3D sound space perception: binaural is so natural...

Rozenn Nicol, Orange
Pascal Rueff, Agence du Verbe/ Feichter Electronic

Séminaire ETIS
16 octobre 2018
3D sound space?

- Sound are localized in 3D:
  - front/ rear
  - left/ right
  - up/ down
3D sound space perception

- How the auditory system is able to localize a sound?

  ⇒ psycho-acoustic **mecanism** of **sound localization**
    - 2 categories of mecanism
      - localization in the **horizontal** plane
      - localization in the **median** plane
  ⇒ based on the analysis of the **cues** of auditory localization
    - **interaural** cues
      - differences between the left and right ears
    - **monaural** cues
      - the information available at each ear is analysed separately and independently
Localization in the horizontal plane
localization in the horizontal plane

- "Duplex" theory (Lord Rayleigh)
  - localization is based on interaural differences
  - **phase/time** difference
    - Low Frequencies (LF)
    - Interaural Phase / Time Difference (IPD / ITD)
  - **level** difference
    - High Frequencies (HF)
    - Interaural Level Difference (ILD)
  - their interpretation by the auditory system is inherited
Interaural cues (1): ITD

ITD is caused by the difference of path length between the two ears.

ITD computed from HRTF (Head Related Transfer Function)
measured for one individual.
Interaural cues (2): ILD

**Diffraction** of the sound wave by the listener's head causes a **shadowing** effect for the controlateral ear.

ILD computed from HRTF (Head Related Transfer Function) **measured** for one individual.
Localization in the median plane
Localization in the median plane

- The **duplex theory fails** to explain localization in the **median** plane, and more generally in any **sagittal** plane
  - how the auditory system is able to spatially discriminate 2 sounds which share the same interaural differences?
Localization in the median plane (1): psychoacoustic evidence

- Psychoacoustic experiments suggest that the perception of elevation is controlled by the spectral content of the auditory event (i.e. Spectral Cues, SC)

Experiment of « directional bands » [Blauert 1969-70]

- A loudspeaker is located in the median plane and radiates a narrow-band noise signal whose center frequency is variable (the listener's head is immobilized).

- The perceived direction of the auditory event depends not on the loudspeaker direction but only on the frequency center of the signal.
Localization in the median plane (2): physical evidence of SC

- the **spectral information** at the **entrance** of the **listener’s ears** varies as a function of **elevation** in the median plane
  - various **spectral features** are observed
    - **notches**
    - **peaks**
  - the frequency of these features varies as a function of the sound's elevation
  - so they can be used by the auditory system as **cues** for the localization of elevation
Localization in the median plane (3): morphological origin of SC

- How are the SC generated?
  - mainly
    - Diffraction/ reflection/ resonances by the pinna
  - but also
    - diffraction by the head
    - reflection by the shoulder and the torso
    - however these cues are little used by the auditory system
Localization in the median plane (4): summary

- what to be kept in mind?
  - spectral encoding of the elevation of the sounds
    - the timbre of the sound source is modified by the acoustic diffraction caused by the listener's morphology (mainly the pinna)
    - these alterations of the timbre strongly depend on the sound's direction
  - the mecanism of SC interpretation is not inherited, but learned
    - the auditory system learns to associate a spectral pattern (i.e. a sound color) to a direction
  - the elevation localization is less accurate than the localization in the horizontal plane
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
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
Localization of distance

- Localization **cues** of distance
  - **Sound level**
    - the louder, the closer…
  - **Direct-to-reverberant ratio**
    - the lower, the farther…
  - **Timbre**
    - Low-Pass filtering by acoustic propagation
      - a sound with HF boost is perceived close (presence effect)
Accuracy of localization
Localisation blur in the horizontal plane

better accuracy for the frontal sources than for the lateral ones

Figure 2.2
Localization blur $\Delta \varphi_{min}$ and localization in the horizontal plane (after Preibisch-Effenberger 1966a and Haustein and Schirmer 1970; 600–900 subjects, white-noise pulses of 100 ms duration, approximately 70 phon, head immobilized).
Localisation blur in the median plane

the accuracy decreases as a function of elevation

Figure 2.5
Localization and localization blur in the median plane for continuous speech by a familiar person (after Damaske and Wagener 1969; 7 subjects, 65 phon, head immobilized). Note that the view is different from that in figure 2.2.
Localization in real life?
Theory vs real life

▪ The theory of sound localization previously presented concerns the case of localizing one single source in free field

▪ Reality is often more complex
  – a sound scene is generally composed of several sources
  – a room effect is present (image sources)

▪ Other mechanisms occur, e.g.:
  – *Precedence effect*
  – *Summing localization*
  – *Auditory scene analysis*
    – *Organization of the sound scene by auditory streams*
    – *Phenomena of fusion / integration or fission / segregation*
Virtual reality
3D audio

- The objective is to create the acoustic signals at the entrance of the listener's ears
  - to give him/her the **illusion** of **sound sources** which are **spatialized**
  - **virtual** or **phantom** sources
Binaural technology
Binaural technologies in a nutshell…

- 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 *(Interaural Time and Level Differences)*
  - SC *(Spectral Cues)*
- **"Full 3D" spatialization**
  - The virtual sound sources are localized in any direction over the 3D sphere
    - azimuth & elevation
Binaural recording

- recording by a **pair of binaural microphones**
  - either with a "natural" head
  - or with an artificial head or dummy head

- Brüel & Kjær HATS
- Head Acoustics HSMIII
- Neumann KU100
- Sound Professionals: SP-TFB-2 binaural microphones
- KEMAR
Binaural reproduction (1): headphones

- Headphones are the **recommended** way to listen to binaural signals
  - the signals are restituted at the closest to the location where they were recorded
  - **headphone calibration** is required
    - to compensate for the transfer function between the headphones and the listener's ears
    - But the problem is more **complex** than it appears at first sight
    - the HPTF depends both on the **individual** (i.e. the listener's morphology) and on the headphone **positioning**
- Keep in mind that it is possible to create **virtual** sources that **cannot be discriminated** from **real** sources
  - provided that the binaural signals are **properly reproduced**
- However **mechanical vibrations**, which are mainly perceived by the listener's body, are missing
Binaural reproduction (2): loudspeakers

- when listening to binaural signals over a stereophonic setup, **crosstalk** occurs
  - each loudspeaker is perceived not only by the ipsilateral ear, but also by the **contraletaral** one
Binaural synthesis

▪ an **alternative** to binaural **recording**
  – to create binaural content
▪ 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** (Head Related Transfer Function) which describes the acoustic propagation between the sound source and the listener's ears
Binaural synthesis: dynamic rendering

▪ How?
  – with a **head-tracker**
    – a system which **measures** any movement of the listener's head
      – acoustic (ultrasound)
      – inertial (accelerometer & gyroscop)
      – magnetic
      – optic (camera)
    – the tracking information is transmitted to the binaural synthesis engine
      – the binaural filters which control the location of the virtual sound source are **updated** in real time to compensate for the listener's movement
HRTF (1): definition

- HRTF = 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
- Depends on
  - the location of the sound source
    - fine sampling of the 3D sphere
    - in order to accurately measure all the spatial dependencies
  - the individual morphology of the listener
    - the measurement of individual HRTFs is require to provide individualized binaural synthesis to any listener
- Usually individual HRTFs are acquired by:
  - acoustic measurement
  - numeric modelling
    - BEM [Katz, 1998] [Kahana, 2000]
HRTF (2): acoustic measurement

Orange Lab equipment (Lannion)

- **Method**
  - Multiple Overlapped Delayed Sweeps
  - 31 loudspeakers arranged in 2 vertical arcs
  - Turning table
- **Both accurate and fast**
  - Around 20 min. per individual
  - 1580 directions (6° step/ azimuth & elevation)
- **Database:** around 110 individuals
HRTF (3): variation as a function of the azimuth angle, in the horizontal plane

HRTF spectrum magnitude (left ear)
HRTF (4): associated HRIR (variation as a function of the azimuth angle, in the horizontal plane)
HRTF (5): variation as a function of the elevation angle, in the median plane

monaural SC:
- frequency peaks
- and notches
HRTF (6): variation as a function of individual

8 individuals for the same direction
(Jean-Marie Pernaux database, Orange Labs)
Which HRTF: individual vs non-individual?

- **Preliminary remark:**
  - Binaural **recording** (dummy-heads) is not as affected by non-individualization as binaural **synthesis**

- **Individual** synthesis is better
  - Audible **artefacts** of non-individual HRTFs
    - poor externalization, front-back reversals
  - On the contrary, with **properly individualized** HRTFs
    - **virtual** sources are not **discriminated** from **real** ones

- **Which individualization?**
  - Acoustic **measurement**
  - Numerical **modelling** (BEM)
  - *Selection* (« ready-to-wear »)/ *adaptation* (« taylor-made ») of HRTF sets taken from **database**
  - Listener **recalibration**
Conclusion
Why working with binaural technologies?

- Binaural technology is an **easy** and **efficient** way to create **3D audio**
  - only **2 signals**
  - reproduction by **conventional headphones**
  - auditory **illusions** which are very **effective**
- More and more **tools** available
  - public HRTF databases
  - AES standard SOFA
  - Binaural engines in DAW
- Increasing use in **artistic** creations
  - radios: BBC, Radio France
  - Live drama
  - Music recording
Thank you for your attention!