

Pau Ballart Godoy

Localització de sons binaurals simulats amb ordinador



2010

Introducció

La capacitat que tenim de localitzar un so en l'espai, tant en l'azimuth (esquerre/dreta) com en l'elevació (amunt/avall) es deu al fet de tenir dues orelles. Una persona que perdi la capacitat auditiva d'una orella, també perd la capacitat de localitzar la procedència d'un so. El fet de poder localitzar un so en l'espai es deu a un seguit de propietats que té el nostre cap, les nostres orelles i el so. La primera és que les dues orelles estan separades entre elles per una distància. Aquesta distància fa que un so arribi amb una petita diferència d'intensitat, freqüència i temps a cada orella. També es nota l'impacte de l'ona sonora en la pell sensible de l'orella, que ens fa notar quina orella està més a prop del so. Un dels factors més importants són els HRTF(Head-Related Transfer Function), que, a grans trets, podem dir que són uns "filtres" que descriuen la modificació que pateix un so en xocar amb el tors, el cap i els pavellons auditius. També són decisius a l'hora d'identificar la procedència d'un so l'atenuació per la distància (un so que vingui de més lluny sona més fluix i amb més atenuació en els aguts) i els rebots de la sala, que ens ajuden, a més de crear una imatge de les dimensions i naturalesa de l'espai, a posicionar el so.

La binauralització és una tècnica d'enregistrament i reproducció de sons que permet identificar-ne la procedència. S'ha escollit aquest tema perquè és una tècnica innovadora, que representa un pas més enllà de l'estèreo, i que permet fer més realistes les pel·lícules, els videojocs i altres tipus d'entorns virtuals.

El principal inconvenient és que per escoltar els sons binaurals és imprescindible l'ús d'auriculars. Cal destacar que s'estan estudiant maneres de fer una sala amb una disposició d'altaveus especial que permeti sentir els sons binaurals i evitar així haver de posar-se auriculars.

A l'apartat 1 descriurem els aspectes metodològics de l'experiment. A l'apartat 2 introduïrem el concepte de "Funció de Transferència HRTF" i al 3 s'explicaran les diferents maneres d'obtenir un so binaural. A l'apartat 4 s'aprofundeix en el procés de simulació per ordinador i s'explica l'experiment; al 5 es mostren els resultats de l'experiment i al 6 s'analitzen tècnicament els sons binauralitzats. A l'apartat 7 s'exposen unes conclusions. L'apartat 8 mostra la bibliografia i el 9, la bibliografia web.

1. Metodologia

En aquest article es parlarà dels sons binaurals, què són i com es poden aconseguir. Es veuran breument les diferents maneres d'aconseguir aquests tipus de sons i s'aprofundirà més en la simulació per ordinador. S'explicarà què són els HRTF i quins s'han utilitzat per l'experiment. L'experiment consistirà a transformar dos sons diferents per tal que doni la

sensació que se sent en un costat o que passa de dreta a esquerra, i es faran dos tipus d'anàlisi. L'anàlisi tècnica consistirà a explicar el significat de les variacions de la forma d'ona del so en funció de com s'ha transformat. La segona anàlisi consistirà a enquestar 20 informants per veure si encerten la localització del so. Finalment s'extrauran conclusions.

2. HRTF's – Head-Related Transfer Functions

HRTF és una funció de transferència usada per sintetitzar sons binaurals des d'una font monoaural. Descriuen la resposta en freqüència a un impuls emès en una posició determinada relativa al subjecte. Un joc complet d'HRTFs consisteix en molts filtres que descriuen un mapa esfèric de les possibles fonts sonores. Cada HRTF conté informació sobre els canvis de freqüència, retard temporal i diferència d'intensitat entre ambdues orelles.

Si es filtra un so amb un HRTF i després s'escolta la filtració mitjançant auriculars, es percep la direcció de la font sonora. Els auriculars són necessaris ja que sense ells no és possible localitzar un so.

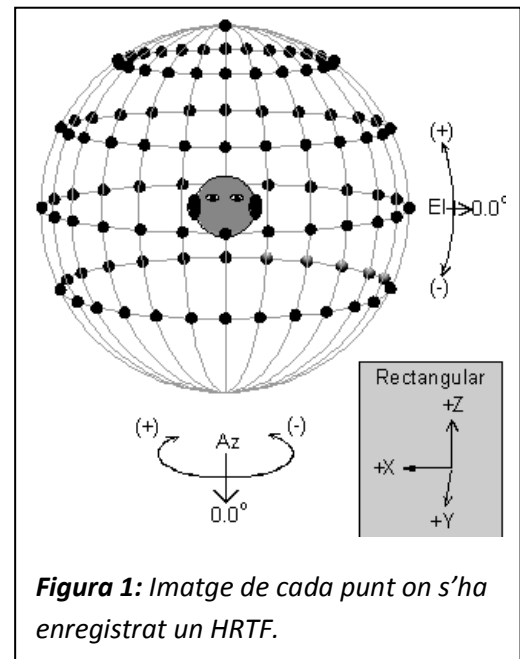


Figura 1: Imatge de cada punt on s'ha enregistrat un HRTF.

3. Enregistraments binaurals

Hi ha tres maneres d'aconseguir un so binaural, és a dir, un so del qual es noti la procedència. La primera és utilitzant un cap de plàstic, un maniquí, que incorpora un micròfon a cada orella i que emula les deformacions del so igual que passa amb un cap humà. És el millor mètode i el que dóna un resultat més acostat a la realitat. El principal inconvenient és que és car; per exemple, el maniquí "NEUMANN KU 100" costa 7.000€.

Una altra opció més econòmica és utilitzar uns micròfons que tenen forma d'auricular. Un voluntari es posa els auriculars a les orelles i el seu cap i les seves orelles substitueixen el cap de plàstic. El principal problema és que els micròfons són molt sensibles i el voluntari ha d'estar-se completament immòbil. Qualsevol moviment, per petit que sigui, pot quedar enregistrat i malmetre la gravació.

La tercera opció és fer la simulació per ordinador. Cal tenir un joc complet d'HRTFs i implementar-los en un so mitjançant un ordinador. Aquest és el mètode que s'utilitzarà en l'experiment. Es farà servir un joc d'HRTFs compilat a l'MIT, en el qual van utilitzar un maniquí anomenat KEMAR. Els investigadors de l'MIT van posar un micròfon a cada orella del maniquí i van registrar sons reproduïts des d'un altaveu situat a 1.4 metres del maniquí. Es van registrar un total de 710 sons utilitzant la freqüència de 44.1 KHz en una sala anti-eco. Les

mesures enregistrades són des de -40° d'elevació fins a $+90^\circ$ (just a sobre el cap). A cada grau d'elevació s'ha enregistrat una volta completa de 360° d'azimuth cada 5° aproximadament. Per exemple, elevació 0 azimuth 0 és directament davant el cap; elevació 0 azimuth 90 és a la dreta del cap, etc.

4. Procés de simulació per ordinador

S'han enregistrat dos sons diferents: una motocicleta passant per un carrer i una noia dient "Hola". A aquests sons se'ls aplicarà un azimuth concret mitjançant un programa creat amb MATLAB per un estudiant de la universitat politècnica de Hong Kong anomenat Ngan Cheuk Yin (2002) [4]. Un cop es tinguin els sons transformats es faran unes enquestes a 20 informants per comprovar si encerten la procedència del so que s'ha escollit. A partir dels resultats, es farà el tant per cent d'encerts de cada una de les 8 transformacions. També es faran unes anàlisis tècniques de la forma d'ona i de l'espectre de cada so i les seves transformacions.

El so de la moto es transformarà 4 vegades:

- La moto passa de dreta a esquerra
- La moto passa d'esquerra a dreta
- La moto dóna la volta sortint de davant i en el sentit de les agulles del rellotge
- La moto dóna la volta sortint de davant i en el sentit contrari de les agulles del rellotge

El so "Hola" també es transformarà 4 vegades:

- El so se sent a la dreta
- El so se sent a l'esquerra
- El so se sent al davant
- El so se sent al darrere

Cada informant escoltarà 8 sons aleatoris que poden estar repetits. Cada so l'escoltarà dues vegades i haurà de marcar a quin dibuix es correspon utilitzant la següent plantilla.









	So Moto				So Hola			
								
T1								
T2								
T3								
T4								
T5								
T6								
T7								
T8								

Figura 2: Plantilla de les enquestes.

La transformació dels sons es fa amb el programa MATLAB, seguint una sèrie de passos utilitzant el programa creat per Ngan Cheuk Yin [1]. S'ha d'escollir el so original, escollir el fitxer de sortida, aplicar-hi un "horitzontal surround" o "vertical surround" i clicar a "process". El procés de transformació es fa amb un algorisme. Si tens el so $x(t)$ i la funció de transferència que representem com $h(t)$, el so transformat $y(t)$ es calcula amb la convolució: $y(t)=h(t)*x(t)$.

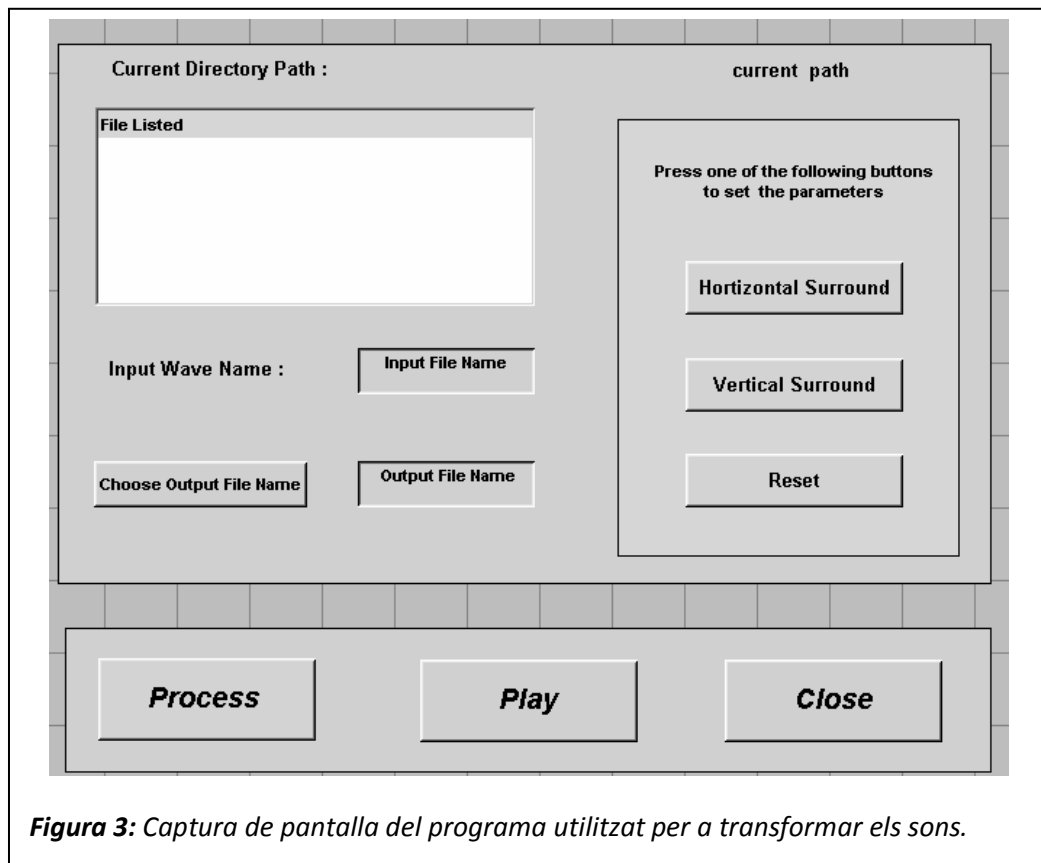


Figura 3: Captura de pantalla del programa utilitzat per a transformar els sons.

5. Resultats de les enquestes

Les enquestes s'han fet a 20 informants d'edats i sexes diferents, que han hagut d'escoltar amb auriculars que es posen dins el canal auditiu els 8 sons transformats i identificar-los a la plantilla de la figura 2.

Els gràfics següents mostren el tant per cent d'encerts dels sons escoltats:

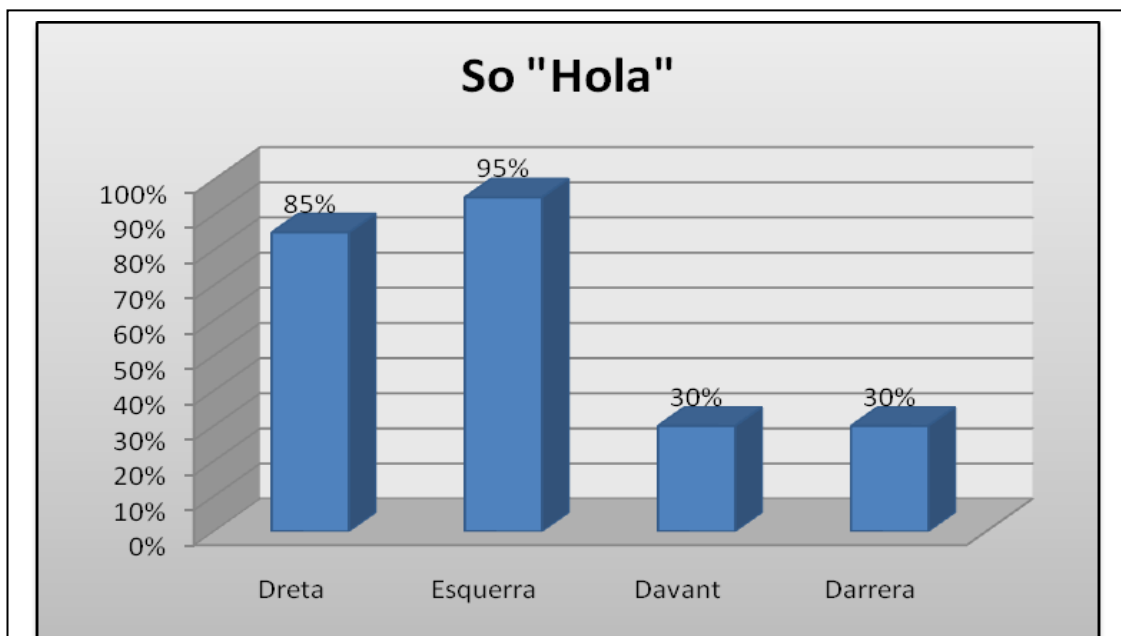


Figura 4: Gràfic del tant per cent d'encerts en la procedència del so "Hola".

Els resultats de les enquestes demostren que la localització d'un so curt i precís com el de la paraula "hola" es localitza fàcilment si prové de la dreta o de l'esquerra. En canvi, costa molt saber si prové de davant o de darrere. Això és a causa que els HRTF emprats per la transformació de davant i la de darrera són molt semblants. Amb una mostra tan petita no es poden extreure conclusions rellevants, però més endavant s'intentarà explicar per què no es diferencien a davant i a darrere mentre que sí ho fan a dreta i esquerra.

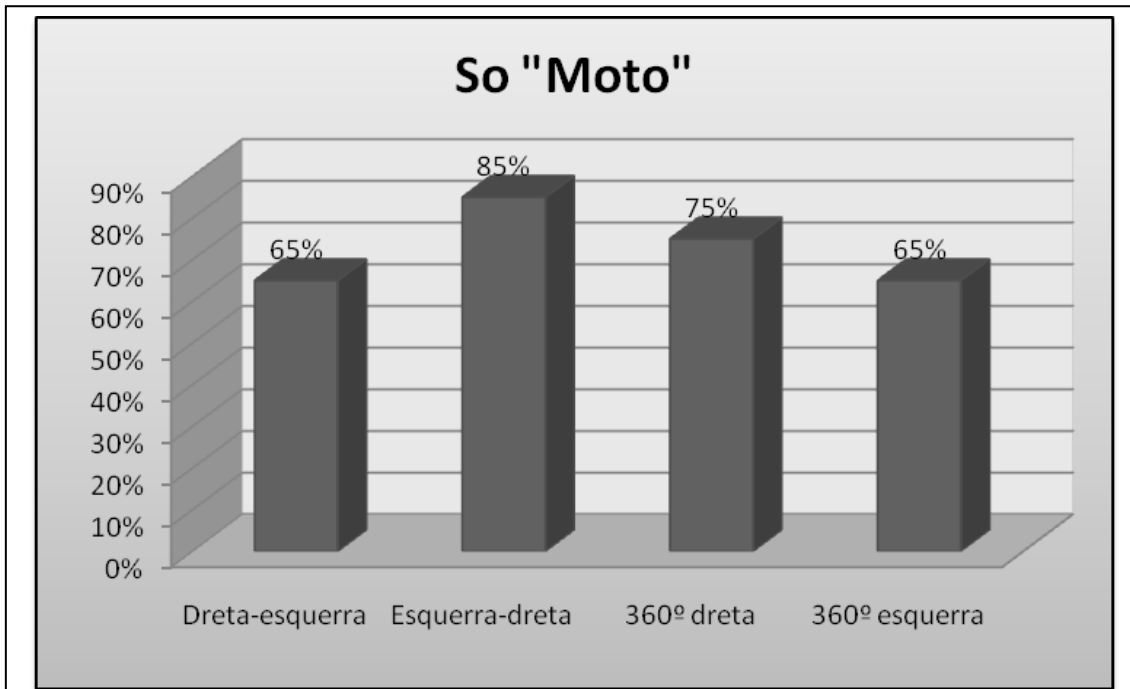


Figura 5: Gràfic del tant per cent d'encerts en la direcció del so "Moto".

Pel que fa al so de la moto, es veu com es reconeix bé la procedència i la trajectòria del so. L'elevat nombre d'encerts es deu probablement al fet que és un so brillant i per tant amb més contingut en el rang de freqüències on les HRTF tenen més influència (8-13kHz). Potser alguns informants han tingut problemes a l'hora d'interpretar els símbols, però en general s'ha encertat.

Un fet que cal tenir en compte és que les enquestes s'han fet a tot tipus de persones, de diverses edats i que alguns potser pateixen alguna deficiència auditiva. També s'ha de tenir en compte que els HRTFs emprats són estàndards i que en realitat cada subjecte, per les seves proporcions físiques, mostra unes HRTFs úniques i són a les que està habituat per posicionar el so, cosa que afegeix una dificultat extra quan s'intenta localitzar un so binauralitzat per software.

6. Anàlisi tècnica dels sons

Espectres dels HRTF usats

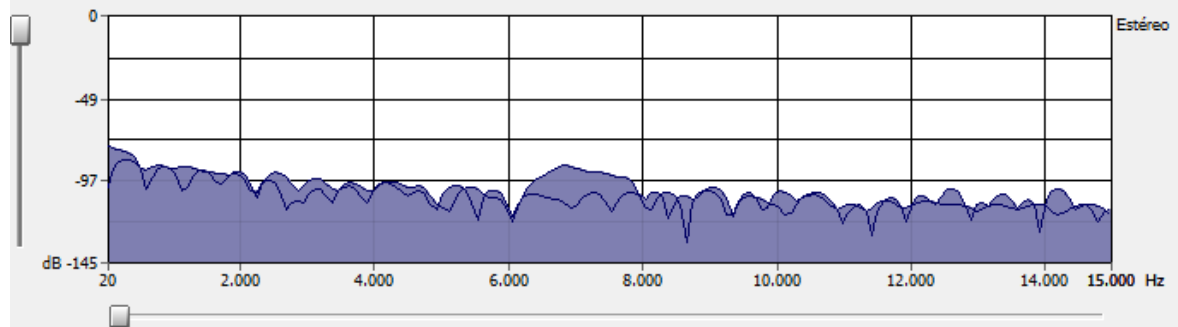


Figura 6: Espectre de l'HRTF elevació 0 i azimuth 90 (dreta)

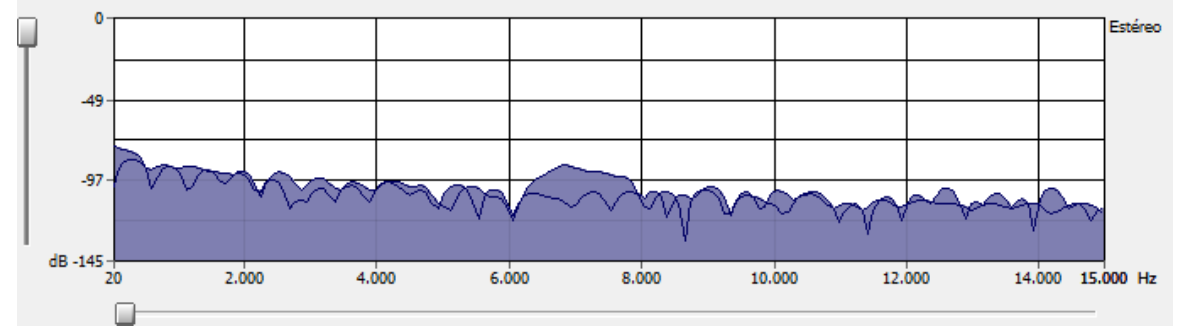


Figura 7: Espectre de l'HRTF elevació 0 i azimuth 270 (esquerra)

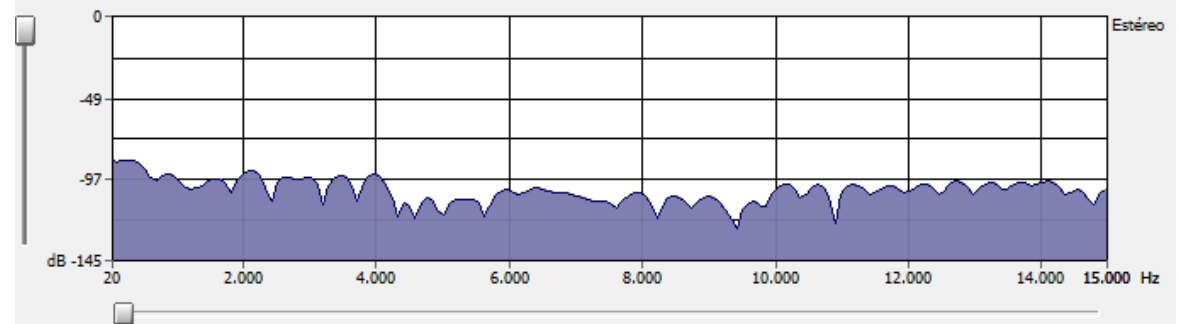


Figura 8: Espectre de l'HRTF elevació 0 i azimuth 0 (davant)

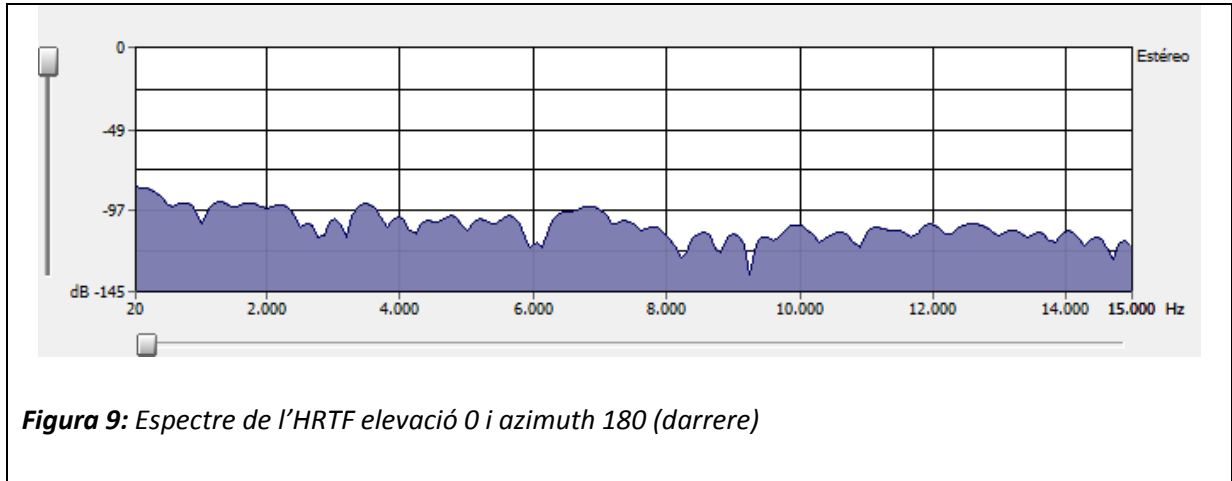


Figura 9: Espectre de l'HRTF elevació 0 i azimuth 180 (darrera)

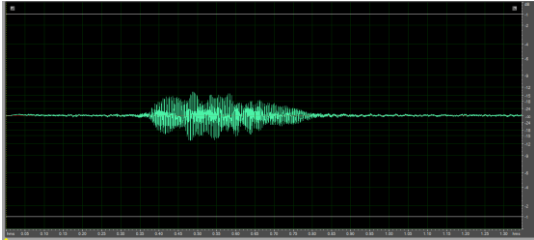
Explicació

Els espectres ens mostren les freqüències a l'eix horitzontal i l'amplitud de cada freqüència a l'eix vertical. Amb aquests espectres podem saber per què un so costa de diferenciar a davant i a darrera mentre que és fàcil saber si ve de la dreta o de l'esquerra. En la figura 6 i 7 podem observar entre les freqüències 6.000 i 8.000 Hz un pic en un dels dos canals. Aquest pic correspon al filtre ocasionat pel cap a l'ona sonora. Aquest pic és el que ens ajuda a diferenciar si el so prové de la dreta o de l'esquerra, ja que en la figura 7 l'espectre aparentment és el mateix que el de la figura 6, però els canals estan invertits. En la figura 6 el pic correspon al canal dret, mentre que en la figura 7 correspon al canal esquerre.

Per contra, en la figura 8 i 9, que corresponen als HRTF usats per filtrar el so a davant i a darrera, només es veu una línia, ja que els espectres dels dos canals són iguals i les dues línies se sobreposen.

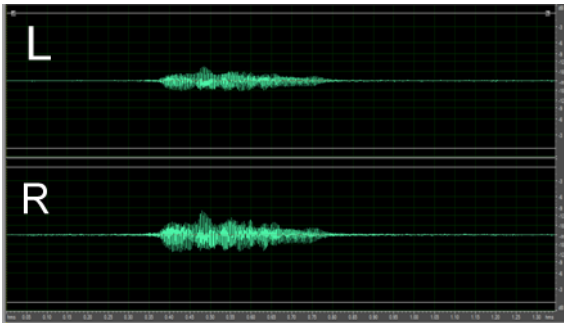
Els espectres de les figures 8 i 9 són lleugerament diferents però no hi ha cap pic que ens permeti diferenciar bé el so, per això és fàcil de confondre'ls. A més, no podem servir-nos de la diferència de temps interaural per localitzar la procedència del so, ja que arriba al mateix temps a les dues orelles.

Forma d'ona dels sons "Hola"

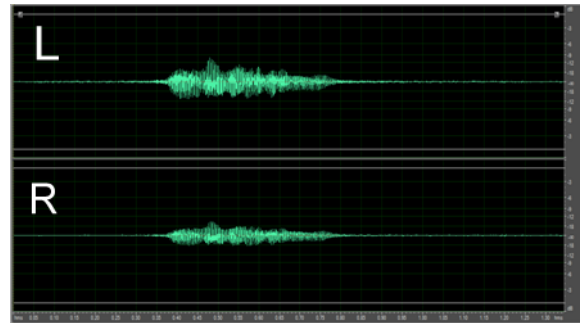


So original mono

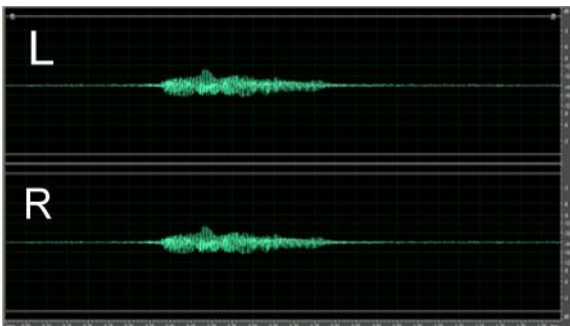
Les imatges mostren la forma d'ona dels sons de cada canal (L=esquerre (left) i R=dreta (right)). Observem com en les transformacions a la dreta o a l'esquerra l'amplitud d'ona és més gran al canal del costat que ha estat transformat. En canvi, en els sons transformats a davant i a darrera, no s'aprecia pràcticament cap canvi.



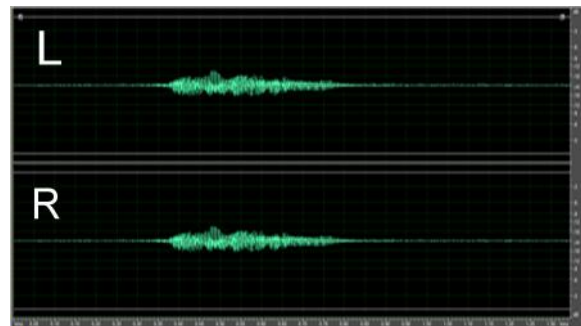
So transformat procedent de la dreta



So transformat procedent de l'esquerra

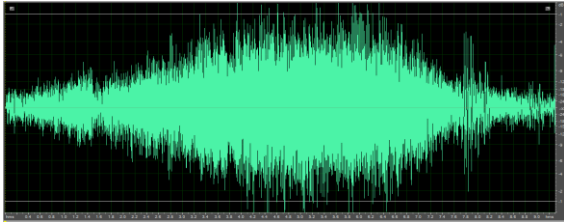


So transformat procedent de davant



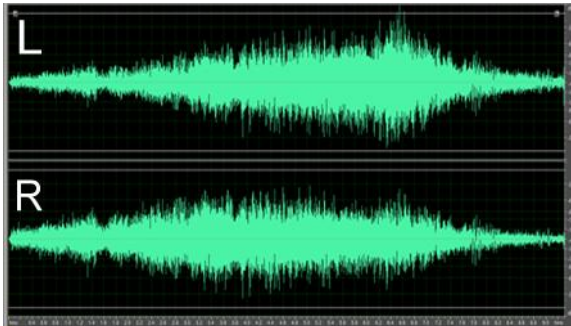
So transformat procedent de darrere

Forma d'ona dels sons "Moto"

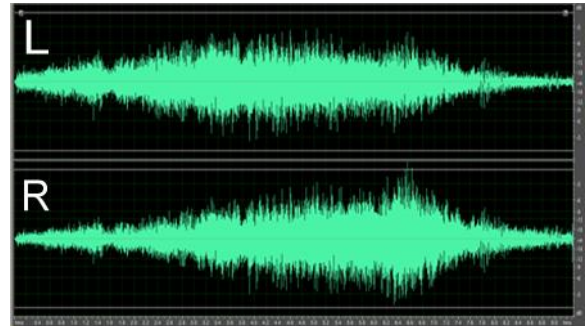


So original mono

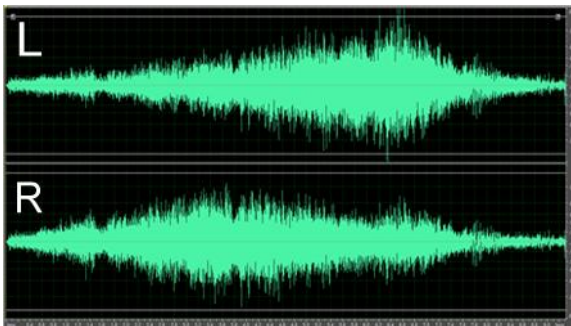
Les imatges mostren la forma d'ona dels sons de cada canal (L=esquerre (left) i R=dreta (right)). Es veu com a l'inici del so hi ha més intensitat en un canal i al final del so la intensitat augmenta en l'altre canal.



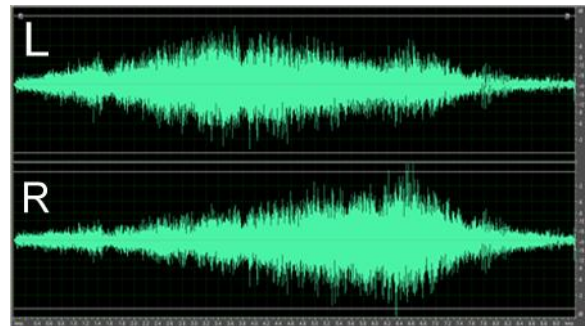
So transformat que va de dreta a esquerra



So transformat que va d'esquerra a dreta



So transformat que dóna una volta sortint de davant i en sentit horari



So transformat que dóna una volta sortint de davant i en sentit anti-horari

7. Conclusions

En aquest article hem vist com es fa per obtenir un so binaural. Hem posat a prova el mètode més econòmic i més artificial: la simulació per ordinador. Hem transformat uns sons amb un programa i hem enquestat 20 informants per comprovar la fiabilitat del programa. A més hem analitzat tècnicament la forma d'ona dels sons obtinguts i l'hem relacionat amb els resultats de les enquestes.

Podem concloure que la binauralització per software és prou efectiva. En la transformació de sons a la dreta o l'esquerra els usuaris no tenen problemes per localitzar la procedència simulada de la font. En el cas de davant i darrere no podem dir el mateix, ja que queda clar que no se'n diferencia la procedència. En els sons amb moviment podem dir que s'identifiquen força bé i que la binauralització és correcta. Per tant, només falla en la diferenciació de davant i darrere. Això s'explica per la semblança dels HRTFs aplicats.

A partir del que s'ha fet, es podria millorar el treball fent proves més exhaustives amb sons més variats i de diferents angles i freqüències. També seria millor si es fessin més enquestes i amb més usuaris per eliminar el marge d'error.

Seria molt interessant poder enregistrar un so amb el maniquí KEMAR i fer la simulació per ordinador del mateix so i veure quines diferències hi ha entre els sons binaurals obtinguts d'una manera i de l'altra, com per exemple la qualitat, el realisme, l'espectre que mostren, etc.

8. Bibliografia

- [1] DeLiang Wang and Guy J. Brown, "Binaural Sound Localization", capítol 5, dins *Computational Auditory Scene Analysis*. John Wiley & Sons, Inc. 2005 ISBN 0-471-45435-4
- [2] Inoue, Nishino, Itou, Takeda, "HRTF Modeling using Physical Features", dins *Forum Acusticum 2005 Budapest*.
- [3] Fakheredine Keyrouz and Klaus Diepold, "An Enhanced Binaural 3D Sound Localization Algorithm", dins *2006 IEEE International Symposium on Signal Processing and Information Technology*.
- [4] Ngan Cheuk Yin, "3-D Sound Based on Head-Related Transfer Functions", de la Universitat Politècnica de Hong Kong

9. Bibliografia web

- Joc complet de mesures d'HRTFs del maniquí KEMAR:
<http://sound.media.mit.edu/resources/KEMAR.html>
- Informació i història dels sons binaurals:
http://deaparatos.com/sonido_binaural_y_holofonia
- Informació dels enregistraments binaurals:
http://en.wikipedia.org/wiki/Binaural_recording
- Informació dels HRTF:
http://en.wikipedia.org/wiki/Head-related_transfer_function
- Informació dels enregistraments amb maniquí:
http://en.wikipedia.org/wiki/Dummy_head_recording
- Desenvolupament d'un sistema de captació per a sistemes 5.1:
<http://www.esmuc.net/sonologia/files/TFC-GerardFont.pdf>
- So en 3D basat en HRTF:
<http://www.eie.polyu.edu.hk/~mwmak/fyp/3D-sound/index.htm>
- Informació sobre fenòmens acústics:
http://www.lpi.tel.uva.es/~nacho/docencia/ing_ond_1/trabajos_03_04/Csound/41.htm#_ftn3
- Guia per a construir un cap de maniquí amb micròfons (com el KEMAR):
<http://digdagga.com/dummy/index.html>

Agraïments

Al meu tutor, Josep Maria Comajuncosas, per la seva paciència i dedicació, suport i ajut en tot moment.

A totes les persones que han tingut l'amabilitat de fer les enquestes.

A la meva mare, per haver-me fet costat i aconsellat des del principi fins al final.