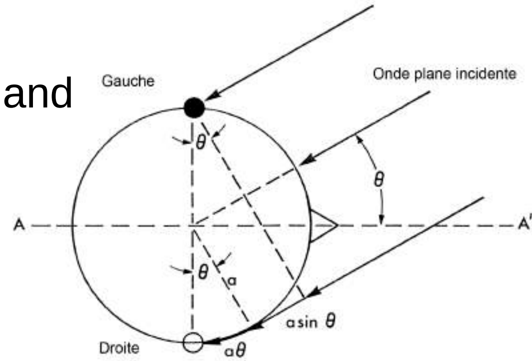
Summary

- some brief **reminders** on
  - spatial hearing
  - and binaural sound
- a short overview of the **BiLi Project**
  - **individualized binaural downmix** (or upmix?) of multichannel audio
    - TV or radio programmes
    - to improve sound immersion
  - a sub-project dedicated to the **QoE of binaural sound**
- an open question:
  - **how to perceptually assess spatial sound?**
- **work in progress** in the BiLi Project
  - new tools/ methods
  - experiments

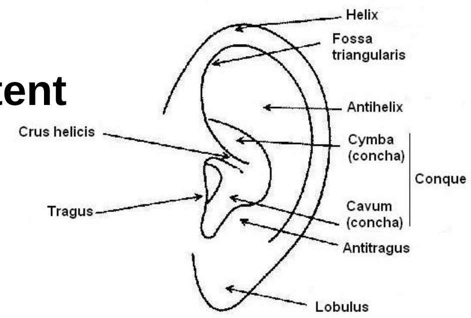
SPATIAL HEARING

How the auditory system is able to localize a sound?

- **localization in the horizontal plane**
  - (mainly) based on **interaural** cues
    - time/phase and time differences between the left and right ears (Duplex theory of Lord Rayleigh)
      - **ITD**: Interaural Time Difference
      - **ILD**: Interaural Level Difference



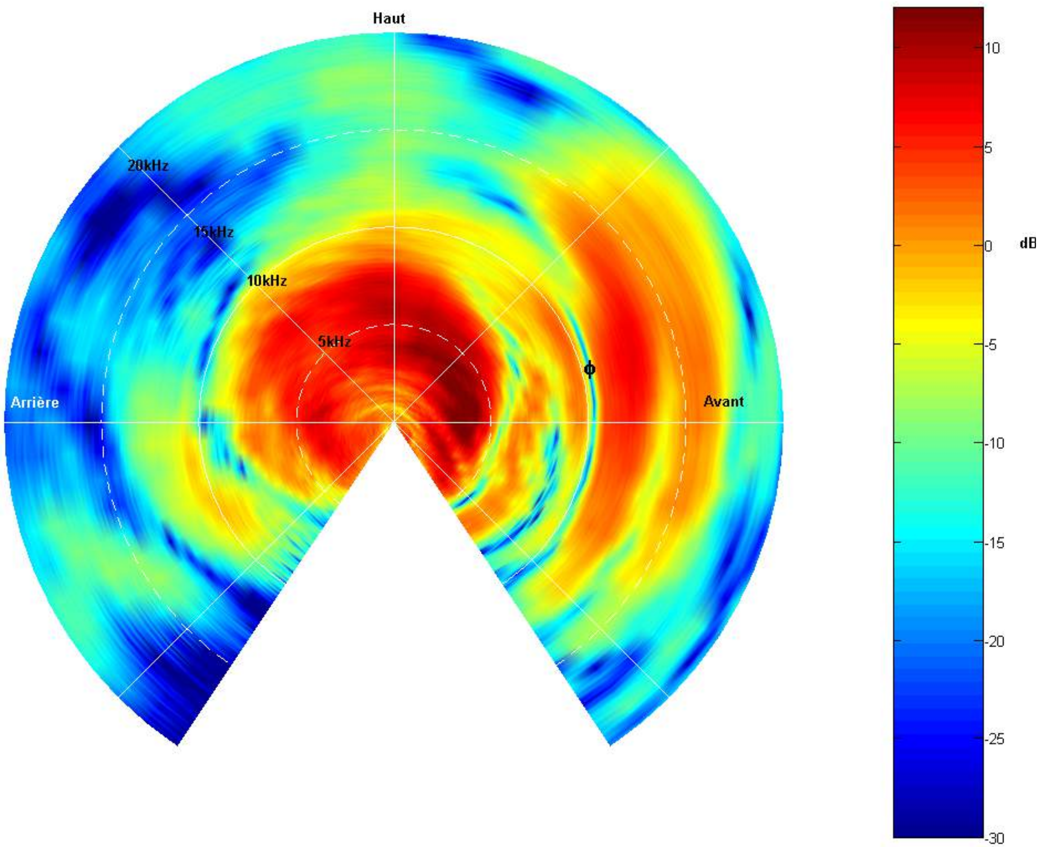
- **localization in the median plane**
  - (mainly) based on **monaural** cues
    - the information available at each ear is analysed separately and independently
  - Experiments of « hearing phantoms » suggest that the perception of elevation is controlled by the **spectral content** of the auditory event
    - concept of **Spectral Cues** (SC)
  - originating from **pinna**'s resonances



HRTF

- = Head Related Transfer Function
  - **transfer function** between the sound source (located in 3D space) and the listener's ears
  - accounts for the **interaction** between the **acoustic wave** and the **listener's body**
    - reflection & diffraction by the head/ pinna/ torso
  - contains all the localization cues
    - **ITD, ILD, SC**

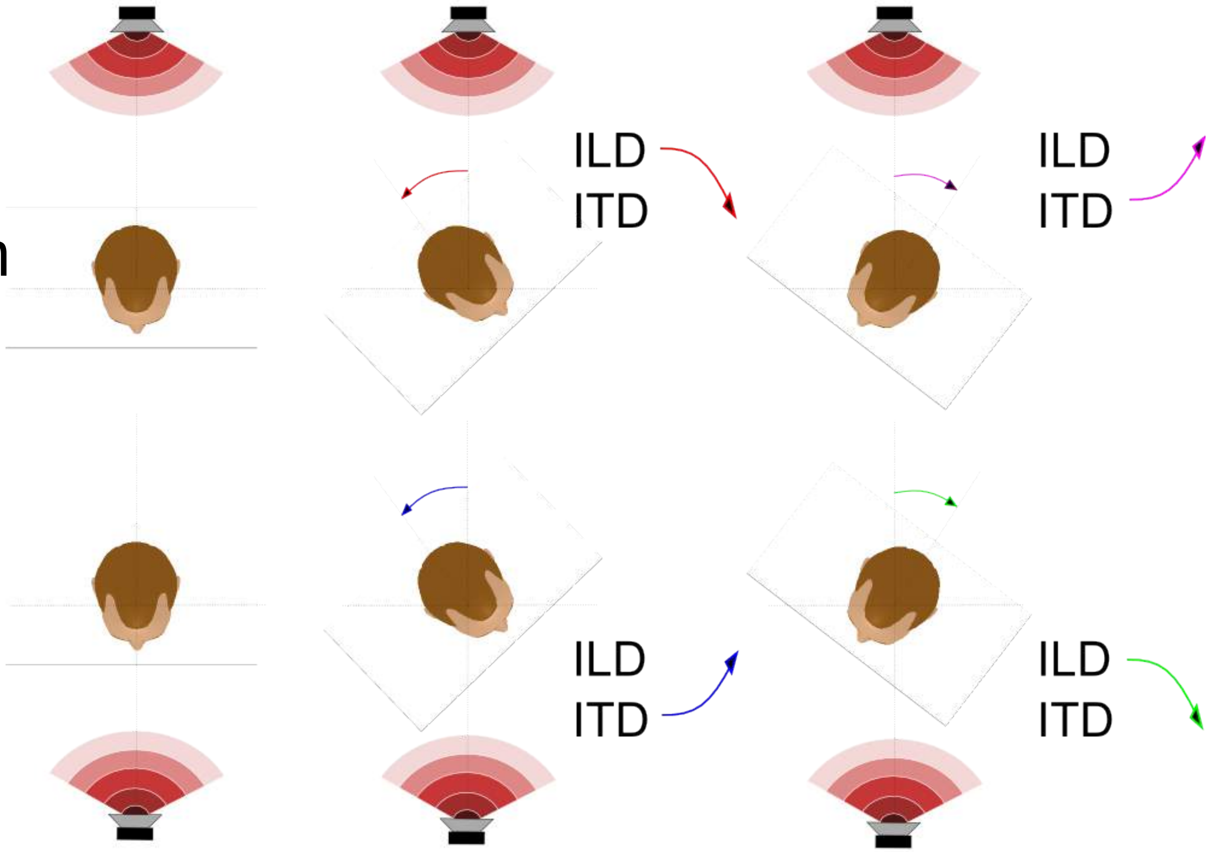
HRTF measured for one individual in the median plane



- the spectrum pattern strongly varies as a function of elevation
  - peaks and notches
- and therefore can be used by the auditory system as an elevation cue

Dynamic cues of localization

- the listener is able to use the information from (very) **little head's movements**
  - mainly to solve **front/ rear ambiguity**



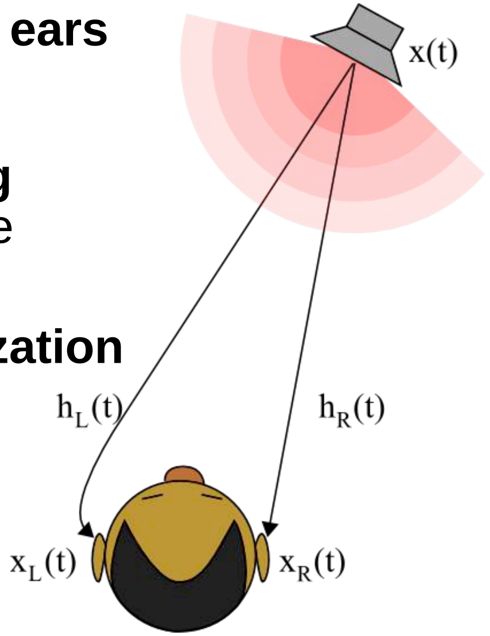
Localization of distance

- the auditory system is **not efficient** to perceive the **absolute** distance of a sound source
  - experiments show that the localization of absolute distance is easy to be **fooled**
    - visual cues
    - cognition
- but the **differential** localization of distance is more accurate
  - main localization **cues**
    - **Sound level**
    - **Direct-to-reverberant ratio**
    - **Timbre**

BINAURAL SOUND

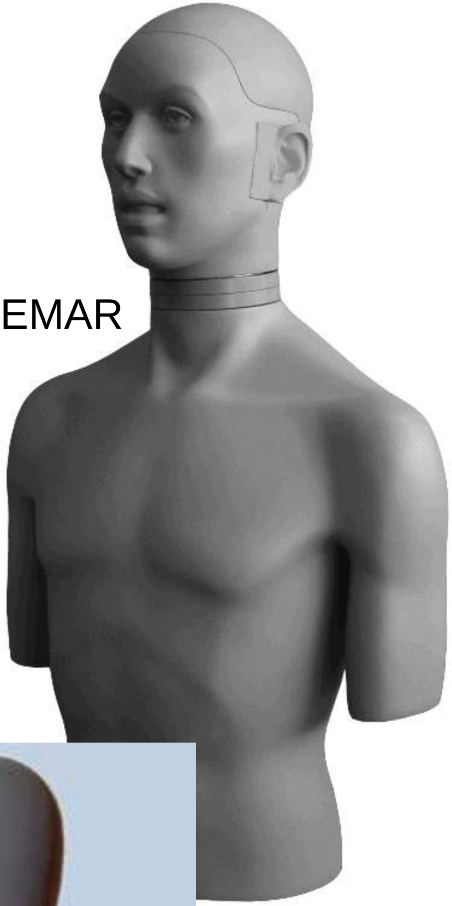
The technology of sound spatialization the closest to natural listening

- Natural listening
  - **3D audio** perception and localization with **only 2 ears**
- Binaural technologies aim at **mimicing** this principle
  - by **recording** (*natural recording*) or **synthesizing** (*binaural synthesis*) the 2 signals perceived at the entrance of the ear canals
- **Spatial encoding** based on the reproduction of **localization cues**
  - **ITD** and **ILD**
  - **SC**
- "**Full 3D**" spatialization
  - The virtual sound sources are localized in any direction over the 3D sphere
    - azimuth & elevation

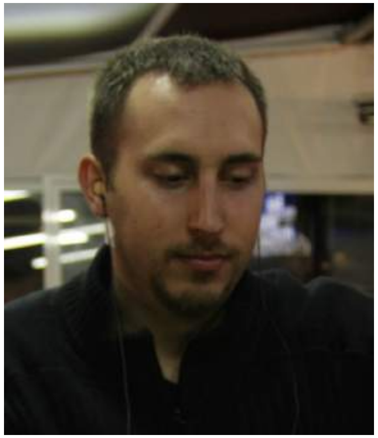


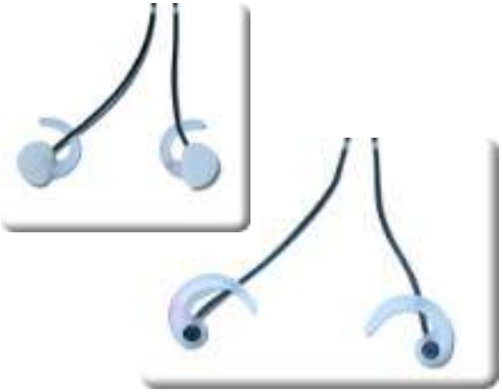
Binaural recording
by natural or dummy heads

Neumann
KU100

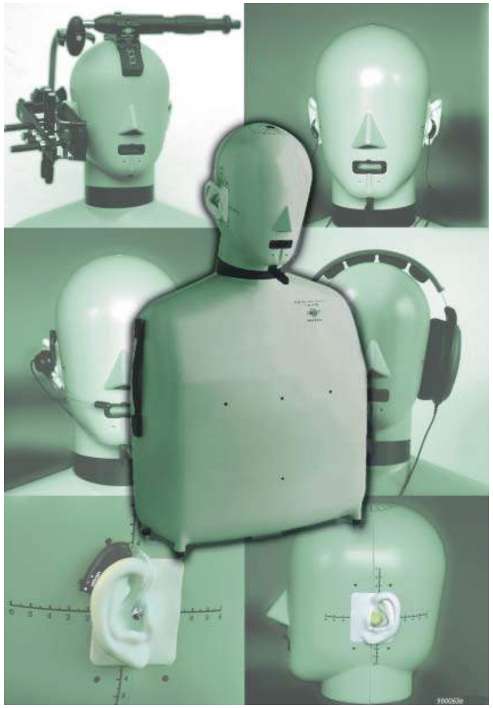


KEMAR





Sound Professionals:
SP-TFB-2 binaural
microphones

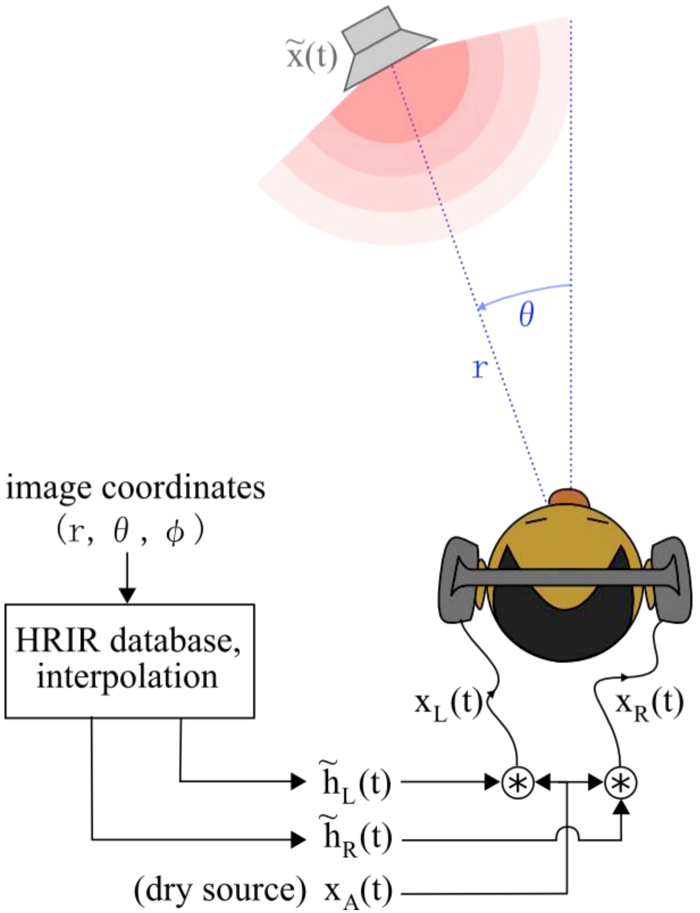


Brüel & Kjær HATS

Head Acoustics HSMIII

Binaural synthesis

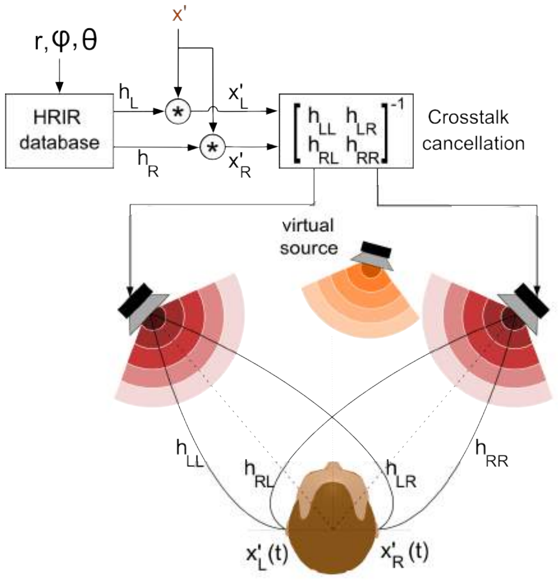
- a **virtual** recording
  - by synthesizing the 2 signals which would have been recorded by 2 microphones inserted at the entrance of the ear canal
- How?
  - by filtering a monophonic signal (1 signal per virtual source) by a pair of **binaural filters**
  - each binaural filter is computed from a **HRTF**
- **static** or **dynamic** synthesis
  - reproduction of the **dynamic cues** of localization



Beyerdynamic Headzone ® (ultrasonic)

Binaural rendering

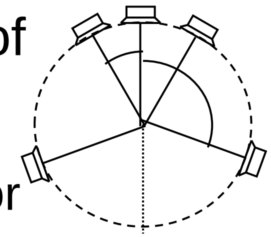
- **headphones** are the **recommended** equipment to listen to binaural signals
  - **headphone calibration** is required
    - compensation of the HPTF
  - but a tricky issue!...
    - depends on the **individual** and the **positioning**
- rendering over **loudspeakers** is also possible
  - crosstalk cancellation is needed
  - examples of system
    - **transaural®** [**Cooper & Bauck** 1988] [Atal & Schroeder 1966]
    - double transaural
    - **stereo dipole**



BILI PROJECT

Main objectives

- ✓ to develop **customized** tools for **binaural conversion** of **multichannel** audio contents
  - – to create a **virtual 5.1 setup** (or 7.1/ 10.2/ 22.2/ HOA) for binaural listening over headphones
    - to promote multichannel audio
    - but without needing new equipments
  - – use of **individualized HRTF**
    - for a better **sound immersion**
- ✓ in combination with a **standard** format of HRTF
  - – **AES 69** (April 2015)
    - **SOFA** representation of HRTF
    - to ensure the compatibility with any device (computer, smartphone, tablette…)



binaural downmix
or
« binauralization »

Consortium

6 Sub-Projects

- SP1: Assessment of the **Quality of Experience** of binaural sound
- SP2: **Production** of contents
- SP3: HRTF **standard**
- SP4: Acquisition/modeling of **individualized** HRTF
- SP5: **Prototypes** of customized binaural sound
- SP6: **Dissemination**

PERCEPTUAL ASSESSMENT OF SPATIAL SOUND

How to assess the perception of spatial sound reproduction?

- 2 issues to be investigated:

  1) to **identify** the perceptual **dimensions** of this perception

  – eg. 2 dimensions are already known:

  – **timbre**

  – **position** of sources

  – but there are potentially other dimensions:

  – **immersion**, **naturalness, plausibility**, **emotion** …

  2) to **develop methods** to **« measure »** the perception as a function of each of these dimensions

Perceptual dimensions

- **methods**
  - **MDS** (Multi Dimensional Scaling)
    - to identify the dimensions of perceptual space
  - **verbal description** of perception
    - eg: *Descriptive Analysis* (DA), *Repertory Grid Technique* (RGT), *Free-Choice Profiling* (FCP)
- **several perceptual attributes** already identified
  - linked to **physics**, eg:
    - timbre
    - position of the source
      - spherical coordinates, ability to localize, externalization
    - spatial width of the source
    - room effect
  - but also linked to **cognitive** and **affective**, eg:
    - naturalness, lisibility, emotion, pleasantness

Perceptual attributes: the example of SAQI

- **SAQI**: Spatial Audio Quality Inventory (Project SEACEN)
  - « *a consensus vocabulary for the perceptual assessment of virtual acoustic environments (VAEs), or of Spatial Audio technology in general* »
  - « to allow qualitatively differentiated assessments of unimodal or supramodal auditory differences between technically generated acoustic environments (VAES) as well as with respect to a presented or imagined acoustic reality »
  - list of **48 verbal descriptors of perceptual** attributes
    - 8 categories: *timbre, tonalness, geometry, room, time behavior, dynamics, artifacts, general impressions*
  - initiated in **german**
    - translated in **english**
    - in progress: translation in **french**

Lindau *et al.*, 2014

how to measure?

- **direct measure of perceptual** attribute without reference
  - to be defined:
    - scale of assessment (range, continuous or not, …)
  - examples
    - description of the acoustic quality of a room (eg: *presence*, *room effect*, *reverberance*, …)
    - absolute localization test
- **direct measure of perceptual** attribute **with a reference**
  - the stimulus under assessment is compared to a reference
    - possibility to use **anchors** in addition
  - examples
    - AB, ABX, MUSHRA test
    - relative localization test (acoustic pointer)

how to measure?

- **indirect measure**
  - **task performance**
    - eg: description of the sound scene (ie. to report the number, the nature, and the location of the sound sources)
    - the QoE is inferred from the success of the task
  - **physiological** measures
    - to observe cognitive, emotional, or behavioral phenomena by analyzing the physiological responses
      - eg: electrodermal activity, heart pulse, skin temperature, or eye activity
  - **brain imagery**
    - MRI, electroencephalogram, magnetoencephalogram, …
    - but some compatibility issue with sound reproduction

PERCEPTUAL EXPERIMENTS WITHIN BILI PROJECT

perceptual assessment of binaural sound in BiLi

- « **conventional** » methods
  - <u>validation of individual measured HRTFs</u>
    - test of **absolute localization**
      - choice of the method to report localization judgment (H. Bahu & O. Warusfel, IRCAM)
  - <u>assessment of modelled HRTFs</u>
    - comparative assessment of various solutions of individualization
    - test of **relative localization**

perceptual assessment of binaural sound in BiLi

- « **innovative** » methods
  - assessment of the quality of spatial trajectories (A. Andreopoulou & B. Katz, LIMSI)
    - to compare individual vs non-individual HRTFs
    - to evaluate the impact of processing: eg. modelling of binaural filters
  - feasibility study of consumer EEG biosensors to measure the QoE of binaural sound (A. Boisadan, L. Gros & R. Nicol, 2014)
  - identification of perceptual attributes of non individual HRTFs
    - comparison of binaural stimuli synthesized with various sets of non individual HRTFs
    - 2 studies: **verbal** description (L. Simon & B. Katz, LIMSI)/ **MDS** (PY. Michaud & R. Nicol, Orange)
  - comparison of the robustness of sound recording systems after binauralization (CNSMDP, RF, Orange)
    - 2 experiments combining various methods (graphical description of source position or trajectory, preference judgement, multi criteria assessment)

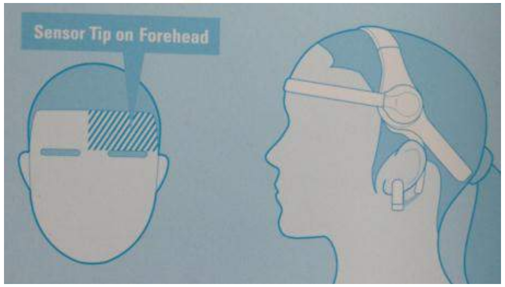
perceptual assessment of binaural sound in BiLi

- « **innovative** » methods
  - assessment of the quality of spatial trajectories (A. Andreopoulou & B. Katz, LIMSI)
    - to compare individual vs non-individual HRTFs
    - to evaluate the impact of processing: eg. modelling of binaural filters
  - feasibility study of consumer EEG biosensors to measure the QoE of binaural sound (A. Boisadan, L. Gros & R. Nicol, 2014)
  - identification of perceptual attributes of non individual HRTFs
    - comparison of binaural stimuli synthesized with various sets of non individual HRTF
    - 2 studies: **verbal** description (L. Simon & B. Katz, LIMSI)/ **MDS** (PY. Michaud & R. Nicol, Orange)
  - comparison of the robustness of sound recording systems after binauralization (CNSMDP, RF, Orange)
    - 2 experiments combining various methods (graphical description of source position or trajectory, preference judgement, multi criteria assessment)

Feasibility study: Mindwave headset by Neurosky

A. Boisadan, L. Gros & R. Nicol, 2014

- **comparison of monaural vs binaural sound**
  - 4 excerpts: hairdryer, thunder, traffic road, fiction story
- 15 participants
- recording of the signal delivered by the electrode
  - estimation of 2 parameters:
    - « **Attention** » and « **Meditation** »
  - **questionnaire** in addition

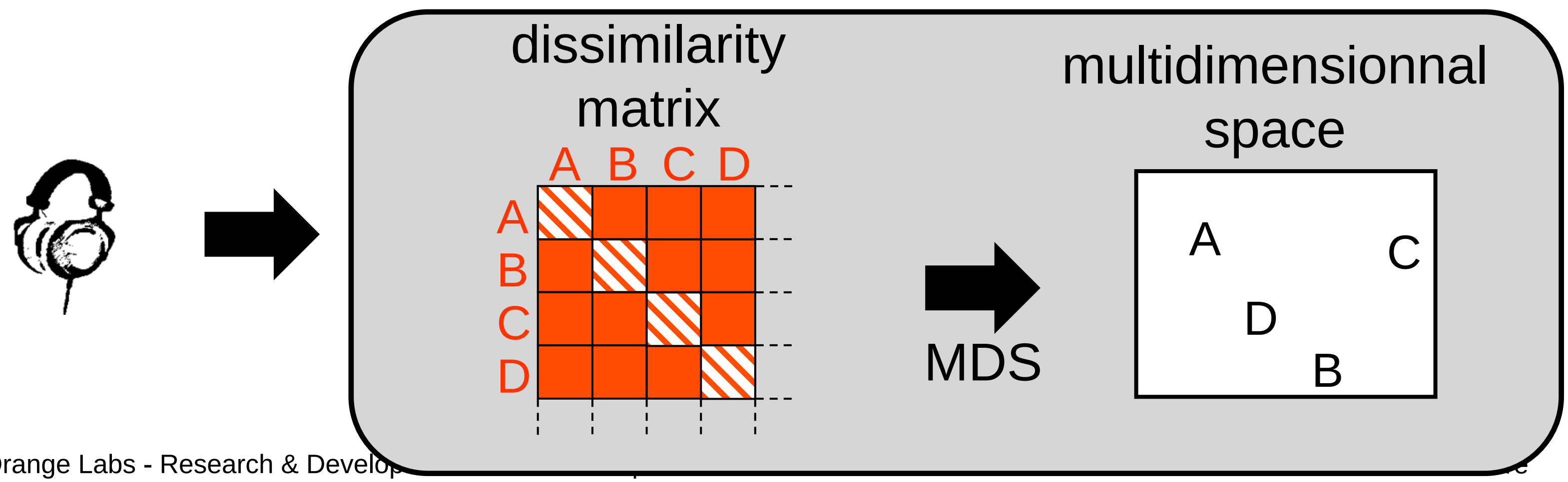


- results
  - the effect of the version (monaural/ binaural) is not significant
  - the signal « Meditation » is significantly affected by the type of excerpt
- next step
  - high resolution EEG (partnership with Univ. Rennes 1)

Identification of the dimensions of the perceptual space of non individual HRTFs

PY. Michaud & R. Nicol, 2014

- Objective:
  - to compare a large number of **non individual HRTF** sets
    - 46 sets (« Listen » Database / IRCAM)
  - to collect **dissimilarity judgements** between these 46 sets
  - in order to build the associate perceptual space by MDS (Multi Dimensional Scaling)
    - to identify the number of dimensions
    - to identify the perceptual(s) attribute(s) associated to each dimension



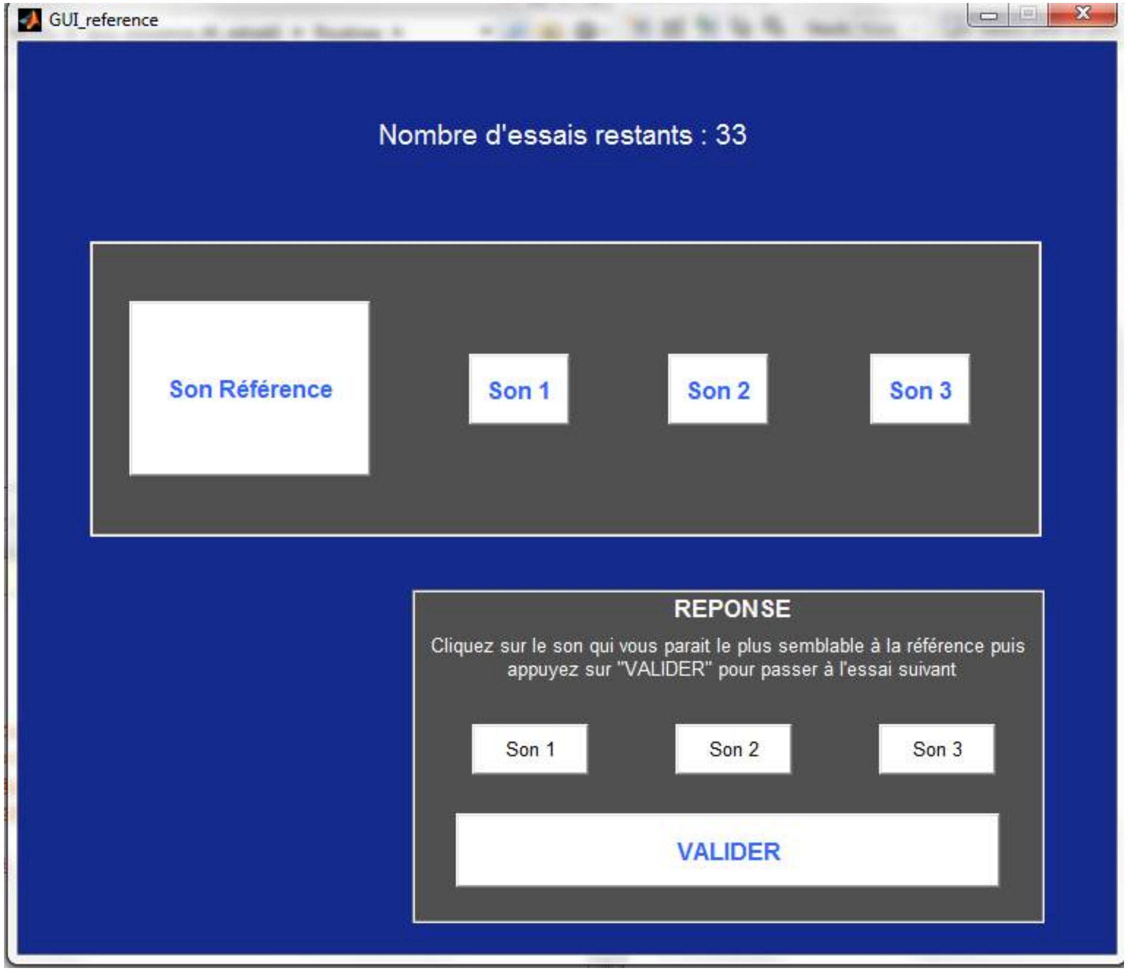
Identification of the dimensions of the perceptual space of non individual HRTFs

- experimental paradigm

PY. Michaud, 2013

  - **SPPR** method
    - *Similarity Picking with Permutation of References*
  - instead of presenting a pair of stimuli
    - for M = 46 versions: M(M-1)/2 = **1035** pairs!
    - a set of **4 stimuli** (« tetrad ») is presented
      - 1 stimulus (reference)
      - 3 stimuli under comparison
        - only 3 because of the limitation of auditory memory
  - the participant is asked to choose among the P=3 stimuli, the stimulus which he(she) perceives as the **most similar to the reference**
  - number of tetrads : M(M-1)/3 = **690**
    - on average 6-7 hours by participant !
  - advantage: **each stimulus** is used as the **reference**

Identification of the dimensions of the perceptual space of non individual HRTFs



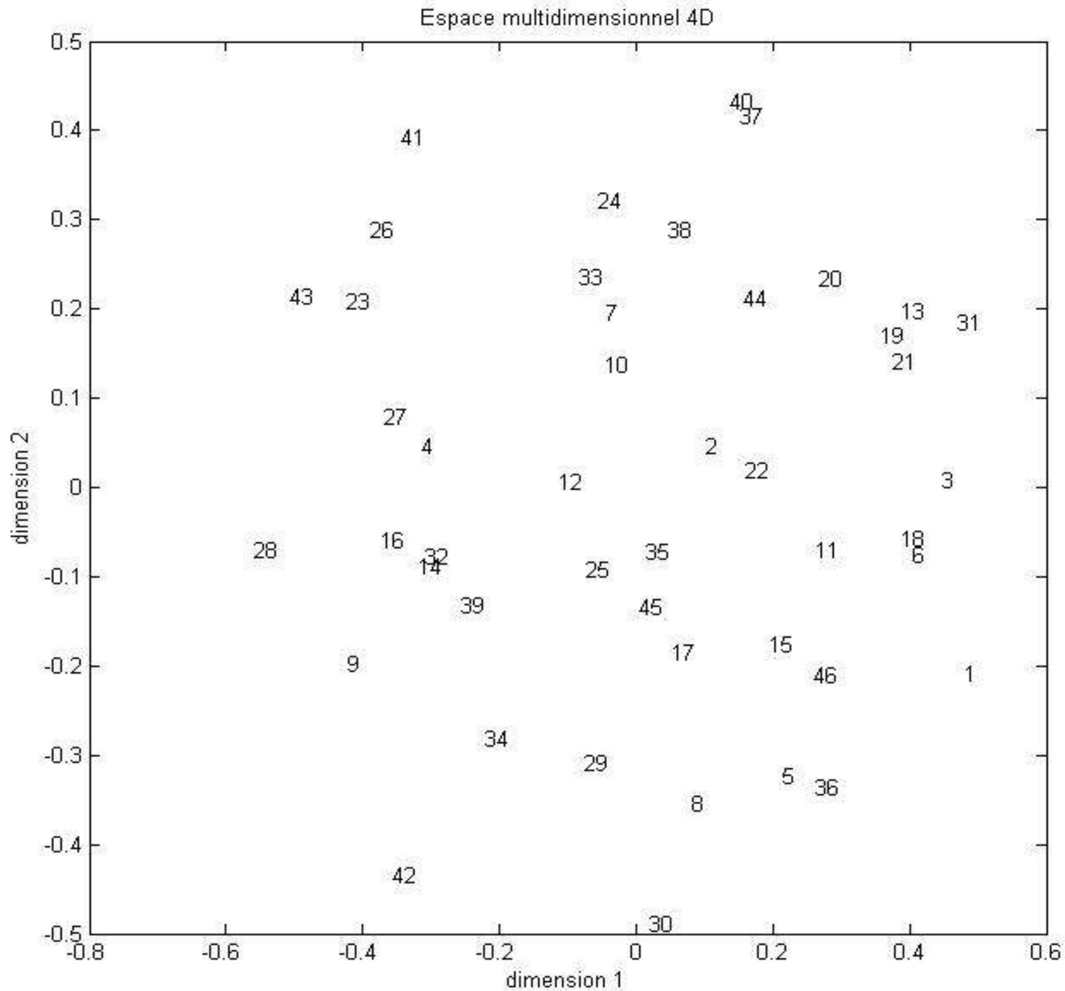
- sound stimulus:
  - excerpt of a radio programme (5.1)
- 10 « expert » participants
  - sound engineers (RF, FTV, CNMSDP),
  - researchers in 3D audio (Orange, LIMSI)

Identification of the dimensions of the perceptual space of non individual HRTFs

- **MDS** analysis of mean matrix of dissimilarities

<u>4 dimensions</u>

- dim. 1: **spectral modification** (amount of high frequencies)
- dim. 2: **size/ volume/ width** of the scene
- dim. 3: **spatial barycenter**
- dim. 4: **depth/ distance** of objects



Identification of the dimensions of the perceptual space of non individual HRTFs

4 dimensions

- dim. 1: **spectral modification** (amount of high frequencies)
- dim. 2: **size/ volume/ with** of the scene
- dim. 3: **spatial barycenter**
- dim. 4: **depth/ distance** of objects

link with the study of verbal description (L. Simon & B. Katz, LIMSI), based on a selection of 7 HRTFs

- 12 attributes
  - Externalization
  - Immersion
  - Spectral modification (low/ high freq.)
  - Realism
  - Depth of the soundfield
  - Accuracy of localization
  - Lateral position
  - Elevation
  - Front/back position
  - Reverberation
  - Sound level
  - Distance

confirmed by further investigations

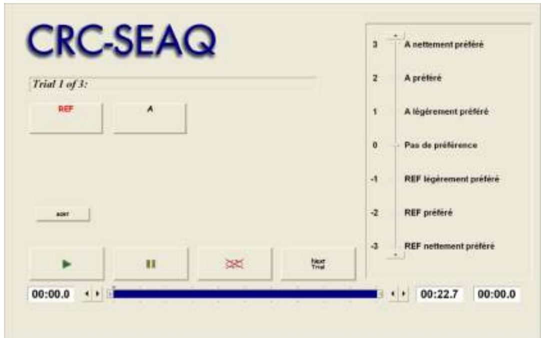
- subset of 19 sets of HRTFs
- « naive » participants
- a different type of stimulus (binaural mix, natural sounds)

Comparison of sound recording systems

- recording of a radio drama (« Deux femmes pour un fantôme », R. de Obaldia) at the CNSMDP
  - 14 systems
    - Neumann KU 100, 2 « home made » heads
    - Eigenmike, Soundfield®
    - stereo: XY, AB
    - various multichannel trees (Decca Tree, IRT, DPA 5100, …)

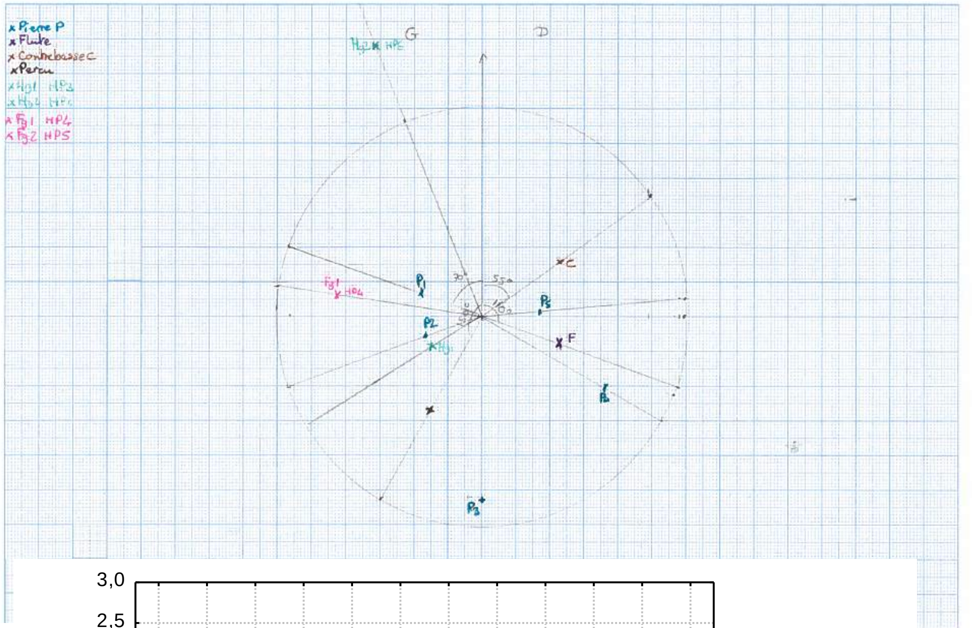

Comparison of sound recording systems

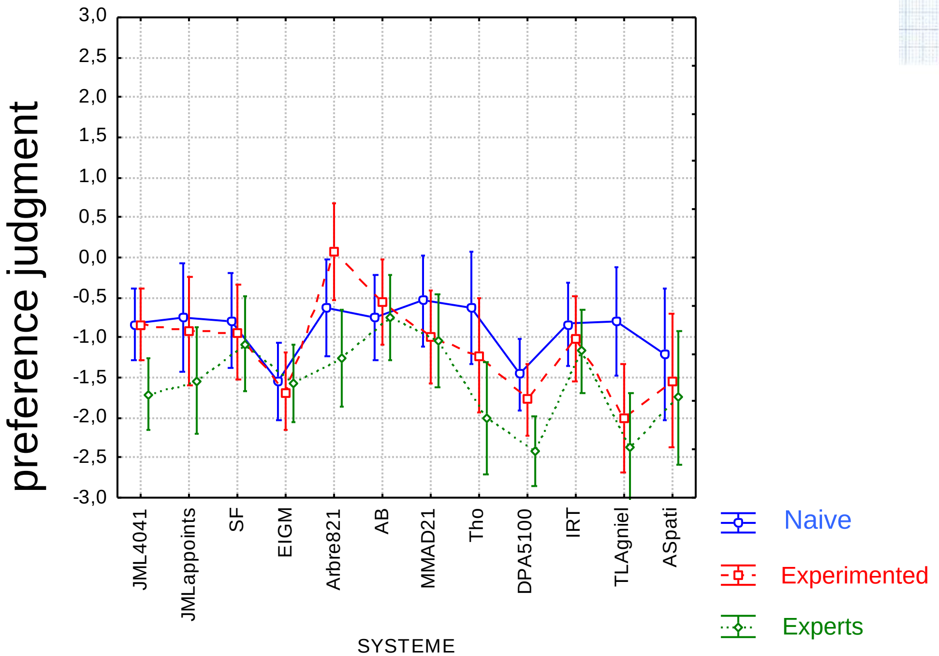
- Objective
  - to compare the **robustness** of sound recording after binauralization
    - the Ambisonics / stereo / multichannel recordings were mixed with the instruction of achieving optimal rendering
- 2 experimental paradigms
  - **test 1**: listening of a long sequence (1min 40sec)
    - **graphical report of source positions and trajectories**
    - in combination with a short questionnaire (3 attributes: realism, sensation of space, distance)
    - 12 participants
  - **test 2**: **pair comparison** of short sequences (20sec)
    - comparison with a reference (KU100)
    - **preference** judgement (7-level scale: -3 to +3)
    - extensive questionnaire at the end of the test
      - list of 10 attributes + general information
    - 24 participants



Comparison of sound recording systems

- some **preliminary results** (analysis still in progress!)
  - **expert vs naive** participants
    - expert judgements are clearly more differenciated
    - both for test 1 & 2
  - test 1
    - the analysis of graphs calls for innovative methods
  - test 2
    - most of preference judgments are < 0
      - KU100 is generally preferred





CONCLUSION

Perception of binaural sound: a lot of questions…

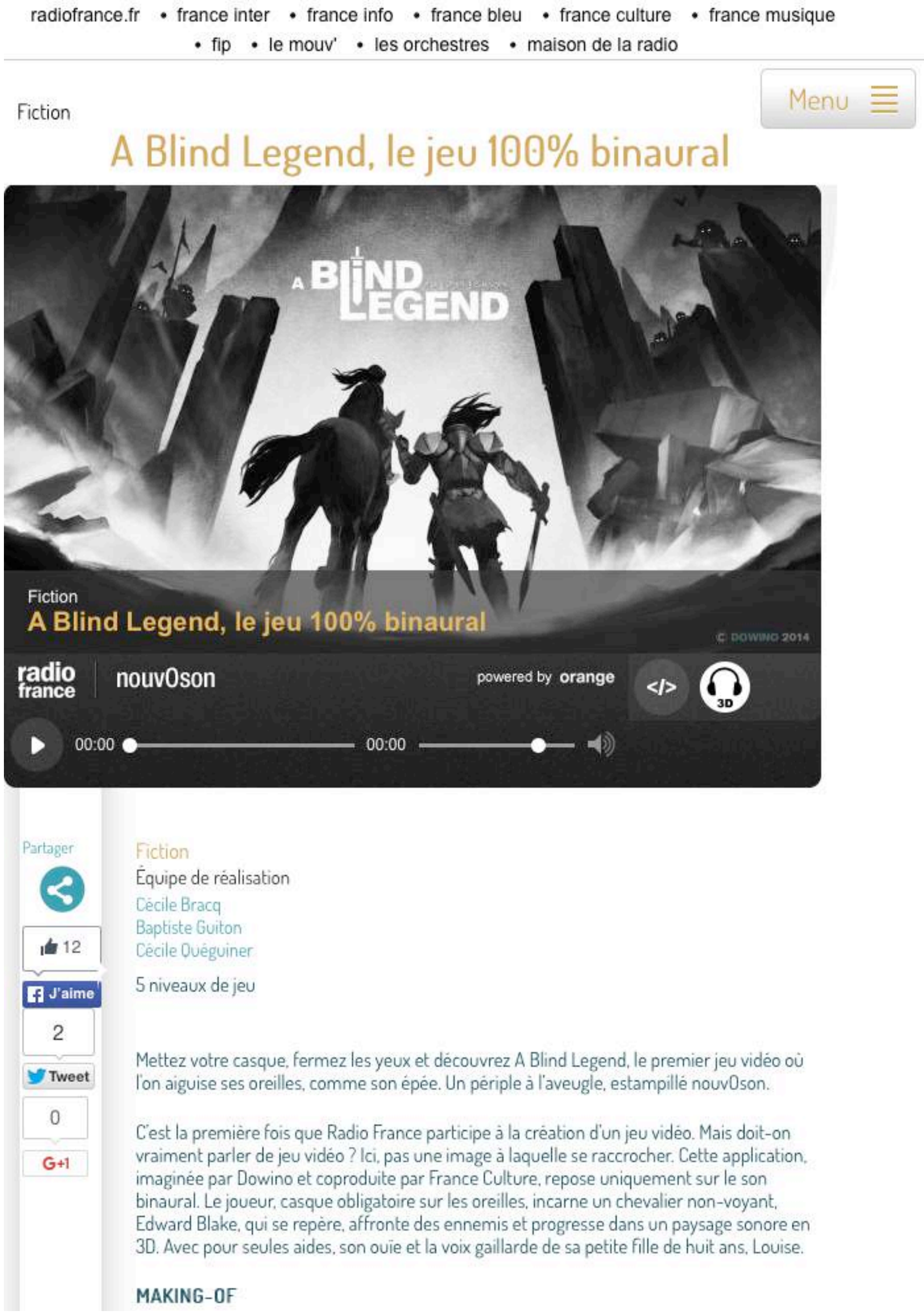
- how to assess the **performances of individualized HRTFs**?
- is it possible to find **non individual HRTFs** which fits me as best as my individual HRTFs?
  - and even better?
- what is the relative benefit of **individualization vs head-tracking**?
- how to achieve the **quality of dummy-head recordings** by binaural synthesis?
- how to improve sound **externalization**?

some « preliminary » answers by BiLi

- provision of a **« toolkit » of perceptual assessment**
  - to allow all the partners to make their own assessment
    - to ensure comparison by common experimental paradigms
- a list of **perceptual attributes**
  - elaborated with binaural sound
- **number of dimensions** of perceptual space
  - 4 dimensions are identified
  - but still under investigation!...
    - what about HRTF **preference**?
    - to be linked with acoustical/ signal **criteria** computed from HRTFs

some sound demos

- Radio France website: **nouvOson**
  - nouvoson.radiofrance.fr
- **radio drama** « deux femmes pour un fantôme »



thank you for your attention!