

Modèles de perception et de production

Majeure SIR – Systèmes Interactifs et Robotiques

Stéphane Rossignol – stephane.rossignol@supelec.fr

2015 – 2016

Les slides sont disponibles ici

[http ://www.metz.supelec.fr/metz/personnel/
rossignol/slides_tout_modeles.pdf](http://www.metz.supelec.fr/metz/personnel/rossignol/slides_tout_modeles.pdf)

Version susceptible d'évolution

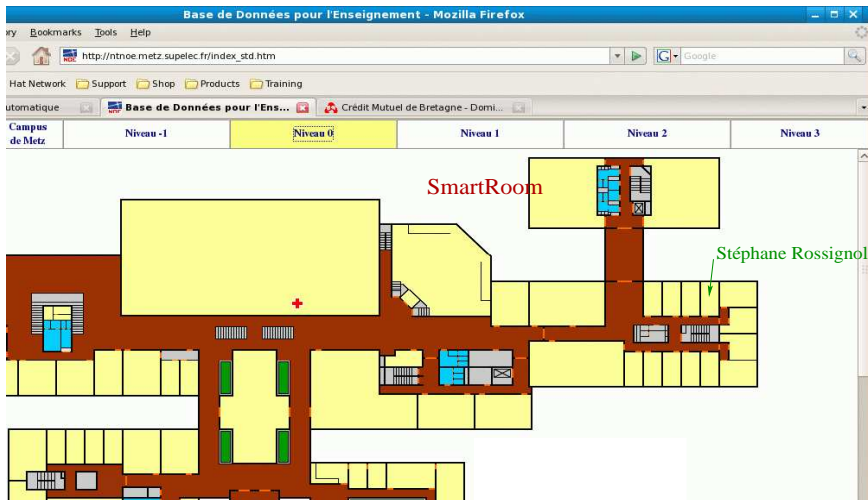
Mes recherches sont liées à ce cours

- ▶ Études à l'Université : DUT EEII, licence et maîtrise d'électronique, DEA signal télécom image radar (option radar), DEA instrumentation en sciences de l'univers
- ▶ Service militaire en tant que scientifique du contingent – détection de masses ferromagnétiques de 300 tonnes se déplaçant sous l'eau
- ▶ Thèse finie en juillet 2000 ; thèse sur la segmentation et l'indexation des signaux sonores musicaux, faite à l'IRCAM et ici, en partenariat avec France Télécom-Orange
- ▶ Modification de la prosodie d'un acteur (Depardieu parlant anglais dans Vatel de Roland Joffé)
- ▶ Postdoc (2000-2001) : modélisation du vibrato dans une université néerlandaise (⇒ aide à la composition, MPEG-7)
- ▶ Postdoc (2001-2005) : mise en place de systèmes de dialogue dans une université néerlandaise (projet européen COMIC)

Mes recherches sont liées à ce cours

- ▶ Postdoc (2006) : mise en place d'un système de reconnaissance de la parole (pour France Télécom-Orange)
- ▶ Postdoc (2007) : caractérisation de défauts de haut-parleurs sur la ligne de production (équipementier automobile)
- ▶ Enseignant-chercheur ici depuis 2008
 - ▶ recherches propres : traitement de la musique (et de la parole)
 - ▶ mise en place de systèmes de dialogue (projets européens CLASSIC, ALLEGRO, HILAIRE)
 - ▶ holophonie
 - ▶ cours concernant principalement le traitement du signal (SIG1, SIG2, RASS, analyse spectrale, traitement de la parole)

Où suis-je ?



Modèles de perception et de production – introduction

Stéphane Rossignol

2015 – 2016

Qu'est-ce que la psychophysique ?

La **psychophysique** est l'étude des relations qui existent entre :

- le *stimulus*, qui est un phénomène physique (par exemple, un poids) provoquant une sensation
- et cette *sensation* elle-même (la sensation de poids)

Par exemple, la **psychoacoustique** (ou acoustique physique) est l'étude des relations entre le phénomène physique que sont les vibrations de l'air et la PERCEPTION que nous en avons ; l'**optique (psycho-)physiologique** est l'étude des relations entre le phénomène physique qu'est la lumière et la PERCEPTION que nous en avons

Production/Reproduction

- ▶ Les modèles de PRODUCTION cherchent à caractériser les sources qu'on trouve dans la nature
 - ▶ modèle physique d'un instrument de musique
 - ▶ modèle physique du système phonatoire
 - ▶ etc.
- ▶ On cherche aussi à reproduire au mieux, artificiellement (avec un ensemble de haut-parleurs/de projecteurs, par exemple), ces sources

Un petit topo sur les sens 1/2

Ch. Sherrington (1857 – 1952) les a classés ainsi :

1. Les sens *intéroceptifs* (impressions venant des surfaces intérieures de l'organisme : tube digestif)
 - sensations qui viennent de l'estomac, de l'intestin
 - sensations viscérales
 - **cœnesthésie** : vagues sensations d'aise ou de malaise
2. Les sens *proprioceptifs* (activité propre de l'organisme, attitudes, mouvements)
 - sens **statique** ou **labyrinthique** (oreille interne) : verticalité du corps, mouvements de rotation et de translation (équilibration)
 - sens **kinésique** ou **kinesthésique** : mouvements proprement dits : déplacements des membres et du corps dans l'espace (permet la reconnaissance des objets par la palpation – toucher actif – ou stéréognosie : lisse, rugueux, grasseux...)

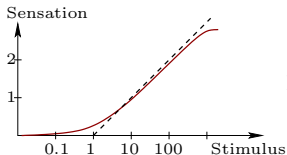
Un petit topo sur les sens 2/2

3. Les sens *extéroceptifs* (impressions venant du monde extérieur)
 - sens impressionnables par contact direct
 - ▶ le **toucher** (décomposé en trois sortes : 1. tactile : contacts, pressions, poids – toucher passif –; 2. thermique; 3. algique : douleur)
 - ▶ sens chimiques : **goût** ; et **odorat** : lié au goût (quand on goûte quelque chose, l'odorat participe : on respire/souffle par le nez)
 - sens impressionnables à distance : ce sont les plus importants
 - ▶ la **vue**
 - ▶ l'**ouïe**

ils sont complémentaires (on ne voit pas derrière soi ; la nuit)

Les premières << lois >> (18ème et 19ème siècles)

1. loi du **seuil** : il existe pour chaque sens une intensité *minima* du stimulus, appelée *intensité liminaire*, au-dessous de laquelle il n'y a pas de sensation (attention : pour l'œil, accommodation)
2. loi du **seuil différentiel** :
 - Forme a. Il existe un *rapport* constant entre l'intensité du stimulus initial et la variation minima qu'il faut lui faire subir pour que la différence soit sentie (Weber)
 - Forme b. Pour que la sensation subisse des accroissements en *progression arithmétique* (0, 1, 2...), il faut faire varier le stimulus en *progression géométrique* (1, 10, 100...); le *rapport* constant est le seuil liminaire. C'est encore la loi logarithmique, ou loi de Fechner



Loi de Fechner et loi pratique

Un peu d'histoire

- En 1860, Fechner crée toute une science : la psychophysique, basée sur les lois précédentes
 - ▶ le début des sciences cognitives
- Ces lois avaient été formulées par Weber en 1846... et par Bouguer en 1760 pour la vision...
- Une remarque. Ces lois :
 1. Ne sont une bonne approximation que pour les stimulus moyens
 2. Ne peuvent pas être considérées comme une méthode de mesure de la sensation *en tant que fait subjectif de la conscience* (en fait, on leur reproche de ne pas tenir compte du contexte : autres sens, organisation, etc.)
⇒ Fechner était trop optimiste, ou trop enthousiaste)

Comment teste-t-on ? Principes de base.

- **Expérimentation**

- ▶ *principe* : on fait écouter au sujet 2 sons (par exemple) en ne modifiant qu'**une** grandeur d'excitation et on demande au sujet de comparer ses sensations
- ▶ *exemple* : niveau double, niveau moitié; hauteur double, hauteur moitié

- **Méthodes de mesure**

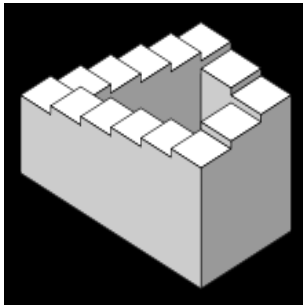
- ▶ *par ajustement* : le sujet ajuste lui-même la valeur d'excitation
- ▶ *ajustement par va-et-vient* : le sujet choisit le sens de variation de la valeur d'excitation, qui varie sans arrêt (audiomètre de von Békésy)
- ▶ *stimulus constant* : le sujet n'agit pas, il compare et répond d'une façon binaire (plus ou moins)

Psychoacoustique – On peut tromper l'oreille !

- ▶ Un son curieux :
 - ▶ son fait par moi (code matlab disponible) : *trompeur3.wav*
 - ▶ trouvé sur le net 1 : *DInfinie.wav*
 - ▶ trouvé sur le net 2 : *Strange_Tones.wav*

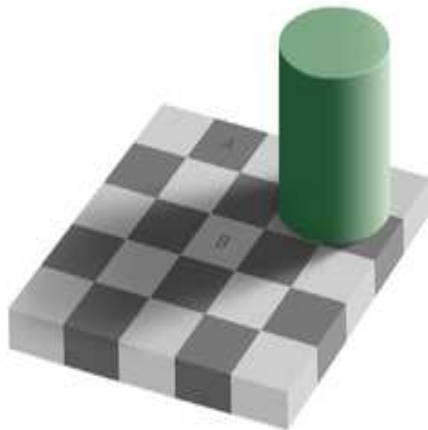
Psychoacoustique – On peut tromper l'oreille et l'œil !

- ▶ on les appelle les sons de Risset (ou de Shepard)
- ▶ il y a confusion entre la hauteur tonale et la hauteur spectrale
- ▶ équivalent en vision : l'escalier de Penrose

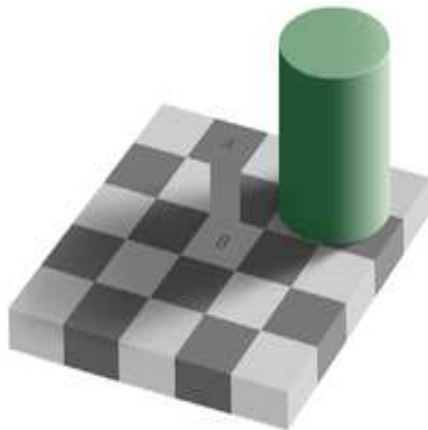


- ▶ voir Eisher (http://artgill.free.fr/MC_Eisher.htm)

Psychophysique – On peut tromper l'œil !



Psychophysique – On peut tromper l'œil !



Psychophysique – domaines d'application 1/2

- ▶ sociologie : musique d'ambiance, de publicité, d'annonce...
- ▶ urbanisme, ergonomie : protection contre la fatigue visuelle/auditive au travail, dans sa voiture, dans la rue, chez soi, sur son ordinateur...
- ▶ psychologie, acoustique, marketing : perception et classification des bruits de voitures, moteurs, portières, klaxons...
- ▶ militaire, ergonomie : aide à la navigation (pour les voitures, les avions, les chars...), message d'information ou d'alarme, reconnaissance des bruits marins, système experts de reconnaissance, aide à la décision, aide à la spatialisation (pilote de chasse)...
- ▶ pêche : reconnaissance des bruits marins...
- ▶ médecine (robotique ; sens compensés ; etc.)

Psychophysique – domaines d'application 2/2

- ▶ art : cinéma, musique (aide à la composition...)
- ▶ traitement du signal : JPEG, codage psychoacoustique : MiniDisque de Sony, DCC (digital compact cassette), DAB (digital audio broadcasting), MPEG, téléphone cellulaire, INMARSAT...
- ▶ physiologie et neurologie : traitement des troubles visuels/auditifs...
- ▶ restitution visuelle, sonore (spatialisation, matériel HIFI...)
- ▶ multimédia : spatialisation (virtualisation de l'espace) des sources visuelles/sonores, richesse de l'environnement visuel/sonore des jeux et des films, synthèse...
- ▶ segmentation : étiquetage, séparation de sources, transcription automatique de partitions...
- ▶ reconnaissance des images, de la parole (avec elle-même 3 millions d'applications, comme par exemple les systèmes de dialogue...)
- ▶ synthèse d'images, de la parole (systèmes de dialogue...)
- ▶ etc.

Bibliographie 1/2

- ▶ **R. Rigal, R. Paoletti, M. Portmann**, *Motricité – approche psychophysiological*, 1974, Presses de l'université du Québec (330 pages)
- ▶ **Delorme et Flückiger**, *Perception et réalité – Une introduction à la psychologie des perceptions*, de Boeck (517 pages)
- ▶ **E. Zwicker et R. Feldtkeller**, *Psychoacoustique*, 1981, Masson (⇒ il est à la bibliothèque du campus!)
- ▶ **R. Feynman**, *Mécanique 2*, 1998 (version française), Dunod (⇒ il est à la bibliothèque du campus!)

Bibliographie 2/2

- ▶ **L. Landau et E. Lifchitz**, *Physique théorique en 10 tomes – Tome 6 – Mécanique des fluides*, 1989, Librairie du globe/MIR
- ▶ **N. H. Fletcher et T. D. Rossing**, *The Physics of Musical Instruments*, 1991, Springer-Verlag
- ▶ **A. Cuvillier**, *Cours de philosophie*; tome 1 ; pages 84 – 85, 541 ; 1954 ; Armand Colin
- ▶ **Émile Bréhier**, *Histoire de la philosophie*; tome 3 ; pages 862 – 864 ; 1964 ; Quadrige – Presses Universitaires de France

Modèles de perception – la lumière

Stéphane Rossignol

2015 – 2016

Le stimulus visuel : la lumière

Les capteurs

Les sensations

Conclusion à l'optique physiologique

Modèles de perception

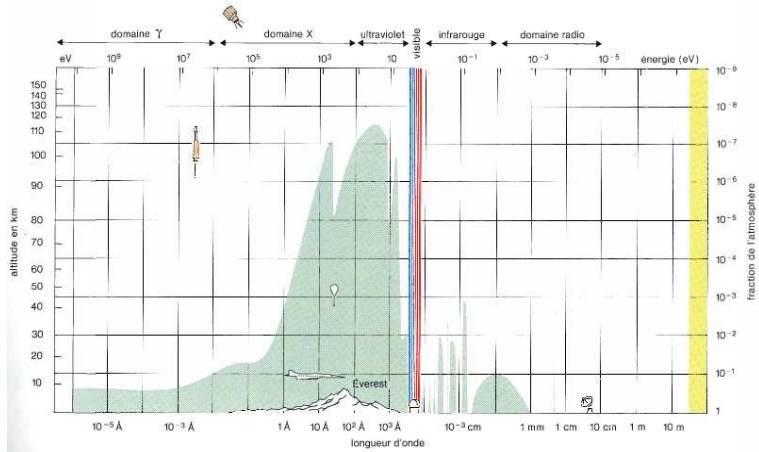
Plan

- ▶ Le stimulus : la lumière

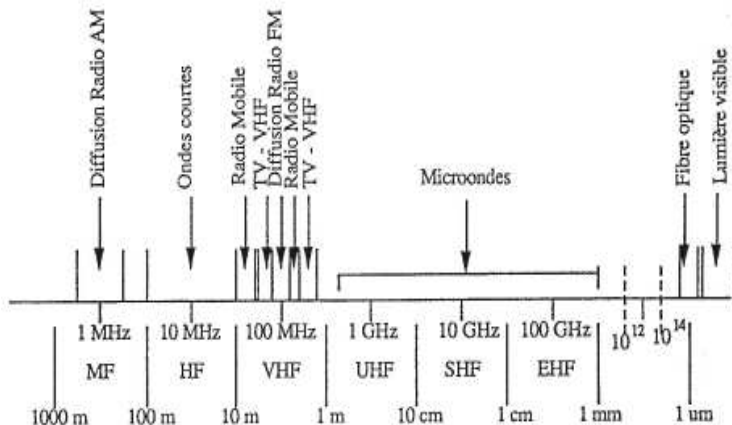
Qu'est-ce que la lumière ?

- ▶ Dualité onde-corpuscule
 - ▶ Théorie ondulatoire
 - ▶ onde électromagnétique
 - ▶ vitesse dans le vide : $c = 300000 \text{ kms/s}$
 - ▶ fonction d'onde : $\frac{\partial^2 E_y}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 E_y}{\partial t^2}$ (slide 128)
 - ▶ Théorie corpusculaire
 - ▶ particule : le photon
 - ▶ énergie : $h\nu$, avec $h = 6.626 \cdot 10^{-34} \text{ js}$
 - ▶ h est la constante de Planck et ν la fréquence
- ▶ Pas de support physique

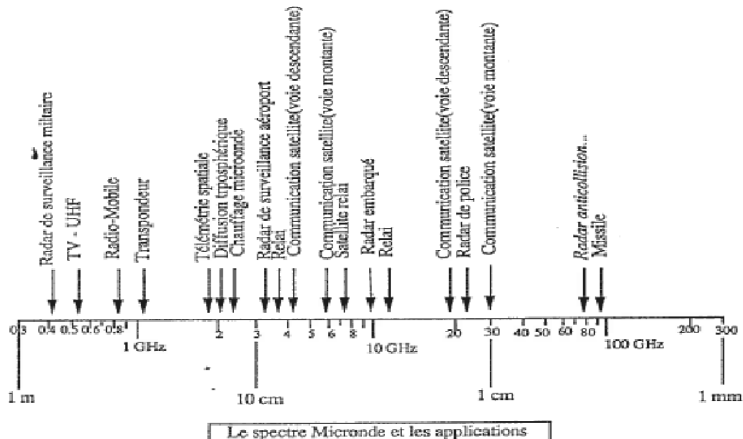
La lumière – spectre électromagnétique complet



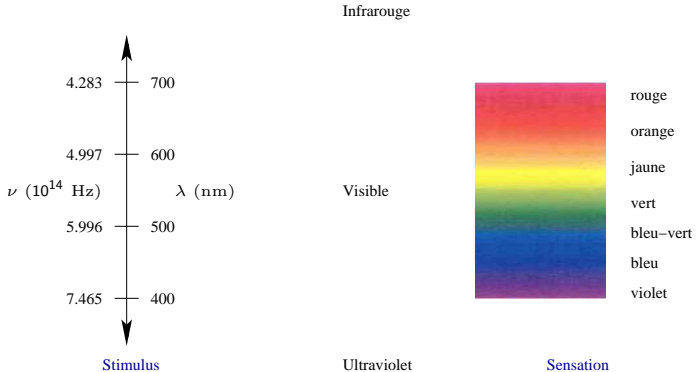
La lumière – spectre ondes radio



La lumière – spectre microondes



La lumière – spectre visible



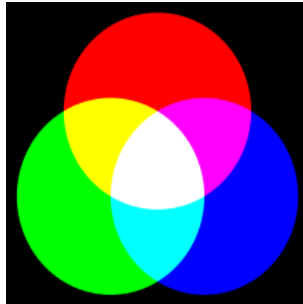
Synthèse additive

- ▶ on combine la lumière de plusieurs sources colorées pour obtenir une nouvelle couleur
- ▶ en général, on utilise 3 sources colorées : le rouge, le vert et le bleu (couleurs primaires)
- ▶ la loi de synthèse trichrome dit qu'il est possible de reproduire n'importe quelle sensation colorée en mélangeant les trois couleurs primaires
- ▶ ceci suggère qu'il existe 3 sortes de cellules photoréceptrices chacune sensible à une couleur primaire

Synthèse additive

- ▶ la synthèse additive est rendue possible par la façon dont l'œil humain fonctionne ; ce n'est pas une propriété intrinsèque de la lumière
- ▶ il y a en effet une grande différence physique entre une lumière jaune de longueur d'onde 580 nm, et un mélange de lumière rouge et verte ; mais les 2 phénomènes stimulent l'œil de façon similaire (couleurs métamères)
- ▶ utilisation de la synthèse additive : écrans cathodiques, écrans LCD, vidéoprojecteurs
- ▶ note : la synthèse additive ne nécessite pas d'éclairage supplémentaire

Synthèse additive



- ▶ Couleurs primaires : rouge, vert, bleu
- ▶ Couleurs secondaires : cyan, magenta, jaune

Synthèse soustractive

- ▶ on combine l'effet d'absorption des couleurs par des objets (filtres) pour obtenir une nouvelle couleur
- ▶ exemple 1 : on superpose sur une table blanche un filtre jaune et un filtre cyan, et on obtient du vert
- ▶ exemple 2 : on superpose sur une table blanche un filtre jaune, un filtre cyan et un filtre magenta, et on obtient du noir (aucune lumière ne passe plus)
- ▶ utilisation de la synthèse soustractive : peinture, imprimerie (trichromie ; la quadrichromie utilise en plus l'encre noire, pour avoir un vrai noir bien profond), impression papier pour la photographie en couleur
- ▶ note : la couleur obtenue est plus sombre que celles de départ

Synthèse soustractive



- ▶ Couleurs primaires : cyan, magenta, jaune
- ▶ Couleurs secondaires : rouge, vert, bleu

Synthèse soustractive

- ▶ l'effet obtenu dépend de l'éclairant (la table blanche des exemples ci-dessus)
- ▶ lumière blanche
 - ▶ objet réfléchissant toute la lumière : apparaît blanc
 - ▶ objet absorbant le rouge : apparaît cyan (blanc-rouge de la synthèse additive)
- ▶ lumière jaune
 - ▶ objet absorbant le rouge : apparaît vert (jaune-rouge de la synthèse additive)
- ▶ lumière magenta
 - ▶ objet absorbant le rouge : apparaît bleu (magenta-rouge de la synthèse additive)

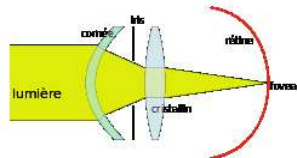
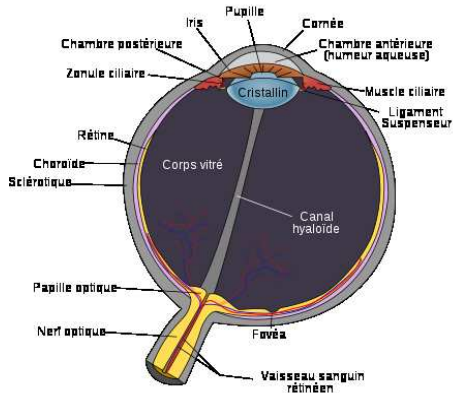
La nature ondulatoire de la lumière...

... nous indique :

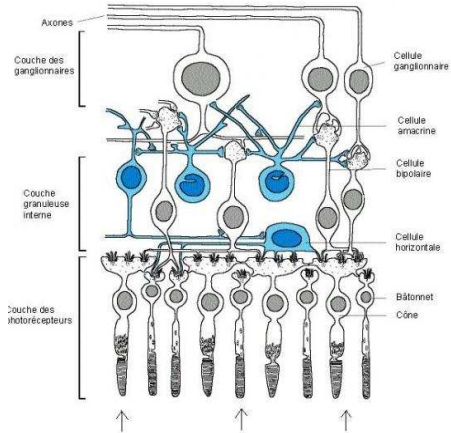
- ▶ qu'il y a propagation de l'énergie
- ▶ qu'à l'interface entre milieux différents, il y a réflexion et réfraction
- ▶ qu'il y a diffraction au niveau des obstacles
- ▶ que dans des milieux de dimensions limitées, il y a des phénomènes de résonances et d'ondes stationnaires, mais ils ne sont pas intéressants à l'échelle humaine macroscopique

- ▶ Le stimulus : la lumière
- ▶ Les capteurs humains (entre autres)

L'œil humain



La rétine



La rétine

- ▶ elle appartient au système nerveux central
- ▶ 3 couches de cellules nerveuses
- ▶ couche 1

Les bâtonnets

110 à 125 M

groupés

1 seul type

formes/luminosité

vision scotopique (nocturne)

vision périphérique

Les cônes

4 à 7 M

liaison unique

3 types : R V B (O ?)

couleurs

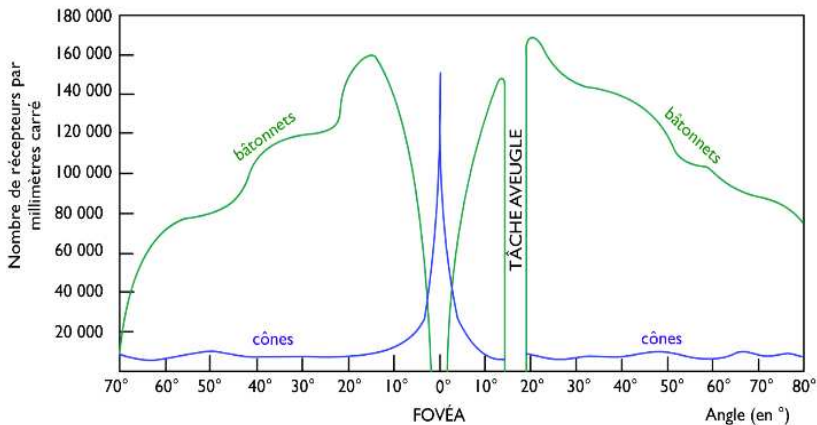
vision photopique (diurne)

acuité visuelle

La rétine

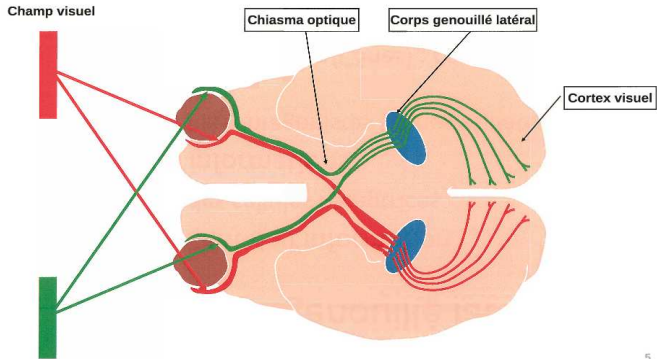
- ▶ couches 2 et 3 – voie directe : des photorécepteurs, aux cellules bipolaires, puis aux cellules ganglionnaires
- ▶ couches 2 et 3 – voie indirecte : les cellules horizontales et amacrines servent à moduler latéralement la transmission d'information ⇒ prise en compte d'événements issus du voisinage
 - ▶ pour s'adapter aux contrastes
 - ▶ aux bords des objets
 - ▶ au mélange des couleurs
- ▶ fovéa : vision précise (point de focalisation), plus détaillée, et sensibilité au mouvement

La rétine – Répartition des cellules



Vers le cerveau

Transmission



Vers le cerveau

- ▶ on constate que des neurones primaires proches l'un de l'autre envoie leur information dans des endroits proches du corps genouillé latéral
- ▶ sauf que, bizarrement, l'information va dans des endroits complètement différents du cerveau (niveau des corps genouillés/géniculés latéraux) suivant qu'on considère une tranche un peu à gauche ou une tranche un peu à droite du juste milieu du champ visuel
- ▶ le cerveau arrive à reconstruire comme il faut (du côté du cortex visuel)

Les corps genouillés/géniculés latéraux

- ▶ ils prétraitent l'information, directement à partir des cellules réceptrices, via les cellules ganglionnaires et amacrines
- ▶ il y a 3 types de cellules, qui correspondent à 3 types d'informations :
 - ▶ neurones de petits calibres : information venant des cônes (couleurs R et V), et concernant les fins détails spatiaux statiques
 - ▶ neurones de gros calibres : information venant des bâtonnets; objets en mouvement
 - ▶ les autres : information venant des cônes (couleurs B et O)

Vision stéréoscopique/binoculaire

- ▶ différence
 - ▶ entre les images issues des deux yeux
 - ▶ donne une information de profondeur (relief)
 - ▶ relief non discerné par certains (hypermétropie)
 - ▶ cas des borgnes de naissance : pas de vision en relief
 - ▶ cas des borgnes accidentels : le cerveau arrive à reproduire l'information manquante (// avec les sourds de naissance et les autres)
- ▶ le mouvement
 - ▶ donne une information de profondeur (relief)
 - ▶ par comparaison avec la scène précédente

D'autres yeux

- ▶ pas mal de mammifères ne perçoivent pas trop bien la couleur (les bovins ne semblent pas les percevoir du tout ; pour le chien, on ne sait pas trop ; le chat ne voit pas le rouge...)
- ▶ les oiseaux perçoivent efficacement la couleur, bien sûr
- ▶ les amphibiens voient bien les couleurs, les reptiles aussi
- ▶ champs visuels :
 - ▶ faucon : 180° (en plus, les rapaces ont deux fovéas ; leur acuité est 7-8 fois meilleure que la nôtre)
 - ▶ chat : 287°
 - ▶ lapin : 360°
 - ▶ homme : 200°

D'autres yeux

- ▶ l'œil du poulpe : le poulpe est un invertébré, il provient donc d'une chaîne évolutive très éloignée de la nôtre ; dans cette chaîne-ci aussi, l'œil optique a été inventé ; il est quasi identique au nôtre ; note : le calamar géant a un œil de 40 cm de diamètre
- ▶ l'œil composé des insectes : on quitte le monde des << yeux optiques >>
 - ▶ la résolution des instruments d'optique dépend du rapport $R = D/\lambda$: les insectes sont trop petits pour que ce soit efficace
 - ▶ or on sait que les abeilles ont une acuité visuelle excellente, et qu'elles voient bien les couleurs (les fleurs n'ont pas des couleurs pour nous, mais pour les abeilles)

D'autres yeux

- ▶ l'œil composé est constitué d'un grand nombre de cellules coniques disposées sur une sphère
- ▶ au sommet, il y a une surface transparente qui guide la lumière vers le fond où se produit l'absorption
- ▶ chaque sommet de cellule ne doit être ni trop large (résolution médiocre), ni trop fin (la diffraction devient gênante : on retombe sur l'optique)
- ▶ on obtient deux formules :

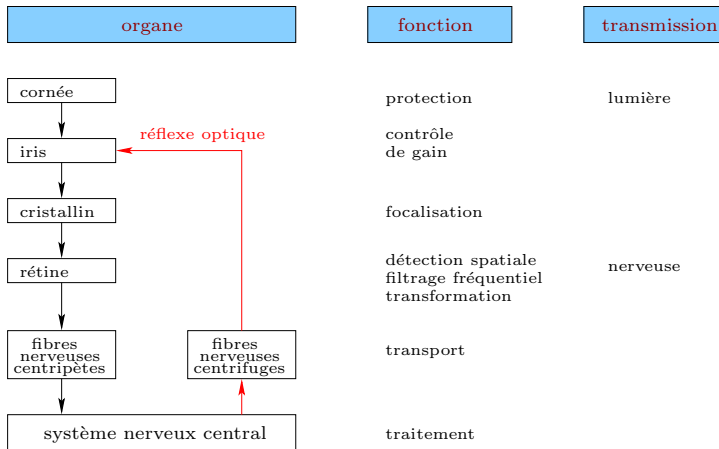
$$\Delta\theta_r = \delta/r \text{ et } \Delta\theta_d = \lambda/\delta$$

où δ est le diamètre de la cellule et r le rayon de la sphère

D'autres yeux

- ▶ il faut minimiser la somme des angles $\Delta\theta_d$ et $\Delta\theta_r$ par rapport à δ
- ▶ on obtient $\delta = \sqrt{\lambda r}$
- ▶ ceci correspond à ce que nous observons chez l'abeille : la formule nous donne $\delta = 35 \mu m$, et la réalité dit $\delta = 30 \mu m$
- ▶ note : certains satellites astronomiques, travaillant dans les longueurs d'onde plus grandes que le visible (rayons X, notamment), fonctionnent sur ce principe

Transmission de l'information



- ▶ Le stimulus : la lumière
- ▶ Les capteurs humains (entre autres)
- ▶ Les sensations

Les sensations visuelles

Elles recouvrent 4 phénomènes :

- ▶ **Le niveau** ; lié à l'énergie du stimulus
- ▶ **La couleur** (monochrome) ; lié à la fréquence du stimulus
- ▶ **La provenance** ; lié à la localisation du stimulus
- ▶ **Le mouvement** ; lié à l'évolution spatio-temporelle du stimulus

Les sensations visuelles – Les trois premiers

- ▶ Le fait que plusieurs bâtonnets soient liés à une cellule bipolaire implique :
 - ▶ un manque de précision sur la **localisation**
 - ▶ que la vision nocturne est floue
 - ▶ note : pas de vision des **couleurs** ; mais grande sensibilité au **niveau**
- ▶ Le fait qu'un seul cône (le plus souvent) soit lié à une cellule bipolaire implique :
 - ▶ qu'on a une grande précision quant à la **localisation** (acuité visuelle)
 - ▶ autour de la fovéa (tache jaune), les deux couches de cellules nerveuses internes sont rejetées encore plus en arrière, pour permettre une encore plus grande concentration de cônes
 - ▶ note : vision des **couleurs** ; sensibilité au **niveau** moindre

Les sensations visuelles – Couleurs

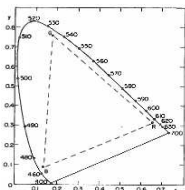
Un peu de subjectivité :

- ▶ couleurs froides, fraîcheur, apaisement, immobilité : de vert à magenta en passant par le bleu
- ▶ couleurs chaudes, chaleur, énergie, violence : de vert à magenta en passant par le rouge

Sensibilité aux couleurs : le rouge est plus lumineux que le bleu, mais il est plus sombre que le vert, qui est la couleur pour laquelle l'œil est le plus sensible

Les sensations visuelles – Couleurs

- ▶ La sensibilité aux couleurs est variable en fonction :
 - ▶ de l'intensité lumineuse
 - ▶ de la position en x et y
 - ▶ **note** : la sensation de provenance (pouvoir séparateur de l'œil) elle aussi dépend de la position
 - ▶ de la couleur, comme déjà dit
 - ▶ \Rightarrow il y aura problème à la restitution, donc
- ▶ les lieux des couleurs équivalentes sont des ellipses (de MacAdam) sur le diagramme de chromaticité



Les sensations visuelles – Le mouvement

- ▶ La sensation elle-même :
 - ▶ dépend de la vitesse d'assimilation d'une scène par le cerveau
 - ▶ réactions chimiques de la rétine \Rightarrow influx nerveux
 - ▶ interprétation par le cerveau
 - ▶ se caractérise par sa fluidité
- ▶ En ce qui concerne les séquences d'images :
 - ▶ nombre d'images/seconde assimilées

Les sensations visuelles – Le mouvement

Le phénomène est bien sûr directement lié à la persistance rétinienne

- ▶ de l'ordre de $1/15$ s
 - ▶ temps de récupération des cellules
 - ▶ temps d'interprétation du cerveau
- ▶ de ce fait, on obtient une sensation de mouvement fluide
 - ▶ à partir de 15 images par seconde

Les sensations visuelles – La persistance rétinienne

- ▶ quand le stimulus est une suite d'impulsions
 - ▶ le passage au noir reste visible jusqu'à 50 images par seconde
 - ▶ impression de papillotement/de battement
 - ▶ la télévision utilise 25 images par seconde \Rightarrow en fait, à cette cadence, certaines personnes perçoivent le battement décrit ci-dessus, et de ce fait on utilise plutôt 50 demi-images par seconde (les deux demi-images étant entrelacées)
- ▶ regardons pendant longtemps une surface/lumière rouge brillante, puis regardons un papier blanc : il nous paraît verdâtre

Les sensations visuelles – La résolution

- ▶ C'est la séparation de deux points contrastés
 - ▶ 1 minute d'arc (vision normale, c'est-à-dire 10/10), soit 1/60ème de degré
 - ▶ à 1 m de distance, ça correspond à 0.29 mm
 - ▶ à 3 m de distance, ça correspond à 0.87 mm
 - ▶ à 10 m de distance, ça correspond à 2.9 mm
 - ▶ sur la Lune : 110 kms
 - ▶ le Very Large Telescope a une résolution de 0.017 seconde d'arc : sur la Lune, ça correspond à 32 m
- ▶ Ça caractérise forcément les systèmes de diffusion :
 - ▶ en pixel/pouce
 - ▶ attention : la résolution nécessaire dépend de la distance

Conclusion à l'optique physiologique

- Il faudrait tenir compte de l'organisation que le cerveau fait de toutes ces informations : la vue n'est pas seule ; en réalité on n'est jamais hors contexte, comme les expériences mises en place pour obtenir les résultats ci-dessus voudraient que l'on soit : *l'observateur modifie ce qu'il observe*. Ceci a deux sens :
 - 1. premier sens (incertitude) :**
 - plus on observe finement la physique, moins on sait ce qui se passe en contexte ;
 - plus on s'intéresse au comportement en contexte, moins on sait ce qui se passe physiquement.

Conclusion

2. second sens :

un auditeur ayant été observé une fois ne peut plus, ou presque, être observé : il apprend... ainsi, vous-mêmes... vous êtes plutôt prêts pour observer, à présent...

⇒ tests spécifiques pour tromper les artistes ; pour vous tromper vous (pour nous tromper nous, physiciens et psychophysiciens)

Conclusion

- Emploi ambigu d'« Acoustique Physique » ? Deux sens à « Physique » :
 - ▶ relatif aux propriétés générales de la matière
 - ▶ relatif au corps humain
- La psychophysique en général est un domaine extrêmement vaste :
 - ⇒ seuls quelques résultats sont disponibles
 - ⇒ seuls quelques-uns d'entre eux sont présentés ici
- Liens entre la mécanique des solides et des fluides, la physique des matériaux, la chimie, l'électronique, l'architecture, l'écriture, la peinture, la psychologie/cognition, la médecine, les sciences de la Terre et de l'atmosphère, etc.

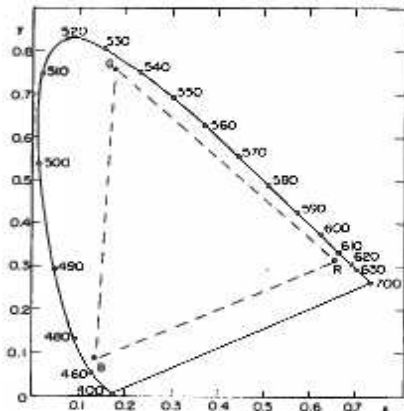
Conclusion – Critères de qualité visuelle

- ▶ image fixe
 - ▶ résolution (< 1 minute d'arc)
 - ▶ fidélité des couleurs
- ▶ image animée
 - ▶ fluidité (cadence des images > 15 images/s)
 - ▶ continuité (rafraîchissement > 50 stimulus/s)
- ▶ restitution de la profondeur
 - ▶ systèmes 3D
 - ▶ systèmes holographiques

- ▶ Les stimulus : la lumière
- ▶ Les capteurs humains (entre autres)
- ▶ Les sensations
- ▶ Modèles de perception

Diagramme de chromaticité (1/4)

Œil classique :



<http://www.tsi.telecom-paristech.fr/>

[pages/enseignement/ressources/beti/correl_couleur/chromaticity.html](http://pages.enseignement/ressources/beti/correl_couleur/chromaticity.html)

Diagramme de chromaticité (2/4)

Œil classique – détails (semsci.u-strasbg.fr/rouedes.htm) :

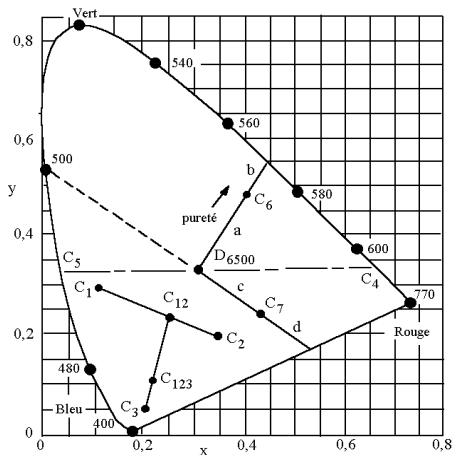


Diagramme de chromaticité (3/4)

Œil protanope (récepteurs rétinaux du rouge absents)

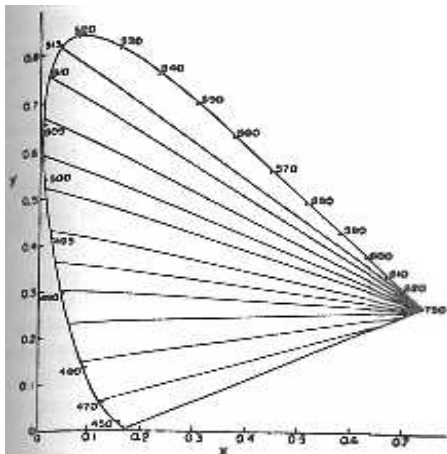
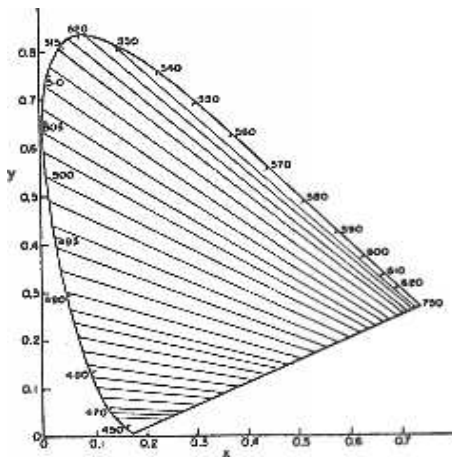
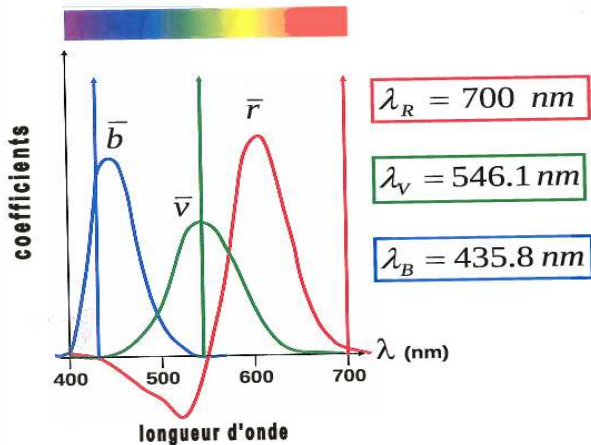


Diagramme de chromaticité (4/4)

Œil deutéranope (récepteurs rétinaux du vert absents)



Système 1 – Primaires RVB



Système 2 – Primaires CIE (ou XYZ)

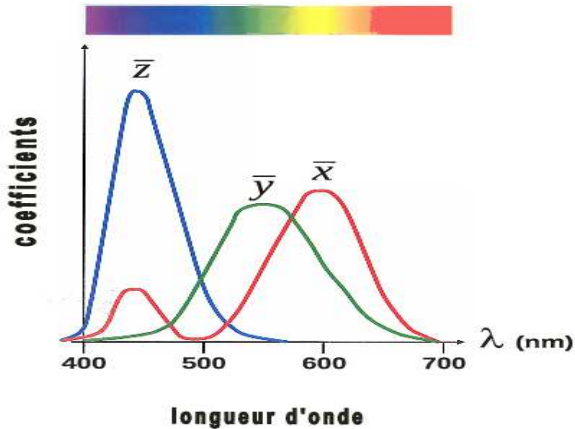
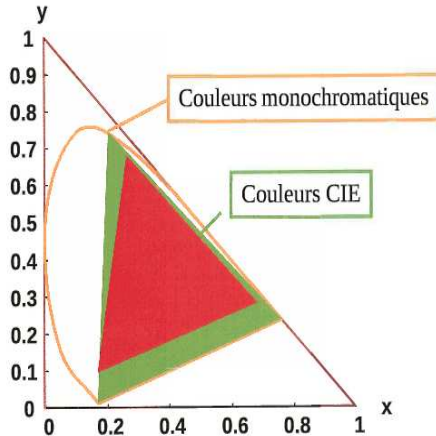


Diagramme de chromaticité et gamuts RVB et CIE



Système 3 – TSL (HSL/HSV)

- ▶ Teinte (ou Hue) : décrit la couleur, de 0 à 360 (orange, jaune, vert, bleu, rose, rouge)
- ▶ Saturation : c'est l'intensité d'une teinte spécifique, sa pureté (proportion de gris), de 0 à 255
- ▶ Luminosité : ajout d'une intensité de blanc ou de noir aux teintes (saturées ou non)

Applications : graphisme, imagerie...

Les couleurs et la vision

- ▶ l'espace des couleurs est non euclidien

- ▶ on ne peut pas écrire :

$$ds^2 = \sqrt{dr^2 + dv^2 + db^2}$$

- ▶ et pas non plus :

$$ds^2 = \sqrt{x^2 + y^2 + z^2}$$

- ▶ et d'aucune autre façon, en fait
- ▶ il est difficile de définir une distance colorimétrique simple

Modèles de production – Lumière

Stéphane Rossignol

2015 – 2016

Modèles de production – la lumière

Modèles de production – modèle sources

Autres sens

Plan

- ▶ Modèles de production pour la lumière

Système RVB

- ▶ Les niveaux de Rouge, Vert et Bleu
 - ▶ sont codés sur 256 valeurs
 - ▶ ce qui est meilleur que la résolution humaine
- ▶ Références
 - ▶ niveaux et couleurs
 - ▶ pas forcément identiques à l'enregistrement et à la restitution
 - ▶ nécessité d'adapter les périphériques

La couleur : RVB en Y/C

- ▶ Y/C est une norme de transport et de stockage de l'information
 - ▶ cassettes S-VHS, Hi8, câbles USHIDEN
- ▶ dans cette norme, les chrominances et la luminance circulent dans deux câbles différents
- ▶ ça permet de faire passer plus d'informations (et donc d'améliorer la qualité de l'image)

La couleur : RVB en Y/C

- ▶ la luminance : $Y = 0.299R + 0.587V + 0.114B$
- ▶ les chrominances : $W = R - Y$, $X = V - Y$ et $Z = B - Y$
- ▶ on choisit le système :
 - ▶ Y, W, Z
 - ▶ chrominances Rouge et Bleu directement disponibles
 - ▶ il est aisément possible de remonter au Vert

La couleur : RVB en Y/C

Voilà la matrice de passage d'un système à l'autre :

$$\begin{pmatrix} Y \\ W \\ Z \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ 0.701 & -0.587 & -0.114 \\ -0.299 & -0.587 & 0.886 \end{pmatrix} \begin{pmatrix} R \\ V \\ B \end{pmatrix}$$

Conséquences (1/3)

- ▶ images fixes
 - ▶ RVB sans pertes (bmp, ...) : lourd
 - ▶ JPEG : codage en Y/C et pas en RVB
 - ▶ JPEG : chrominances sous-échantillonnées
 - ▶ résolution fonction de la distance
- ▶ images animées
 - ▶ résolution
 - ▶ SD : 4/3 en 720×576 (1024×576 en anamorphose)
 - ▶ HD : 16/9 en 1280×720 ou 1920×1080
 - ▶ MPEG dérive du JPEG

Conséquences (2/3)

Exemple : vidéo en full HD

- ▶ débit nécessaire :
 - ▶ $1920 \times 1080 \times 3$ octets
 - ▶ 25 images/s
 - ▶ $\Rightarrow 1920 \times 1080 \times 3 \times 8 \times 25 = 1.24$ Gbits/s
- ▶ nécessité de compresser, donc
 - ▶ ce en conservant la qualité perçue !
 - 1. sous-échantillonnage des chrominances (d'un facteur 2 le plus souvent : 4 :2 :2)
 - 2. codage perceptuel M-JPEG, MPEG, etc.

Conséquences (3/3)

Compression/stockage :

- ▶ il faut comprendre ce qui est pertinent et ce qui ne l'est pas, notamment perceptuellement (codage avec perte)
- ▶ images : JPG versus BMP/GIF (taux de compression $\simeq 20/25$)
- ▶ vidéos : MPG versus AVI (taux de compression $\simeq 30$)
- ▶ *note au sujet des sons : MP3 versus WAV (taux de compression $\simeq 10$)*

Plan

- ▶ Modèles de production pour la lumière
- ▶ Modèles de production – modèle sources

Modèles sources

- ▶ il ne suffit pas de bien produire/reproduire les sons et la lumière
- ▶ il faut tenir compte aussi du contexte :
 - ▶ les acteurs
 - ▶ l'environnement
 - ▶ le mouvement
- ▶ ce pour une immersion aussi totale/parfaite que possible

Modèles sources – Les acteurs

- ▶ Caractéristiques visuelles
 - ▶ géométrie
 - ▶ courbes d'absorption
 - ▶ textures
- ▶ Caractéristiques acoustiques
 - ▶ les sons émis (monophoniques)
 - ▶ type d'émission : directivité

Modèles sources – L'environnement

- ▶ Visuel – Les éclairants
 - ▶ source ponctuelle
 - ▶ source diffuse
 - ▶ courbe d'émission
- ▶ Acoustique – les parois et les objets (passifs ou actifs)
 - ▶ absorptions, réflexions, réfractions...

Modèles sources – Le mouvement

Les acteurs (notamment) se meuvent dans l'environnement

- ▶ chaque source a une trajectoire
 - ▶ translations
 - ▶ rotations
- ▶ l'apparence dépend du temps
 - ▶ apparence visuelle
 - ▶ apparence auditive
 - ▶ modifiées par les positions relatives des objets et des sources

Modèles sources – La scène finalement

Elle doit conjuguer :

- ▶ une représentation objet : les acteurs, l'environnement où ils se trouvent et les trajectoires de tout ce petit monde
- ▶ et une représentation signal
 - ▶ image 2D ; 3D ; holographie
 - ▶ son stéréo ; n.1 ; holophonie (avec contrôle de l'acoustique de l'environnement)

Plan

- ▶ Modèles de production pour la lumière
- ▶ Modèles de production – modèle sources
- ▶ Autres sens

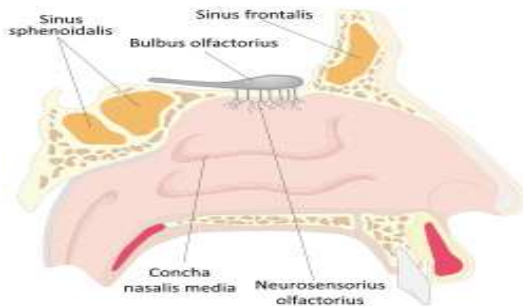
Et les autres sens alors ?

Page intentionnellement laissée blanche

Le nez : l'olfaction

- ▶ forte connection avec le goût
- ▶ au-delà de sa fonction perceptive : participe à la phonation
- ▶ le nez a 55 millions d'années (primates supérieurs et chauves-souris) ; les autres mammifères ont un rhinarium (truffe)
- ▶ les récepteurs ne commencent à être compris que depuis la fin du 20ème siècle... et on n'en sait pas beaucoup plus (l'environnement (génétique, culturel, etc.) distord tout)

Le nez : l'olfaction



Le goût

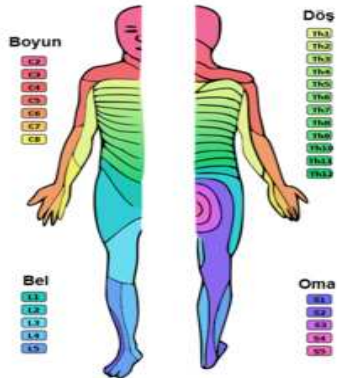
- ▶ forte connection avec l'olfaction
- ▶ 5 récepteurs différents chez l'homme (en première approximation) :
 - ▶ sucré (saccharose ; Adolph Fick, 19ème)
 - ▶ salé (chlorure de sodium ; Adolph Fick, 19ème)
 - ▶ amer (quinine ; Adolph Fick, 19ème)
 - ▶ acide (citron ; Adolph Fick, 19ème)
 - ▶ umami=savoureux (glutamates ; Kikunae Ikeda, 20ème)
- ▶ il y a mieux que l'homme : la souris a des récepteurs pour les acides gras et le calcium (chou, pavot)
- ▶ il y a des récepteurs pour le piquant (poivre, gingembre, piment), l'âpre/astringent (vin rouge, thé)

Le toucher

- ▶ pas très étudié, en fait (on sait qu'il est important pour l'apprentissage des autres sens comme la vision et l'audition...)
- ▶ la peau chez l'homme couvre 2 m^2 et pèse 5 kgs
- ▶ 4 types de récepteurs :

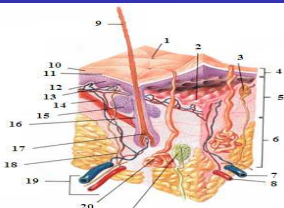
récepteur	perception	forme	durée de perception	grandeur du champ	fréquences
Merkel	pression lente : légère tape, petits détails	disque	tant que le stimulus est là	petit (1 mm)	0.3 à 3 Hz
Meissner	pression rapide : tremblement, frottement	boudin	ne dure pas	petit	3 à 40 Hz
Ruffini	étirement	cyindre	tant que le stimulus est là	grand (8 mm)	15 à 400 Hz
Pacini	vibration rapide	oignon	ne dure pas	grand	10 à 500 Hz

Le toucher : la peau



Dermatome – Chaque couleur correspond à la partie de la peau innervée par un nerf différent, à partir de la moelle épinière (lésions cérébrales)

Le toucher : la peau



- | | |
|---------------------------------------|-------------------------------|
| 1- Pore de transpiration | 2- Jonction dermo-épidermique |
| 3- Terminaison nerveuse (toucher) | 4- Épiderme (anatomie) |
| 5- Derme | 6- Hypoderme |
| 7- Veine | 8- Artère |
| 9- Poil | 10- Cornée |
| 11- Couche pigmentée | 12- Kératinocytes |
| 13- Mélanocytes | 14- Muscle érecteur du poil |
| 15- Glande sébacée | 16- Follicule pileux |
| 17- Bulbe | 18- Nerf |
| 19- Système lymphatique et vasculaire | 20- Glande sudoripare eccrine |
| 21- Corpuscule de Pacini | |

C'est la fin

Page intentionnellement laissée blanche

Les slides sont disponibles ici

[http ://www.metz.supelec.fr/metz/personnel/
rossignol/slides_tout_modeles.pdf](http://www.metz.supelec.fr/metz/personnel/rossignol/slides_tout_modeles.pdf)

Version définitive mise en ligne demain matin

Examen

N'oubliez pas l'examen, commun avec celui sur les interfaces innovantes et immersives, en janvier prochain

- ▶ est-ce que vous avez réussi à récupérer les slides ?
- ▶ je préfère ne pas les faire imprimer (même en noir et blanc, 4 slides par page)

Ce cours est un pré-requis pour le cours de traitement des images de Jean-Luc Collette

fin

Modèles de perception – le son

Stéphane Rossignol

2015 – 2016

Le stimulus auditif : le son

L'équation d'onde

Onde progressive

Onde stationnaire

Les capteurs

Les sensations auditives

Perception de l'intensité

Perception de la hauteur

Perception du timbre

Perception de l'espace

Conclusion à la psychoacoustique

Modèles de perception

Plan

- ▶ Le stimulus : le son

Qu'est-ce que le son ?

- ▶ au sens << commun >>, sensation perçue par l'oreille humaine
 - ▶ nous entendons environ entre 16 Hz et 16 kHz (10 octaves)
- ▶ au sens physique, phénomène physique provoquant, le cas échéant, cette sensation
 - ▶ variations périodiques (!) de la pression d'un milieu : l'onde acoustique
 - ▶ phénomène regroupant indifféremment les sons communs (y compris, en plus de nous, baleines, oiseaux...), les infrasons et les ultrasons (sonars ; chauves-souris, chiens, dauphins...)

Hypothèses simplificatrices

- ▶ le milieu de propagation (gaz, liquide, solide) est supposé être un fluide parfait (non visqueux) au repos
- ▶ les vibrations sont de faible amplitude (1/1000 pression normale)
- ▶ les transformations du fluide sont supposées adiabatiques réversibles (pas d'échange de chaleur)
- ▶ \Rightarrow la réalité est plus complexe ; dans l'absolu, on ne peut résoudre les équations que par une approche informatique : équations de Navier-Stokes (fluides visqueux) résolues par des méthodes aux éléments finis

L'équation d'onde

Page intentionnellement laissée blanche

Détermination de l'équation d'onde

On va la déterminer dans le cas d'une onde plane (déplacement selon x)

Déjà, qu'est-ce qui se passe ?

- ▶ si un objet se déplace en un endroit dans un fluide, nous observons une perturbation qui se déplace au travers de ce fluide
- ▶ quel type de perturbation ? \Rightarrow un changement de pression
- ▶ si l'objet est déplacé doucement, le fluide s'écoule autour de l'objet, et il n'y a pas de son
- ▶ si l'objet est déplacé suffisamment vite, le fluide est comprimé, la pression augmente localement, et elle pousse sur une quantité supplémentaire de fluide, et etc., et une onde se propage

Détermination de l'équation d'onde

- ▶ les particules (molécules/atomes) s'échappent d'une région A de densité plus élevée (crête de pression) vers une région B voisine de densité moins élevée (creux de pression)
- ▶ les particules venant de A donnent une quantité de mouvement aux particules de B
- ▶ afin que le son naisse, les régions A et B doivent être beaucoup plus grandes que le libre parcours moyen des particules, sinon celles-ci se déplaceraient librement depuis les crêtes vers les creux et effaceraient immédiatement l'onde
- ▶ du coup, on n'a pas besoin de considérer les particules individuellement (méc. quantique), mais on peut considérer des petites régions du fluide (méc. classique/phys. stat./méc. des fluides)

Détermination de l'équation d'onde

La physique du phénomène des ondes sonores comporte 3 équations :

- I le fluide se déplace et change de densité
- II le changement de densité correspond à un changement de pression
- III les inégalités de pression engendrent le mouvement du fluide

... dont il faut trouver les expressions ; et leur mélange va donner l'équation d'onde.

L'équation d'onde – Considérons d'abord II

- ▶ pour un gaz, un liquide ou un solide, la pression P est fonction de la densité ρ
- ▶ avant l'arrivée de l'onde sonore (à l'équilibre) : $P_0 = f(\rho_0)$; quand l'onde sonore passe : $P = f(\rho)$
- ▶ la variation de pression due au son est P_e ; il lui correspond la variation de densité ρ_e
- ▶ or, P_e est très petite
- ▶ donc, on a, bien sûr : $P = P_0 + P_e$, $\rho = \rho_0 + \rho_e$, et surtout

$$P_0 + P_e = f(\rho_0 + \rho_e) = f(\rho_0) + \rho_e f'(\rho_0)$$

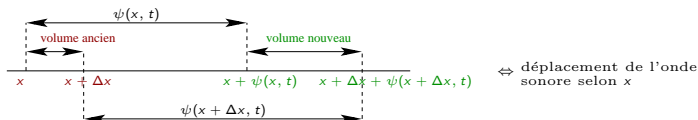
où $f'(\rho_0)$ est la dérivée de f en ρ_0

- ▶ P_e est lié linéairement à ρ_e :

$$\text{II} : P_e = \kappa \rho_e$$

L'équation d'onde – Considérons I (1/2)

- ▶ en $t = 0$, la position d'une tranche du fluide non perturbée par l'onde est x ; son déplacement au temps t dû au son est $\psi(x, t)$
- ▶ en $t = 0$, la position d'une tranche voisine est $x + \Delta x$; son déplacement au temps t est $\psi(x + \Delta x, t)$



- ▶ en $t = 0$, la quantité de fluide par unité de surface est $\rho_0 \Delta x$; cette quantité de fluide se retrouve au temps t entre $x + \psi(x, t)$ et $x + \Delta x + \psi(x + \Delta x, t)$

L'équation d'onde – Considérons I (2/2)

- ▶ Δx est petit, donc on peut écrire :

$$\psi(x + \Delta x, t) - \psi(x, t) = \frac{\partial \psi}{\partial x} \Delta x$$

- ▶ on a alors : $\rho_0 \Delta x = \rho \left(\frac{\partial \psi}{\partial x} \Delta x + \Delta x \right)$

- ▶ on peut négliger $\rho_e \frac{\partial \psi}{\partial x}$ devant $\rho_0 \frac{\partial \psi}{\partial x}$ et on obtient (note : $\frac{\partial \psi}{\partial x}$ lui aussi est petit) :

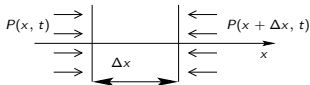
$$I : \rho_e = -\rho_0 \frac{\partial \psi}{\partial x}$$

- ▶ si les déplacements $\psi(x, t)$ varient avec x , il y a des changements de densité

L'équation d'onde – Considérons III

- ▶ prenons une fine portion de fluide de longueur Δx ; sa masse est $\rho_0 \Delta x$ et son accélération $\frac{\partial^2 \psi}{\partial t^2}$; nous avons donc une 1ère expression pour la force s'exerçant sur cette portion de fluide
- ▶ une 2ème expression pour cette force s'obtient aisément ; on a effet :

$$P(x, t) - P(x + \Delta x, t) = -\frac{\partial P}{\partial x} \Delta x = -\frac{\partial P_e}{\partial x} \Delta x$$



- ▶ et on a :

$$\text{III} : \rho_0 \frac{\partial^2 \psi}{\partial t^2} = -\frac{\partial P_e}{\partial x}$$

L'équation d'onde

▶ I : $\rho_e = -\rho_0 \frac{\partial \psi}{\partial x}$; II : $P_e = \kappa \rho_e$; III : $\rho_0 \frac{\partial^2 \psi}{\partial t^2} = -\frac{\partial P_e}{\partial x}$

- ▶ on peut les lier pour se ramener à une seule variable, disons ψ : on élimine P_e de III en utilisant II, puis on utilise I pour éliminer ρ_e ; et on obtient l'équation d'onde :

$$\frac{\partial^2 \psi}{\partial t^2} = \kappa \frac{\partial^2 \psi}{\partial x^2}$$

- ▶ note 1 : $\kappa = c_s^2$ est le carré de la célérité de l'onde
- ▶ note 2 : ρ_e et P_e obéissent à la même équation, et d'ailleurs pour le son la forme :

$$\frac{\partial^2 P_e}{\partial x^2} = \frac{1}{c_s^2} \frac{\partial^2 P_e}{\partial t^2}$$

est beaucoup plus utilisée

L'équation d'onde

- ▶ pour une onde plane de lumière se propageant selon x et polarisée de telle sorte que le champ électrique soit dans la direction y , on aurait :

$$\frac{\partial^2 E_y}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 E_y}{\partial t^2}$$

ceci découle des équations de Maxwell

- ▶ les équations de l'électrodynamique (pour la lumière) conduisent à la même équation d'onde que les équations de la mécanique (pour le son)
- ▶ la forme générale de l'équation d'onde pour le son est :

$$\Xi = \nabla^2 - \frac{1}{c_s^2} \frac{\partial^2}{\partial t^2} \quad \Xi P_e = 0 \quad (\text{d'Alembertien})$$

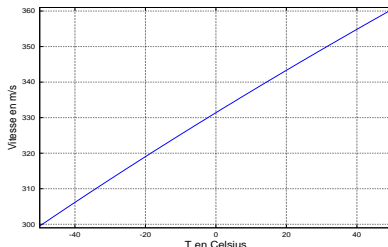
La nature ondulatoire du son...

... nous indique :

- ▶ qu'il y a propagation de l'énergie
- ▶ qu'à l'interface entre milieux différents, il y a réflexion et réfraction
- ▶ qu'il y a diffraction au niveau des obstacles
- ▶ que dans des milieux de dimensions limitées, il y a des phénomènes de résonances et d'ondes stationnaires

Célérité du son dans l'air

$$c_s = 331.4 \sqrt{1 + \frac{T}{273}} \quad (T \text{ en degrés Celsius})$$



Ainsi, pour $T = 20^\circ C$, on a $c_s = 343 \text{ ms}^{-1}$

Longueurs d'onde (dans l'air et ailleurs)

$$\lambda = \frac{c}{f}$$

pour $c_s = 340 \text{ ms}^{-1}$

fréquence	longueur d'onde
3 mHz	2500 m
1 Hz	340 m
10 Hz	34 m
100 Hz	3.4 m
1 kHz	34 cm
10 kHz	3.4 cm
16 kHz	2.1 cm
100 kHz	3.4 mm

⇐ *héliosismologie* ($c_s = 7 - 8 \text{ km/s}$)

⇐ sismologie, géologie

⇐ architecture

⇐ outils

Haut-parleurs, enceintes, micros, ...
oreille, audition, phonation

Solutions de l'équation d'onde

Nous allons considérer deux cas :

- ▶ celui des ondes progressives : les ondes se déplacent en champ libre
- ▶ celui des ondes stationnaires : les ondes sont enfermées dans une région de dimension finie (contraintes aux limites, sur les bords)

On va se placer dans le cas à une dimension (onde plane) : l'onde se déplace selon x

Onde progressive

Page intentionnellement laissée blanche

Solutions de l'équation d'onde – Onde progressive

- ▶ est-ce que la perturbation se déplace à vitesse constante ?
- ▶ est-ce que la perturbation peut aller soit vers la gauche soit vers la droite ?
- ▶ est-ce que deux perturbations peuvent se déplacer au travers l'une de l'autre (principe de superposition) ?

Tout ça doit être contenu dans l'équation d'onde.

Solutions de l'équation d'onde – Onde progressive

- ▶ une perturbation quelconque en forme d'onde plane qui se déplace à vitesse constante c_s s'écrit : $f(x - c_s t)$
- ▶ il faut vérifier que $P_e = f(x - c_s t)$ est une solution de l'équation d'onde
- ▶ la dérivée seconde de f par rapport à x est $f''(x - c_s t)$; la dérivée seconde de f par rapport au temps est $c_s^2 f''(x - c_s t)$: donc, en effet, toute perturbation sonore se propage avec la vitesse c_s , qui est liée à une propriété du fluide :

$$c_s = \kappa^{1/2} = \left(\frac{dP}{d\rho_{\text{en},\rho_0}} \right)^{1/2}$$

Solutions de l'équation d'onde – Onde progressive

- ▶ $g(x + c_s t)$, qui se déplace dans la direction opposée, satisfait également l'équation d'onde
- ▶ on vérifie aisément aussi que si : $h(x, t) = f_1(x, t) + f_2(x, t)$, où f_1 et f_2 satisfont l'équation d'onde, on a bien :

$$\frac{\partial^2 h}{\partial x^2} = \frac{\partial^2 f_1}{\partial x^2} + \frac{\partial^2 f_2}{\partial x^2} \quad \text{et} \quad \frac{\partial^2 h}{\partial t^2} = \frac{\partial^2 f_1}{\partial t^2} + \frac{\partial^2 f_2}{\partial t^2}$$

donc pas de souci avec le principe de superposition

- ▶ note :
 - ▶ si à certain temps $t = 0$ en un certain point x du fluide la pression a une valeur déterminée, au bout d'un temps t le fluide aura la même pression en un point $x + c_s t$ du fluide
 - ▶ on dit que l'image du mouvement se propage dans le milieu le long de l'axe des x avec une vitesse c_s appelée vitesse du son

Solutions de l'équation d'onde – Onde progressive

- ▶ dans les fluides parfaits, la vitesse du fluide est dirigée suivant la direction de propagation de l'onde
- ▶ de ce fait, on dit que les ondes sonores se propageant dans les fluides parfaits sont des ondes longitudinales
- ▶ en sismologie, on a aussi des ondes transversales (la Terre ne peut pas être considérée comme étant un fluide parfait)
- ▶ les vagues à la surface de la mer sont un mélange extrêmement compliqué d'ondes transversales et longitudinales (parce que notamment l'eau est (quasi)-incompressible)

Solutions de l'équation d'onde – Onde progressive

Pour résumer, dans une onde progressive :

- ▶ la pression du fluide varie ainsi :

$$P_e(x, t) = P_1(x - ct) + P_2(x + ct)$$

(somme d'une onde progressive et d'une onde rétrograde)

- ▶ la vitesse moyenne des particules du fluide (qu'il faut distinguer de la vitesse du son c_s !) varie ainsi :

$$v(x, t) = \eta_0 (P_1(x - ct) + P_2(x + ct))$$

- ▶ v est plus petit que c_s (voir les hypothèses simplificatrices)

- ▶ l'impédance de rayonnement est : $Z = \frac{P_e}{v} = \frac{1}{\eta_0}$ (réelle)

Solutions de l'équation d'onde – Onde progressive

- ▶ un cas important est le cas des ondes monochromatiques, pour lequel on a :

$$P_1(x, t) = A \cos(k(x - c_s t) + \phi_1)$$

où $k = \frac{\omega}{c_s} = \frac{2\pi f}{c_s} = \frac{\lambda}{2\pi}$ est le nombre d'onde

- ▶ on a bien sûr quelque chose de similaire pour P_2
- ▶ on manipule ceci souvent avec les complexes, du coup :

$$P_1(x, t) = R_e (A \exp(jk(x - c_s t) + \phi_1))$$

- ▶ dans le cas à général, k est un vecteur : c'est le vecteur d'onde
- ▶ toute onde peut être décomposée en un ensemble d'ondes monochromatiques ayant des vecteurs d'onde et des fréquences différentes : Fourier, analyse spectrale

Onde stationnaire

Page intentionnellement laissée blanche

Solutions de l'équation d'onde – Onde stationnaire

- ▶ jusqu'à présent, on a considéré un milieu illimité : des ondes de toutes fréquences peuvent s'y propager
- ▶ à présent, il faut respecter des conditions aux limites (sur les parois, etc.) : du coup, seules certaines fréquences sont permises (ce sont les *fréquences propres* ou *modes propres* du fluide dans un récipient donné)
- ▶ intuitivement, la longueur d'onde la plus grande est de l'ordre de l , la dimension du récipient

Solutions de l'équation d'onde – Onde stationnaire

- ▶ la solution onde progressive est toujours valable, sauf qu'elle est mal adaptée au calcul et à l'interprétation des phénomènes
- ▶ de façon plus appropriée, on cherche une solution à variables séparées
- ▶ on a alors :

$$p(x, t) = X(x)T(t)$$

- ▶ l'équation d'onde s'écrit alors, bien sûr :

$$\frac{1}{c_s^2 T} \frac{\partial^2 T}{\partial t^2} = \frac{1}{X} \frac{\partial^2 X}{\partial x^2}$$

Solutions de l'équation d'onde – Onde stationnaire

- ▶ pour qu'une fonction de t seul soit identique à une fonction de x seul, il est nécessaire que ces fonctions soient égales à une même constante, que nous notons (au hasard) $-k^2$:
- ▶ il vient alors, en posant $k^2 c_s^2 = \omega^2$ donc $\left(k = \frac{\omega}{c_s}\right)$, en considérant l'équation d'onde ci-dessus :

$$\frac{\partial^2 X}{\partial x^2} + k^2 X = 0$$

$$\frac{\partial^2 T}{\partial t^2} + \omega^2 T = 0$$

- ▶ k et ω sont réels (sinon, on aurait des ondes amorties)

Solutions de l'équation d'onde – Onde stationnaire

Dans le cas général, on peut alors écrire :

$$\begin{aligned}p(x, t) &= A \cos(kx + \phi) \cos(\omega t + \phi') \\v(x, t) &= B \sin(kx + \phi) \sin(\omega t + \phi')\end{aligned}$$

Ce qui s'écrit aussi, bien sûr :

$$\begin{aligned}p(x, t) &= A \cos(kx + \phi) \cos(\omega t + \phi') \\v(x, t) &= B \cos\left(kx + \frac{\pi}{2} + \phi\right) \cos\left(\omega t + \frac{\pi}{2} + \phi'\right)\end{aligned}$$

(À la rigueur on peut toujours s'arranger pour que ϕ et/ou ϕ' soi(en)t nulle(s) en choisissant convenablement l'origine des coordonnées et du temps ; mais elles servent aussi à ce que les conditions aux limites soient respectées)

Solutions de l'équation d'onde – Onde stationnaire

Remarque 1 :

- ▶ progressive versus stationnaire :
 - ▶ dans l'*onde progressive*, la pression p et la vitesse des particules v sont en phase (impédance réelle)
 - ▶ dans l'*onde stationnaire*, la pression p et la vitesse des particules v sont en quadrature (les ventres de vitesse correspondent aux nœuds de pression et inversement ; impédance imaginaire pure)
- ▶ note : pour démontrer ça, il faudrait passer par l'équation d'Euler $\frac{\partial p}{\partial x} = -\rho_0 \frac{\partial v}{\partial t}$, le potentiel des vitesses Φ tel que $v = \frac{\partial \Phi}{\partial x}$ et l'étude de la réflexion parfaite des ondes sur une surface...

Solutions de l'équation d'onde – Onde stationnaire

Remarque 2 :

- ▶ dans l'*onde progressive*, les phases $kx - \omega t + \alpha$ des oscillations en différents points de l'espace à un même instant sont différentes, et elles ne sont égales qu'en des points séparés par des distances égales à la longueur d'onde
- ▶ dans l'*onde stationnaire*, tous les points du fluide vibrent à tout instant avec la même phase $\omega t + \alpha$; on ne peut donc pas parler proprement de propagation (d'énergie, notamment) d'une telle onde

Onde stationnaire – 2 parois rigides

- ▶ Considérons deux parois rigides parallèles séparées de L
- ▶ Dans ce cas, la vitesse est nulle sur les parois :

$$\forall t \text{ on a : } v(0, t) = 0 \text{ et } v(L, t) = 0$$

- ▶ on a donc : $kL = n\pi$ avec $n \in \mathbb{Z}$, et $\phi = 0$
- ▶ c'est-à-dire : $\omega = kc = \frac{c\pi}{L}n$, ou $f = \frac{\omega}{2\pi} = \frac{c}{2L}n$

Onde stationnaire – 2 parois rigides

On obtient :
$$p(x, t) = \sum_{n=0}^{+\infty} A_n \cos\left(n\pi \frac{x}{L}\right) \cos\left(n\pi \frac{ct}{L} + \phi'_n\right)$$

Série harmonique de fondamentale $f_0 = \frac{c}{2L}$

- ▶ toutes les harmoniques sont présentes
- ▶ pression maximale sur les parois
- ▶ ça correspond à la résonance de salles
- ▶ ça correspond aussi à des instruments à vent fermés aux deux bouts
 - ▶ certains tuyaux d'orgue, flûte traversière

Onde stationnaire – 2 parois molles

- ▶ Considérons deux parois molles parallèles séparées de L
- ▶ Dans ce cas, la pression est nulle sur les parois :

$$\forall t \text{ on a : } p(0, t) = 0 \text{ et } p(L, t) = 0$$

- ▶ on a donc : $kL = n\pi$ avec $n \in Z$, et $\phi = \frac{\pi}{2}$
- ▶ c'est-à-dire : $\omega = kc = \frac{c\pi}{L}n$, ou $f = \frac{\omega}{2\pi} = \frac{c}{2L}n$

Onde stationnaire – 2 parois molles

$$\text{On obtient : } p(x, t) = \sum_{n=0}^{+\infty} A_n \cos\left(n\pi \frac{x}{L} + \frac{\pi}{2}\right) \cos\left(n\pi \frac{ct}{L} + \phi'_n\right)$$

Série harmonique de fondamentale $f_0 = \frac{c}{2L}$

- ▶ toutes les harmoniques sont présentes
- ▶ vitesse maximale sur les parois
- ▶ ça correspond aux tubes sans pertes, du type conduit vocal
- ▶ ça correspond aussi à des instruments comme la flûte de pan
- ▶ ça correspond aussi aux tunnels

Onde stationnaire – 1 paroi rigide et 1 molle

- ▶ Considérons deux parois parallèles séparées de L , une rigide (en $x = 0$) et une molle (en $x = L$)
- ▶ Dans ce cas, la vitesse est nulle sur la première et la pression est nulle sur l'autre ; soit : $\forall t$ on a : $v(0, t) = 0$ et $p(L, t) = 0$
- ▶ on a donc :
 - ▶ pour la pression :
 - ▶ $k0 + \phi \neq \frac{\pi}{2} + m\pi$
 - ▶ $kL + \phi = \frac{\pi}{2} + n\pi$ 2
 - ▶ pour la vitesse :
 - ▶ $k0 + \frac{\pi}{2} + \phi = \frac{\pi}{2} + n\pi$
 - ▶ $kL + \frac{\pi}{2} + \phi \neq \frac{\pi}{2} + m\pi$

Onde stationnaire – 1 paroi rigide et 1 molle

▶ soit :

▶ de 2 : $kL + \frac{\pi}{2} + \phi = (n + 1)\pi \Rightarrow k = \frac{2n + 1}{2} \frac{\pi}{L}$

▶ $\phi = 0$

▶ c'est-à-dire :

▶ $\omega = kc = \frac{2n + 1}{2} \frac{c\pi}{L}$

▶ ou $f = \frac{\omega}{2\pi} = \frac{2n + 1}{4} \frac{c}{L}$

Onde stationnaire – 1 paroi rigide et 1 molle

On obtient :

$$p(x, t) = \sum_{n=0}^{+\infty} A_n \cos\left((2n+1)\pi \frac{x}{2L}\right) \cos\left((2n+1)\pi \frac{ct}{2L} + \phi'_n\right)$$

Série harmonique de fondamentale $f_0 = \frac{c}{4L}$

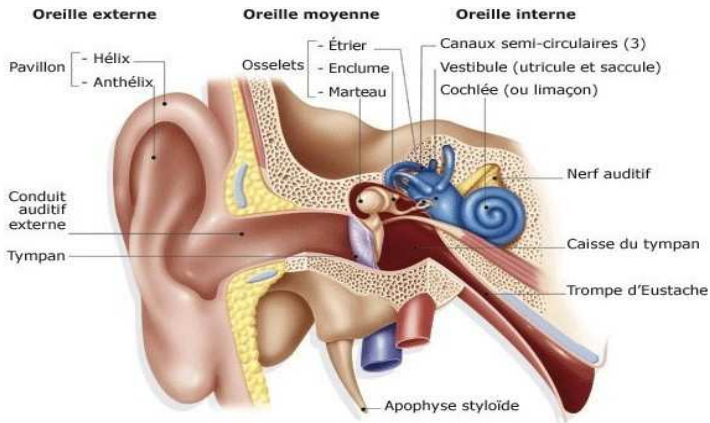
- ▶ seules les harmoniques impaires sont présentes
- ▶ ça correspond à des instruments << quintoyants >>, comme la clarinette, le didgeridoo, la corne de brume, les avertisseurs sonores des locomotives...
- ▶ ça correspond aussi aux bouteilles d'eau, aux cavités...

Onde stationnaire – La réalité, en ce qui concerne ces diverses parois

- ▶ Paroi imparfaitement molle
 - ▶ impédance de rayonnement non nulle
 - ▶ augmentation virtuelle de la longueur du tube
 - ▶ baisse de la fréquence de résonance
- ▶ Paroi imparfaitement rigide
 - ▶ impédance non infinie
 - ▶ impédance fonction de la fréquence
 - ▶ vitesse non nulle sur la paroi

- ▶ Le stimulus : le son
- ▶ Les capteurs humains (entre autres)

L'oreille – Un peu d'anatomie



Copyright © sanofi-aventis france

L'oreille externe

- ▶ composition
 - ▶ pavillon : forme et taille selon les individus ; utilisation pour la localisation des sources sonores
 - ▶ conduit auditif : longueur $\simeq 25$ mm ; diamètre $\simeq 6-10$ mm ; contourné : angle $\simeq 160-210^\circ$
 - ▶ tympan : ellipse d'axes $\simeq 9$ et 10 mm ; épaisseur $\simeq 0.1$ mm
- ▶ rôle
 - ▶ capter, amplifier (résonance due à la forme du conduit) et diriger les vibrations vers le tympan, qui vibre
 - ▶ protection de l'oreille moyenne et interne

L'oreille moyenne

► composition

- le tympan et la caisse tympanique
- la chaîne des osselets : le marteau, l'enclume et l'étrier
- la trompe d'Eustache

► rôle

- régulation de l'amplitude acoustique ; on passe d'un milieu aérien (grande amplitude et faible force) à un milieu aqueux (amplitude faible et efficacité)
- équilibrage de la pression (par la trompe d'Eustache, via la déglutition et le bâillement)

L'oreille interne

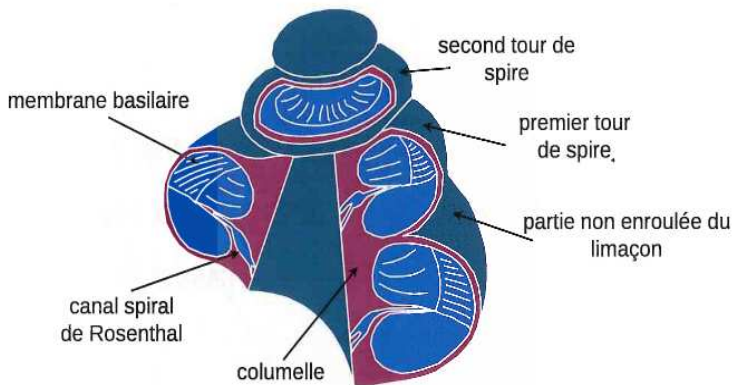
► composition

- le labyrinthe osseux
- le labyrinthe membraneux ou cochlée ou limaçon (2 tours et demi)
- liaison avec les nerfs crâniens via la membrane basilaire (entre 1/6 de mm et 1/2 mm)

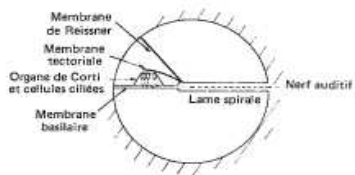
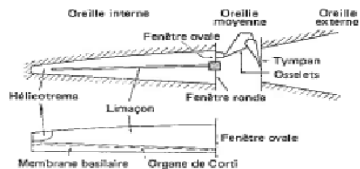
► rôle

- traitement fréquentiel de l'information acoustique
- transformation en influx nerveux : cellules ciliées externes (12000) : intensité; internes (3000-5000) : fréquence
- note : vestibule et canaux semi-circulaires servent à l'équilibration

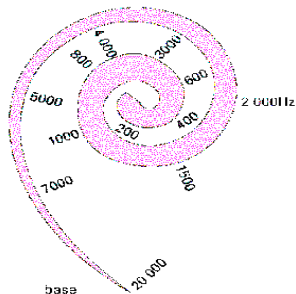
La cochlée



Modèles



La cochlée – Conversion position/fréquence

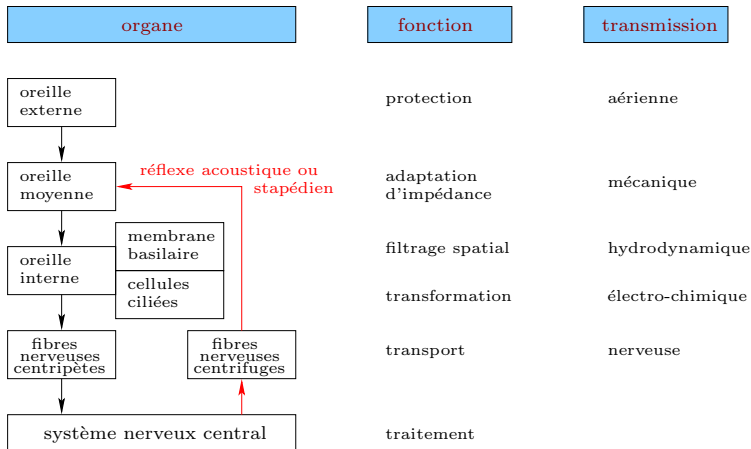


⇒ C'est von Békésy (1899 – 1972) qui comprend ça (prix Nobel de médecine en 1961)

Les cellules ciliées

- ▶ les cellules externes sont les plus fragiles
 - ▶ diminution avec l'âge
 - ▶ sensibles aux toxiques (médicaments)
 - ▶ atteintes en cas de diabète
- ▶ une diminution du nombre des cellules internes entraîne une diminution de la discrimination des fréquences ; **le sujet entend mais ne comprend pas**
- ▶ une atteinte des cellules externes entraîne une diminution de la sensation d'intensité ; **le sujet ne comprend pas bien, il entend une sorte de bruit désagréable**
- ▶ s'il y a encore des cellules internes, une audition dégradée peut être améliorée par des amplis

Transmission de l'information



D'autres oreilles

- ▶ chat : audition jusqu'à 65 kHz
- ▶ chien : audition jusqu'à 45 kHz
- ▶ chauve-souris : audition jusqu'à 120 kHz
- ▶ éléphant : entend les infrasons en-dessous de 10 Hz
- ▶ dauphin : audition jusqu'à 200 kHz (mais c'est dans l'eau, où la célérité du son est 4 fois plus élevée, or $\lambda = \frac{c}{f}$)
- ▶ cétacés : audition entre 10 Hz et 150 kHz

- ▶ Le stimulus : le son
- ▶ Les capteurs humains (entre autres)
- ▶ Les sensations auditives

L'influx nerveux

- ▶ Variation de potentiel cellulaires
 - ▶ potentiel d'action d'environ 100 *ms*, pendant environ 1 *ms*
 - ▶ identique quelle que soit la sensation
 - ▶ la fréquence est liée à l'intensité
- ▶ Localisation des sensations
 - ▶ cortex : zones de l'audition et de la vision
 - ▶ positionnement = informations de qualité et de localisation de la source

Ces sensations recouvrent 4 phénomènes acoustiques

1. La perception de l'**intensité** (énergie) d'un son : dépend de l'amplitude des vibrations de l'air
2. La perception de la **hauteur** (fréquence) d'un son pur : dépend de la fréquence des vibrations de l'air
3. La perception du **timbre** d'un son : notamment, dépend des amplitudes respectives des sons accessoires, les partiels, qui viennent s'ajouter au son pur
 - son pur = sinusoïde seule de fréquence f_0 = vibrations sinusoïdales de l'air
 - son composé (ayant un timbre) = fondamentale (f_0) + partiels (if_0 , i entier) = vibrations périodiques quelconque de l'air
4. La perception de l'**espace** : dépend des positions des sources émettant les vibrations

Sensations auditives

- ▶ Chaque sensation est liée à plusieurs paramètres du stimulus à la fois
- ▶ Tout est intriqué de façon plutôt compliquée
 - ▶ déjà au niveau de la perception (oreille physique et entrée dans le cerveau)
 - ▶ et ensuite au niveau de l'interprétation par le cerveau
- ▶ On va le voir (en partie) dans la suite

Perception de l'intensité

Page intentionnellement laissée blanche

Perception de l'intensité – Définitions

1. Le son en un point fixe de l'espace est décrit par la pression acoustique $p(t)$, où t est le temps. $p(t)$ décrit les variations de la pression de l'air en ce point. $p(t)$ est exprimée dans les mêmes unités que la pression de l'air : 1 Pascal = 1 N/m^2 . En fait, on considère la valeur efficace p_{eff} de $p(t)$.
2. Plus petite pression acoustique audible : $p_{eff} = 10^{-5} Pa$
Plus grande pression acoustique audible sans douleur : $p_{eff} = 10^2 Pa$
3. L , niveau de pression acoustique :

$$L = 20 \log_{10} \frac{p_{eff}}{p_0} \text{ dB}$$

où $p_0 = 2,0 \cdot 10^{-5}$.

- Notez que la pression atmosphérique normale est 101330 Pa

Le stimulus auditif : le son
Les capteurs
Les sensations auditives
Modèles de perception

Perception de l'intensité
Perception de la hauteur
Perception du timbre
Perception de l'espace
Conclusion à la psychoacoustique

intensite.m

Perception de l'intensité – Qualitativement

inaudible
juste audible
faible
moyen
fort
très fort
assourdissant
douloureux

seuil d'audition

DANGER

SEUIL DE DOULEUR (DESTRUCTIF)

Perception de l'intensité – Qualitativement

- 15 dB feuilles légères agitées par un vent doux dans un jardin silencieux
- 20 dB chuchotement/studio d'enregistrement/jardin paisible
- 25 dB conversation à voix basse entendue à 1.5 m
- 30 dB appartement dans un quartier paisible
- 35 dB bateau à voile
- 40 dB lieu calme/bureau dans un lieu calme
- 45 dB appartement normal avec les bruits minimaux de la rue
- 50 dB bruit d'une voiture au ralenti entendu à l'intérieur
- 60 dB conversation courante/grands magasins/rue résidentielle/bateau à moteur
- 65 dB valeur limite du bruit de l'environnement (routes, autoroutes) captée par les façades
- 70 dB restaurant bruyant/circulation importante
- 80 dB klaxon de voiture/mixer
- 85 dB atelier de tournage et d'ajustage

Perception de l'intensité – Qualitativement

À partir de ce seuil, le bruit est facteur de troubles auditifs

95 dB rue au trafic intense/atelier de forgeage

À partir de ce seuil, le bruit est pénible à entendre

100 dB baladeurs/scie à rubans/moto sans silencieux/marteau piqueur entendu à moins de 5m

105 dB discothèque (avec des crêtes à 120 dB)/raboteuse/méto (à l'intérieur)

À partir de ce seuil, le bruit est difficile à supporter

110 dB atelier de chaudronnerie

À partir de ce seuil, le bruit devient douleur

120 dB moteur d'avion à quelques mètres/concert rock ou techno

130 dB décollage d'un avion/marteau pilon

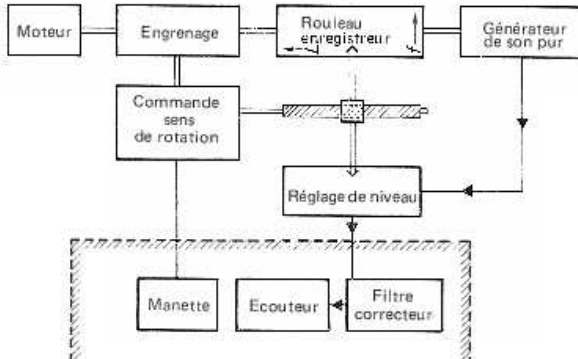
À partir de ce seuil, la loi exige une protection sociale

140 dB turbo-réacteur au banc d'essai

190 dB fusée au décollage

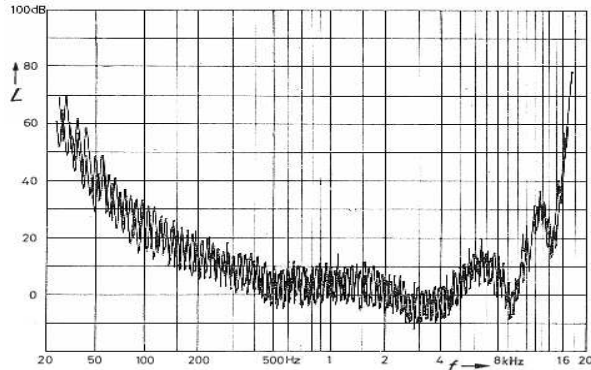
Perception de l'intensité – Seuil d'audition (loi du seuil)

Tracer en fonction de f_0 (son pur) la limite entre niveaux audibles et niveaux inaudibles; l'appareil de mesure (l'audiomètre de von Békésy) :



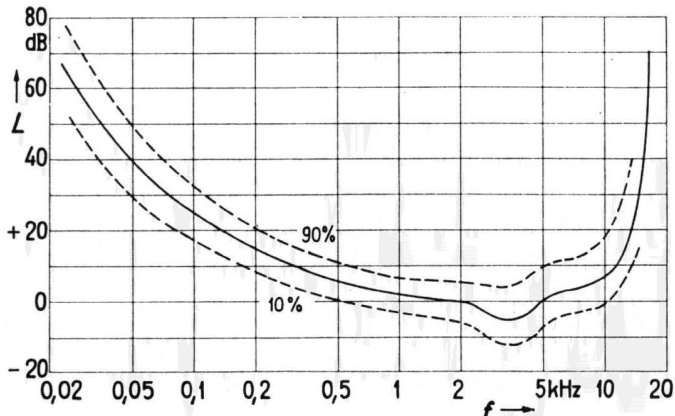
Perception de l'intensité – Seuil d'audition (loi du seuil)

Tracer en fonction de f_0 la limite entre niveaux audibles et niveaux inaudibles ; on obtient (pour un individu donné, plusieurs essais) :



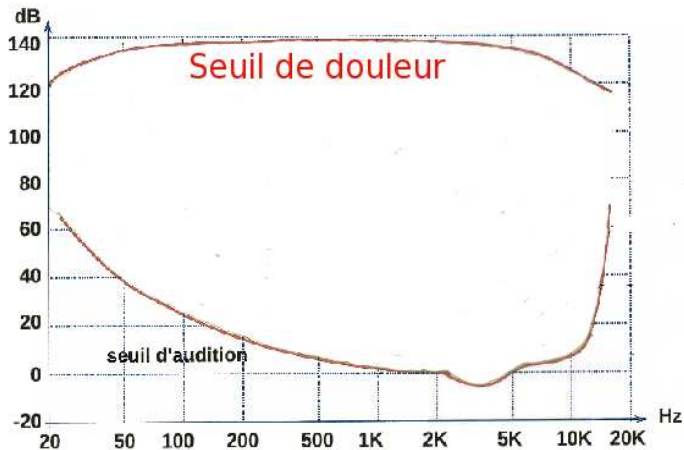
Perception de l'intensité – Seuil d'audition (loi du seuil)

Tracer en fonction de f_0 la limite entre niveaux audibles et niveaux inaudibles ; on obtient sur un ensemble d'individus :



seuilintens.m

Perception de l'intensité – Seuil de douleur



Perception de l'intensité – Courbes d'isophonie

Remarque 1. Pour mesurer la force sonore d'un son pur de fréquence x Hz, on prend une référence, par exemple un son pur de 1 kHz et de niveau $N = 40dB$ (ou phones ici), et on fait varier ensuite le niveau n du son dont on veut mesurer la force sonore jusqu'à ce qu'il *paraisse* (à un auditeur/sujet) aussi fort que le son de référence.

$n = f(x)$ (moyenne des résultats obtenus avec un grand nombre d'auditeurs) nous donne la courbe d'isophonie pour le niveau de référence N .

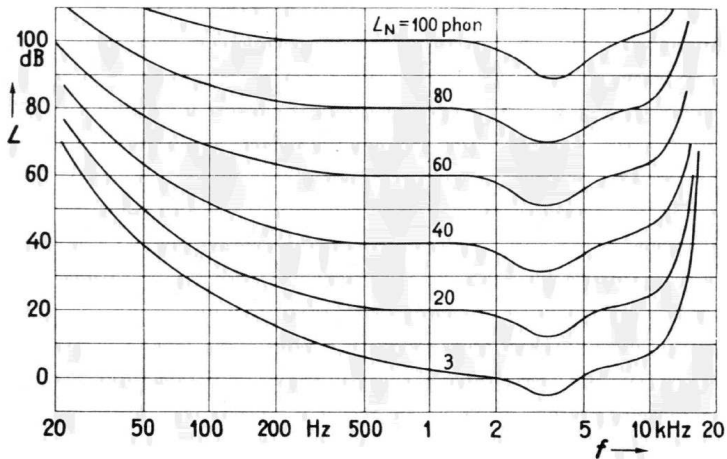
Ainsi, pour $N = 3$, on obtient de nouveau le seuil d'audition présenté sur la figure précédente.

Perception de l'intensité – Courbes d'isophonie

Remarque 2. Les courbes d'isophonie ne sont pas translatées l'une de l'autre

- ▶ De ce fait, la restitution d'un son enregistré ne peut se faire fidèlement qu'à l'intensité où il a été enregistré.
- ▶ Sinon, par exemple, on perd dans les graves si on baisse le volume sonore.
 - ▶ pour $L_N = 40 \text{ dB}$, il y a 25 dB entre 1 kHz et 50 Hz
 - ▶ pour $L_N = 30 \text{ dB}$, il y a 30 dB entre 1 kHz et 50 Hz
- ▶ En fait, on change le **timbre** de l'instrument (voir plus loin dans l'exposé).

Perception de l'intensité – Courbes d'isophonie



Perception de l'intensité – Seuil différentiel de perception

Il s'agit de déterminer la taille des variations d'amplitude juste audibles. Deux mesures sont utilisées.

Mesure 1 : Modulation d'amplitude

L'idée est de moduler une sinusoïde pure en amplitude, et ce suivant un taux de modulation m variable, et de demander à un auditeur de dire à partir de quel taux il entend que l'amplitude est modulée.

Signal :

$$s(t) = I(1 + m \cos(2\pi f_{tre} t + \phi_{tre})) \cos(2\pi f_0 t + \phi_0)$$

Perception de l'intensité – Seuil différentiel de perception

Signal :

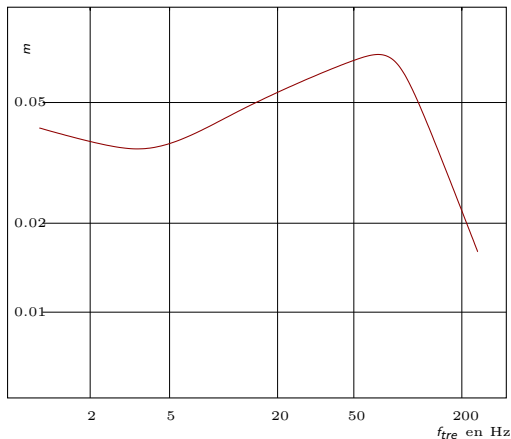
$$s(t) = I(1 + m \cos(2\pi f_{tre} t + \phi_{tre})) \cos(2\pi f_0 t + \phi_0)$$

En fait, ce taux-seuil dépend beaucoup de l'intensité moyenne I du son. Plus le son est fort, plus l'oreille est sensible aux variations d'amplitude de la sinusoïde.

Le phénomène est encore un peu plus compliqué que ça : le taux de modulation tout juste perceptible dépend aussi de f_{tre} . L'oreille est très sensible aux variations d'amplitude pour des f_{tre} situées dans une zone autour de 4Hz.

modulation.m

Perception de l'intensité – Seuil différentiel de perception



Seuil différentiel m en fonction
de f_{tre} . $L = 60$ dB. $f_0 = 1$ kHz

Perception de l'intensité – Seuil différentiel de perception

Mesure 2 : Phénomène de masquage

On a 1 *son parasite* et 1 *son test* masqué par le *son parasite*. Dans la pratique :

- Le *son parasite* considéré est un bruit à bande étroite de fréquence centrale 1 *kHz* et de largeur de bande 160 *Hz*
- Le *son test* est une sinusoïde pure de fréquence f_0

On demande à l'auditeur de déterminer, pour chaque f_0 , à partir de quel niveau acoustique du *son test* il entend ce *son test*. On détermine les courbes de masquage pour plusieurs niveaux L_G du *son parasite*.

Perception de l'intensité – Seuil différentiel de perception

Remarques :

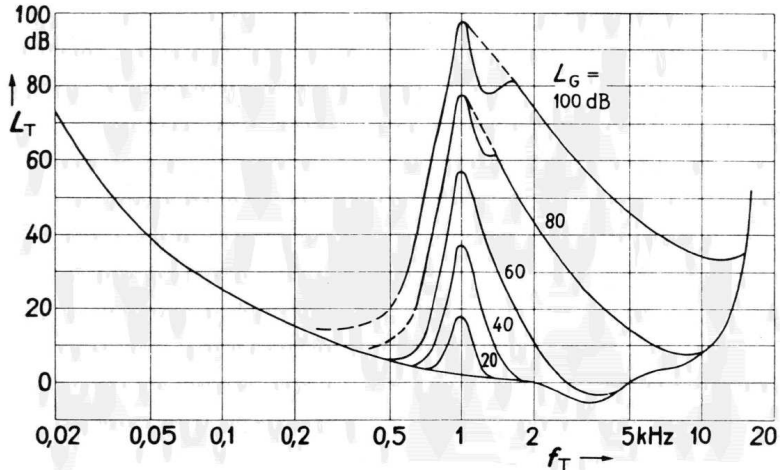
- Pour les fréquences du *son test* très supérieures ou très inférieures à 1 *kHz*, on rejoint le seuil d'audition
- On constate que, quand la fréquence du *son test* est la fréquence centrale du bruit, pour un bruit de niveau x *dB*, on entend un *son test* de niveau $x - 4$ *dB*

Le stimulus auditif : le son
Les capteurs
Les sensations auditives
Modèles de perception

Perception de l'intensité
Perception de la hauteur
Perception du timbre
Perception de l'espace
Conclusion à la psychoacoustique

masque.m

Perception de l'intensité – Seuil différentiel de perception



Perception de l'intensité – Pourquoi des bruits pour masquer ?

On utilise des bruits de bande comme sons masquants pour deux raisons :

- **Raison 1** : Le mixage de deux sons purs de fréquences très proches nous donne un son qui bat, c'est-à-dire un son modulé en amplitude :

$$\cos(2\pi f_1 t) + \cos(2\pi f_2 t) = 2 \cos\left(2\pi \frac{(f_1 - f_2)}{2} t\right) \cos\left(2\pi \frac{(f_1 + f_2)}{2} t\right)$$

Porteuse : $\frac{(f_1 + f_2)}{2}$ proche de f_1 ; modulation : $\frac{(f_1 - f_2)}{2}$ très petite, et m très grand devant 1

battement.m

Perception de l'intensité – Pourquoi des bruits pour masquer ?

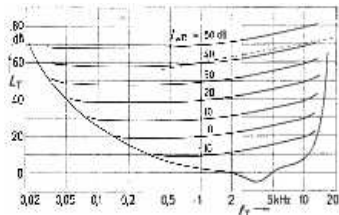
- **Raison 2** : Cela permet de mettre *l'auditeur musicien* dans les mêmes conditions que *l'auditeur qui ne connaît rien à la musique*. Par exemple, si on demande à un auditeur naïf de déterminer le son pur qui lui semble moitié moins haut qu'un son pur de 8 kHz, il donnera un son ayant pour fréquence environ 1.4 kHz (cette expérience nous donne la loi de **tonie** : voir la suite, *perception de la hauteur*). Le musicien utilisera ses connaissances musicales : il passera à l'octave inférieure et donnera la << bonne >> solution (4 kHz), qui en fait n'est pas celle qui est intéressante du point de vue de la psychoacoustique. Ça ne se passe plus ainsi si on utilise des bruits de bande : égalité des sujets.

Perception de l'intensité – Divers effets de masque

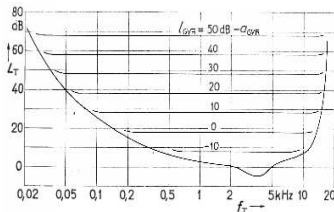
- ▶ un grand nombre d'effets de masque peuvent être mesurés
 1. on a vu ce qui se passe quand on masque avec un **bruit à bande étroite** large de 160 *Hz* et de fréquence centrale 1 *kHz*, dont l'intensité varie, un son test pur dont f_0 varie
 2. masquage avec un **bruit blanc**, dont l'intensité varie, d'un son test pur dont f_0 varie (seuils d'audition en présence de bruit)
 3. masquage avec un **bruit blanc uniformément masquant** (densité spectrale de puissance uniforme jusqu'à 500 *Hz*, puis décroissante linéairement avec f), dont l'intensité varie, d'un son test pur dont f_0 varie
 4. masquage avec un **bruit à bande étroite** et d'intensité fixe, dont la fréquence centrale varie, d'un son test pur dont f_0 varie
 5. masquage avec des sons purs ou des combinaisons d'harmoniques

Perception de l'intensité – Effets 2. et 3.

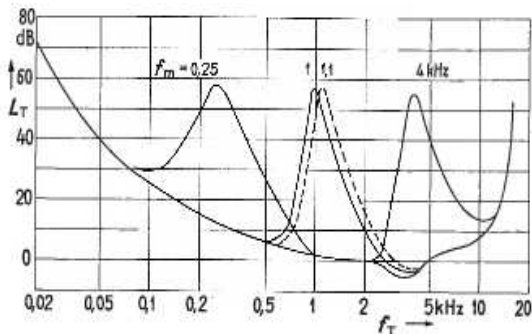
bruit blanc



bruit blanc uniformément masquant



Perception de l'intensité – Effet 4. – $L_G = 60 \text{ dB}$



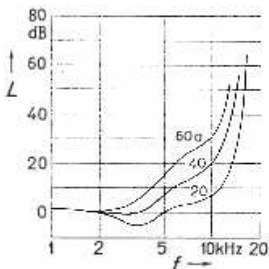
- ▶ peut être obtenu aussi en masquant avec un bruit passe-bas et un bruit passe-haut de même fréquence de coupure

Perception de l'intensité – Effets de masque

- ▶ les courbes de masquage des bruits blancs sont particulièrement plates et se translatent quasi
- ▶ les courbes de masquage par des bruits à bande étroite ne sont plus aussi simples
- ▶ les courbes de masquage obtenues pour des sons purs sont encore plus tourmentées et difficiles à interpréter
- ▶ de toutes ces mesures, on peut dire qu'en général des sons purs aigus et faibles sont masqués par des sons purs graves très intenses, mais pas par des sons purs aigus et intenses ; c'est notamment pourquoi on trouve dans un chœur mixte moins d'hommes que de femmes
- ▶ **ces effets de masquage sont intensément utilisés pour le codage à perte bien connu mp3**

Perception de l'intensité – Effet du vieillissement

Pour des personnes en bonne santé et dont l'ouïe n'a pas été endommagée par des travaux quotidiens en ambiance bruyante, on obtient ce type d'effets dus au vieillissement :



Perception de la hauteur

Page intentionnellement laissée blanche

Perception de la hauteur – Bandes critiques, mels, barks

Tout à l'heure, le choix d'une largeur de bande de 160Hz n'était pas innocent : il est lié à notion de *bande critique*. L'étude du phénomène de masquage fait apparaître :

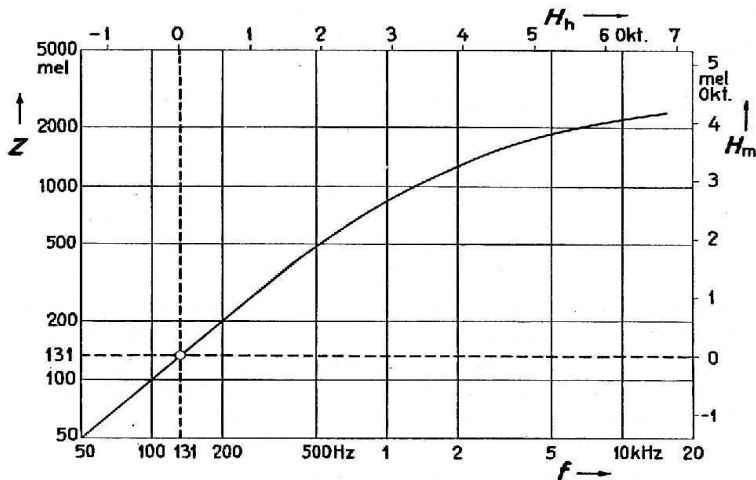
- qu'à l'intérieur d'une bande critique un son pur peut être masqué
- qu'à l'extérieur de cette bande il n'est pas masqué

Notre oreille est équipée de récepteurs sélectifs en fréquence, traitant des zones fréquentielles dont la largeur est précisément la largeur de la bande critique. Donc deux sons séparés de plus d'une bande critique excitent des récepteurs complètement disjoints. Ils sont quelles que soient leurs intensités respectives toujours discriminés (en 1ère approximation).

Perception de la hauteur – Bandes critiques, mels, barks

- ▶ La largeur de la bande critique n'est pas une constante en fonction de la fréquence.
- ▶ L'échelle des mels (ou barks ; 1 bark = 100 mels) est une échelle déduite de l'échelle des fréquences (en Hz) de telle façon que la bande critique soit de largeur constante.
- ▶ Expérimentalement, on constate que l'échelle des mels correspond à celle des Hz jusqu'à 500 Hz, puis la relation entre les deux échelles devient logarithmique.

Perception de la hauteur – Échelle des mels

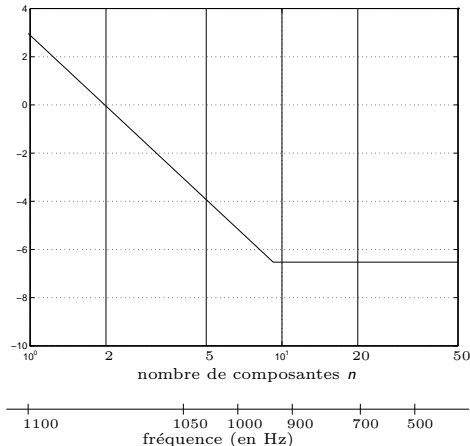


Perception de la hauteur – Bandes critiques

Une autre approche :

- ▶ un sujet écoute un signal sinusoïdal de fréquence f_0
- ▶ on ajoute une 2ème sinusoïde de même amplitude que la 1ère et de fréquence $f_0 + \Delta f$, où Δf est petit (20 Hz)
- ▶ on demande au sujet d'ajuster le volume de la somme pour que la perception de l'intensité du son qu'il a soit la même que pour le 1er son : normalement, on a A qui décroît de 3 dB
- ▶ on ajoute une 3ème sinusoïde de même amplitude que les 2 autres et de fréquence $f_0 + 2\Delta f$
- ▶ le sujet ajuste le volume : normalement, on a A qui décroît de $10 \log_{10} \left(\frac{1}{3} \right) = 4.77 \text{ dB}$
- ▶ on continue, jusqu'à ce qu'on constate qu'on n'a plus besoin d'ajuster le volume : on est sorti de la bande critique

Perception de la hauteur – Bandes critiques, une autre approche



⇒ Malheureusement, cette expérience n'est possible que pour des f_0 appartenant à la zone quasi linéaire du seuil d'audition

autreapproche.m

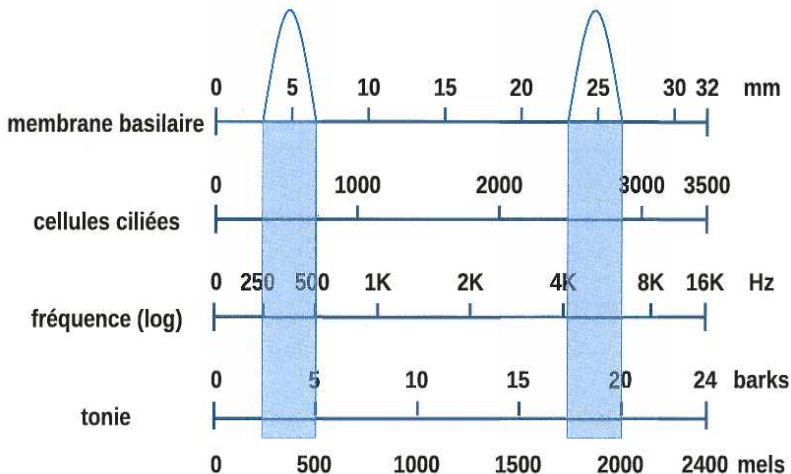
Perception de la hauteur – Bandes critiques

Nr	f_c	Δf		Nr	f_c	Δf
1	60	80		13	1850	280
2	150	100		14	2150	320
3	250	100		15	2500	380
4	350	100		16	2900	450
5	455	110		17	3400	550
6	570	120		18	4000	700
7	700	140		19	4800	900
8	845	150		20	5800	1100
9	1000	160		21	7000	1300
10	1175	190		22	8500	1800
11	1370	210		23	10500	2500
12	1600	240		24	13500	3500

Perception de la hauteur – Bandes critiques

- ▶ les bandes critiques n'ont pas une place fixe sur l'échelle des fréquences
- ▶ c'est le son test qui détermine leur lieu fréquentiel
- ▶ il y en a malgré tout toujours 24
- ▶ les bandes critiques sont adjacentes
 - ▶ note : dans le tableau précédent, en partant pour borne inférieure de la 1^{ère} bande critique 20 *Hz*, on peut déterminer toutes les limites de bandes

Perception de la hauteur – De l'anatomie à la tonie



Perception de la hauteur – Définition de la loi de tonie

- On a 1 *son pur de référence* de fréquence f_0 ; et 1 *son pur de test* dont on peut faire varier f_0 . Un auditeur écoute alternativement les 2 sons. On lui demande de dire quand le *son test* lui semble 2 fois plus haut que le *son référence*.
- Exemple : si on demande à un auditeur naïf de déterminer le son pur qui lui semble moitié moins haut qu'un son pur de 8 *kHz*, il donnera un son ayant pour fréquence environ 1.4 *kHz*.

Perception de la hauteur – Loi de tonie

Commentaires :

- il faut remarquer que la loi de tonie est équivalente à l'échelle des mels. 8 *kHz* correspond environ à 2000 mels; et 1.4 *kHz* à 1000
- la courbe précédente (échelle des mels) pourrait s'appeler : tonie Z (en mels) d'un son pur en fonction de sa fréquence f_0
- La hauteur en mels correspond à ce qu'on appelle la **hauteur mélodique**

Perception de la hauteur – Hauteur tonale

Définition dans le domaine temporel :

- La périodicité temporelle T_0 d'un son composé est le principal phénomène physique à mettre en rapport avec la perception de hauteur.
- Par exemple, tous les sons ayant une période de $T_0 = 2ms$ seront jugés comme étant des sons ayant la même hauteur (**hauteur tonale**), et en particulier la même hauteur qu'un son pur de fréquence $f_0 = \frac{1}{T_0} = 500Hz$.

Perception de la hauteur – Hauteur tonale

Définition dans le domaine fréquentiel :

- Du point de vue fréquentiel, la périodicité d'un son entraîne une répartition harmonique de ses composantes sinusoïdales.
C'est-à-dire que le son est la somme de sinusoïdes de fréquences f_0 , $2f_0$, $3f_0$...
- Donc, si on devait énoncer la méthode la plus simple possible pour mesurer la hauteur tonale perçue d'un son périodique à partir de son spectre, on dirait qu'il s'agit de déterminer le plus grand commun diviseur (PGCD) des fréquences de tous les partiels harmoniques
- Cette fréquence s'appelle *fréquence fondamentale* (ou encore la *fondamentale*, ou le *pitch* pour la parole) d'un son : c'est f_0

Perception de la hauteur – Hauteur tonale

- Dans la réalité, bien sûr :
 - ▶ les mesures du spectre sont entachées d'erreurs (les pics ne sont pas tout-à-fait où il faut)
 - ▶ la physique des instruments (des producteurs de sons en général) fait que les partiels ne sont pas tout-à-fait en $2f_0, 3f_0\dots$, mais en $2f_0 + \epsilon_2, 3f_0 + \epsilon_3\dots$
 - ▶ la physique des instruments (des producteurs de sons en général) fait que certains partiels manquent :
 - ▶ tous les partiels pairs ($2f_0, 4f_0\dots$) ; par exemple, la clarinette
 - ▶ le ou les premiers partiels manquent (f_0 , et les suivants) ; par exemple, le téléphone, qui coupe à 400 Hz , alors que le pitch d'une voix d'homme est typiquement de 100 Hz
 - ▶ etc.

Perception de la hauteur – Théorie simplifiée de l'harmonie

son monophonique : un son harmonique

- ▶ On utilise un générateur de sinusoides.
- ▶ On ajoute successivement des sinusoides aux fréquences $f_0 = 500\text{Hz}$, puis $2f_0$, $3f_0...$
- ▶ La première sensation consiste à entendre chacun des partiels harmoniques entrer séparément dans le son. Mais, rapidement, tous les partiels se fondent, pour ne plus former que la sensation d'un son complexe, de même hauteur que le son pur. Il n'est plus possible de distinguer séparément chacun des partiels du son.

harmonie.m

Perception de la hauteur – Théorie simplifiée de l'harmonie

sons polyphoniques : somme de plusieurs sons harmoniques

- ▶ Un accord musical sonne d'autant mieux que les sons fusionnent correctement. On a vu ci-dessus que des sons purs harmoniques avaient tendance à fusionner sans qu'on puisse les distinguer. L'analyse d'un accord se fait en superposant la représentation fréquentielle des sons le constituant, et en observant comment se superposent les partiels des sons.
- ▶ Dans le cas d'un accord d'octave ($1/2$), 1 partiel sur 2 fusionnent. C'est l'accord le plus consonant (par opposition à dissonant). Dans le cas d'un accord de quinte ($2/3$), à peu près 1 partiel sur 3 fusionnent. C'est un des accords les plus consonants après l'accord d'octave.

harmonie2.m

Perception de la hauteur – Théorie simplifiée de l'harmonie

sons polyphoniques : somme de plusieurs sons harmoniques (suite)

- ▶ Quand deux partiels se superposent mal, disons avec un écart de 10 Hz, ils produisent des battements, c'est-à-dire une modulation d'amplitude, dans le cas présent de 10 battements par seconde. Ce type de battement est :
 - ▶ trop rapide pour être perçu comme un phénomène temporel : au-delà de 180-200 bpm (environ 3 Hz), c'est difficile de suivre !
 - ▶ et trop lent pour être perçu comme un phénomène fréquentiel

C'est le phénomène de rugosité. La rugosité entraîne une ambiguïté de perception qui induit un stress et une dissonance de l'accord.

battement.m

Perception de la hauteur – Détermination hauteur tonale

Quelques pièges à éviter quand vous mettez en place un pitch-tracker

- ▶ La fréquence fondamentale d'un son n'est pas :
 - La fréquence du premier grand pic dans le spectre. Il peut en effet arriver que la *fondamentale* soit absente.
 - L'écartement entre grands pics consécutifs dans le spectre. Il peut arriver qu'il manque un grand nombre de partiels dans le son. Par exemple, il peut manquer tous les partiels pairs (clarinette).
 - La position dans le spectre du pic d'amplitude maximale. Il s'agit là de la *hauteur spectrale*, qui est liée surtout à la caractérisation du timbre. Il n'y pas de raison particulière pour que ce soit la fondamentale qui ait le plus d'énergie, même si plus la fréquence d'un mode est élevée, plus l'énergie requise pour l'exciter est grande.

Perception de la hauteur – Détermination hauteur tonale

- ▶ En fait, la perception de la hauteur tonale des sons complexes est difficile à expliquer.
- ▶ Physiquement c'est relativement simple si le signal est harmonique : il s'agit de déterminer le f_0 qui, d'un point de vue probabiliste, explique au mieux les partiels présents, en sachant bien qu'il peut y avoir un léger écart entre la position théorique nf_0 du partiel de numéro d'ordre n et sa position réelle f_n (c'est toujours vrai pour le piano, par exemple).
- ▶ Mais, ça, ce sont des problèmes de traiteurs de signaux. Des problèmes d'analyse spectrale.

hauteurspectrale.m

(On note aussi l'insensibilité de l'oreille à la phase des signaux.)

Perception de la hauteur – Détermination hauteur tonale

- 1 D'ailleurs, on peut faire intervenir dans ces problèmes de la psychoacoustique :
 - les déviations sur les premiers partiels doivent être plus petites que pour les autres, puisque on prouve que plus la fréquence d'un partiel est élevée plus on a du mal à juger de sa mauvaise position
 - plus l'amplitude du partiel est petite, plus on a du mal à juger de sa mauvaise position

Perception de la hauteur – Détermination hauteur tonale

- De plus, généralement, les amplitudes des partiels d'ordre élevé étant plus petites en moyenne, les deux points qui précèdent se rejoignent
- Avec certains sons, il est de réels problèmes de perception de la hauteur tonale. Notamment avec les sons de cloche, par exemple, pour lesquels différents auditeurs peuvent percevoir différentes hauteurs tonales. Admettons qu'un tel son soit polyphonique (et pas complètement consonant).

Perception de la hauteur – Seuil différentiel de perception

- ▶ Il s'agit de déterminer la taille des variations de fréquence juste audibles. On utilise la modulation de fréquence.
- ▶ L'idée est de moduler une sinusoïde pure en fréquence, et ce suivant un taux de modulation A_{vib} variable, et de demander à un auditeur de dire à partir de quel taux il entend cette modulation de fréquence.
- ▶ Le signal est :

$$s(t) = \cos \left(2\pi f_0 t + \phi_0 + \frac{A_{vib}}{f_{vib}} \sin (2\pi f_{vib} t + \phi_{vib}) \right)$$

- ▶ Note : c'est un vibrato !

Perception de la hauteur – Seuil différentiel de perception

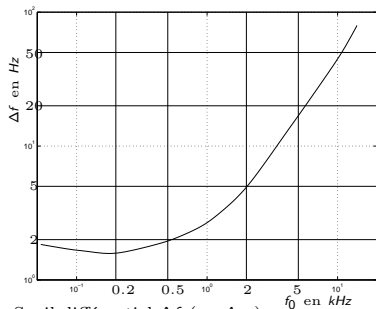
En fait, ce taux-seuil dépend :

- de f_0 : au-dessous de 500 Hz il ne varie pas, au-delà il croît linéairement avec f_0
- de l'intensité du son : jusqu'à $L = 30 \text{ dB}$, plus le son est fort moins il est besoin que la modulation soit grande pour être perçue ; au-delà de $L = 30 \text{ dB}$, l'intensité n'a presque plus d'influence
- de f_{vib} : oreille sensible aux variations de fréquence pour $f_{vib} \in [2 \dots 5] \text{ Hz}$

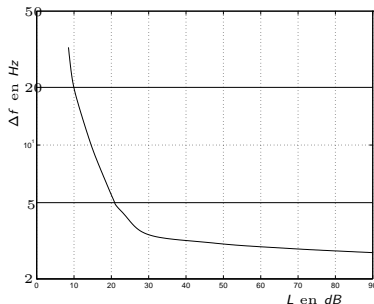
modulation2.m

voir thèse p. 119-120, pour développement en fonctions de Bessel

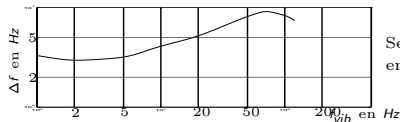
Perception de la hauteur – Seuil différentiel de perception



Seuil différentiel Δf (ou A_{vib})
en fonction de f_0 . $L = 70$ dB. $f_{vib} = 4$ Hz



Seuil différentiel Δf (ou A_{vib})
en fonction du niveau. $f_0 = 1$ kHz. $f_{vib} = 4$ Hz



Seuil différentiel Δf (ou A_{vib})
en fonction de f_{vib} . $L = 70$ dB. $f_0 = 1$ kHz

Lier la perception de l'intensité, la perception de la hauteur et la perception du timbre

Nous écoutons une sinusoïde pure dont la fréquence décroît exponentiellement de 1000 Hz vers 0 Hz. Au bout d'un certain temps, nous ne percevons plus rien : nous ne sommes plus dans la bande audible. Note : la fin du son est signalée par une courte note à 500 Hz.

Nous considérons à présent un son plus riche en partiels, dont la fondamentale est la sinusoïde à fréquence décroissante précédente. Bien sûr, plus il y a de partiels, plus nous entendons longtemps le son. Et nous reconstituons la fondamentale.

⇒ De ce fait, quand on mesure le seuil d'audition avec l'audiomètre de von Békésy, le signal doit être le plus sinusoïdal possible : il doit être sans timbre

dispartition.m

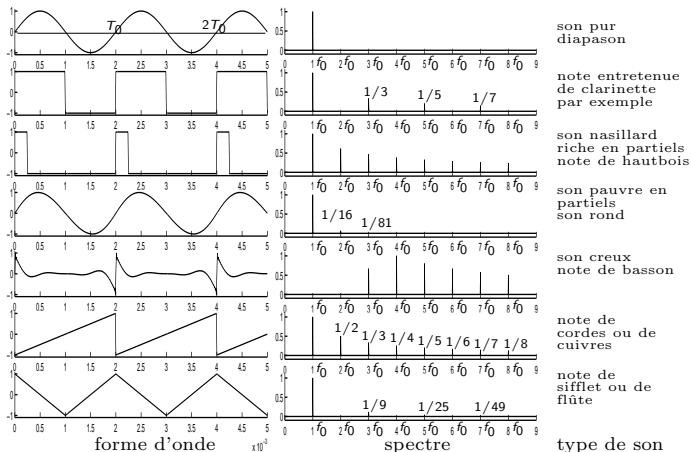
Perception du timbre

Page intentionnellement laissée blanche

Perception du timbre – Définition du timbre

- Le timbre, selon l'Association Américaine de Normalisation (1960), est tout ce qui permet à l'auditeur de différencier deux sons de **même intensité** (paramètre 1) de **même hauteur** (paramètre 2), présentés de façon similaire : même durée, même << **espace** >> (paramètre 4), même lieu, même auditeur...
- En fait, le timbre est la signature de la source du son : de la voix, de l'instrument qui l'a produit. C'est ce qui nous permet de différencier les instruments.
- A priori, la définition de l'AAN se limite aux sons entretenus.

Perception du timbre – Formes d'onde et spectres types



type de son

Perception du timbre – Sons types correspondants

- ▶ sinuspur.wav
- ▶ carre.wav
- ▶ rectangulaire.wav
- ▶ rond.wav
- ▶ moinsdeux.wav
- ▶ dentsdescie.wav
- ▶ triangle.wav

Perception du timbre – Espace des timbres

a. Quelques remarques :

- Tout ces noms (sons creux, sons ronds, etc.) ne sont en fait pas complètement prégnants.
- D'ores et déjà, on peut dire que l'étude des amplitudes respectives des partiels (c'est-à-dire ce que nous avons fait pour le moment) n'est pas suffisante pour caractériser les timbres.
 - ▶ par exemple, on va voir que les *attaques* de notes sont très importantes
 - ▶ qu'en est-il de la voix (pseudo-stationnarité assurée sur quelques dizaines de millisecondes seulement) ?
- Mais, en fait, la caractérisation des timbres est encore à faire.
Quelques voies sont données ci-dessous.

Perception du timbre – Espace des timbres

b. Mesures subjectives

La mesure de l'espace des timbres consiste à demander à des auditeurs de juger du degré de dissemblance entre des sons d'instruments calibrés (même intensité, même hauteur tonale...). Il est noté sur une échelle de 1 à 10 :

- 10 représente 2 sons très différents : la trompette et le piano par exemple
- 1 représente 2 sons très semblables : le saxophone et la clarinette par exemple

Il s'agit alors de ranger les instruments dans un espace à n dimensions, en prenant en compte ces distances. Puis il faut essayer d'expliquer physiquement chacune de ces n dimensions.

Perception du timbre – Espace des timbres

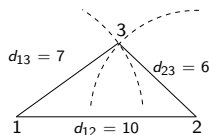
c. Un problème \Rightarrow il faut prendre n le plus petit possible!!!!!!!

Si l'on a M instruments :

- La plupart du temps¹ on parvient à les ranger dans un espace à $n = M - 1$ dimensions : cette solution n'est pas très intéressante!

$$M = 3$$

$$n = M - 1 = 2$$



- Mais dans 1 espace à $n = M - N$ (pour $N > 1$) dimensions? Il faut relâcher les contraintes sur les distances : ajout d'1 barre d'erreur sur chaque distance.

1. Il me semble qu'il suffit que pour tout couple i, j on ait $d_{ik} + d_{kj} \geq d_{ij} \forall k \neq i$ ou j , où d_{ij} est la distance entre les instruments i et j .

Perception du timbre – Espace des timbres

d. Ce qui existe : un espace à 3 dimensions

Ainsi, les 3 premières dimensions perceptives (ou les 3 plus significatives) de l'espace des timbres

correspondent/correspondraient physiquement :

dimension 1. Au temps d'attaque. Sa qualité est primordiale pour reconnaître un son d'un autre

Exemple

son1.wav

son2.wav

Perception du timbre – Espace des timbres

d. **Ce qui existe** : un espace à 3 dimensions

dimension 2. Au centre de gravité du spectre. C'est à mettre en rapport avec le maximum d'énergie dans le spectre et avec ce qu'on appelle la *hauteur spectrale*, qui n'est pas la *hauteur tonale*.

dimension 3. Au flux spectral, c'est-à-dire la mesure de l'évolution du spectre dans le temps. C'est l'aire de la différence entre deux spectres d'amplitude calculés pour deux portions/trames/frames temporelles du son larges de quelques dizaines de millisecondes et décalées de quelques millisecondes.

Perception du timbre – Formants

- La différence entre un << e >> et un << a >> chantés avec la même intensité et à la même hauteur tonale, ou parlés en essayant de tenir la voyelle afin d'avoir une plus grande plage de pseudo-stationnarité, se situe dans l'enveloppe spectrale
- ▶ L'enveloppe spectrale est la courbe qui relie les grands pics du spectre, qui correspondent aux partiels harmoniques
- ▶ Les formants sont des maximums d'énergie dans le spectre, c'est-à-dire des maximums dans l'enveloppe spectrale
- ▶ Chaque formant couvre donc plusieurs partiels

Perception du timbre – Formants

- On caractérise chaque voyelle communément à partir des 3 premiers formants. Ainsi, leurs positions sont, dans le cas d'une voix d'homme, environ :
 - ▶ pour le << i >> : 308 Hz, 2064 Hz et 2976 Hz
 - ▶ pour le << e >> : 365 Hz, 1961 Hz et 2644 Hz
 - ▶ pour le << a >> : 684 Hz, 1256 Hz et 2503 Hz
- On peut par exemple modéliser chaque formant par une gaussienne, d'amplitude et de largeur différentes pour chaque formant

formants.m

formants2.m

Phonème : l

formants3.m

Phonème E

Remarques :

- ▶ ça marche beaucoup moins bien que pour le l
- ▶ il y a des fausses détections : 1200 Hz, 1300 Hz, etc.
- ▶ du coup, l'interpolation par spline est plus chaotique

formants4.m

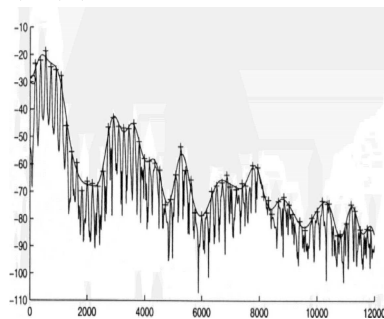
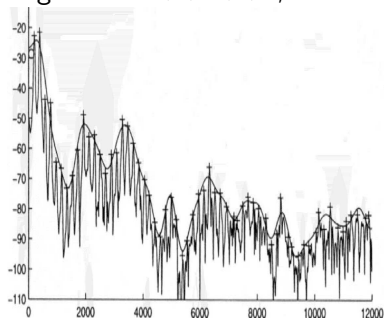
Phonème A

Remarques :

- ▶ ça marche beaucoup moins bien que pour le l
- ▶ il y a des fausses détections : 1000 Hz
- ▶ du coup, l'interpolation par spline est plus chaotique

Perception du timbre – Formants

À gauche : << e >> ; à droite : << a >> .



⇒ Remarque de Beneth Smith

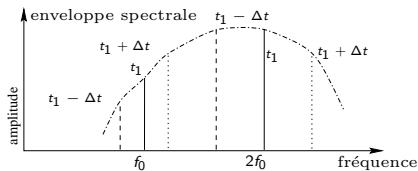
Perception du timbre – Autres caractéristiques

Le timbre est une mesure très subjective, prenant en compte :

- Des caractéristiques fréquentielles, comme les formants (qui ne résument pas complètement l'étude des amplitudes respectives des partiels : cas où il manque les partiels pairs)
- Des caractéristiques temporelles, comme l'attaque d'une note.
- Elle prend aussi en compte des caractéristiques d'autres natures. En particulier, un son, pour être vivant, doit être modulé, doit vibrer. Il y a deux types de vibrations :
 - ▶ **le vibrato**, qui est une modulation, sinusoïdale le plus souvent, de la fréquence (hauteur)
 - ▶ **le trémolo**, qui est une modulation de l'énergie (intensité)

Perception du timbre – Modulations

Les deux modulations sont liées. On admet qu'en première approximation les formants sont fixes, c'est-à-dire qu'ils ne dépendent pas de la fondamentale f_0 à laquelle est dite ou chantée la voyelle. Ainsi, si f_0 est modulée, l'amplitude de chaque partiel change dans le temps : elle suit le tracé des formants. Et donc l'amplitude du signal est elle aussi modulée, à la même fréquence que l'est f_0 .



$$\Delta t = \frac{1}{4f_{vib}}$$

Perception du timbre – Modulations

Notes :

- Ainsi, il existe deux sortes de trémolo, simultanément présents :

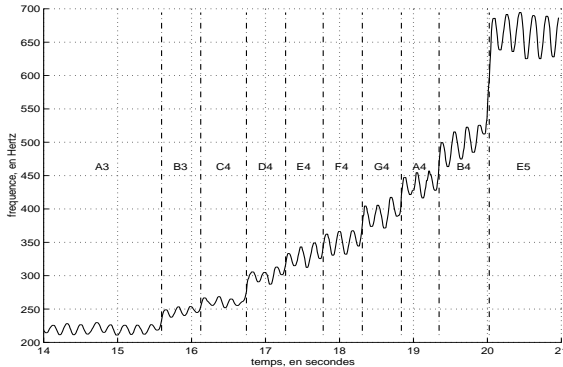
- L'un induit par le vibrato
- L'un << naturel >>

⇒ Comment les séparer ?

- L'oreille est sensible autour de 4 – 5 Hz aussi bien pour le vibrato f_{vib} que pour le trémolo f_{tre} . Ça tombe bien : c'est la fréquence la plus souvent choisie par les chanteurs.

Perception du timbre – Un exemple : analyse (voiceP.wav)

On a le trajet de f_0 original pour un extrait de voix chantée (2ième moitié du son : après que la chanteuse a eu repris son souffle) :



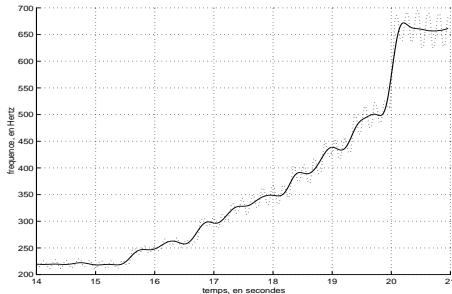
Perception du timbre – Un exemple : synthèse

Le son peut être resynthétisé avec un logiciel de l'IRCAM qui s'appelle Diphone. On a besoin :

- Du trajet de la fondamentale
- Du trajet de l'énergie : ici, on a pris la même énergie tout le long du signal, sauf au moment où la chanteuse reprend son souffle où on l'a mise à 0 \Rightarrow c'est la première transformation effectuée
- Du phonème chanté : ici, un << a >>
- De la voix du chanteur : est-ce un chanteur ou une chanteuse ? un ténor ou un soprano ou etc. ? (il s'agit de placer les formants où il faut)

Perception du timbre – Un exemple : 2 sons resynthétisés

- Premier temps, avec le trajet de f_0 original
- Second temps, en modifiant le trajet de f_0 ; notamment, on a essayé d'éliminer le vibrato (méthode explicitée plus loin), et resynthèse du son avec ce nouveau trajet de f_0



Exemple

- ▶ **son original** (voiceP.wav)
 - amplitude du vibrato et amplitude du trémolo grandes
 - dernière note chantée avec beaucoup d'intensité
- ▶ **son resynthétisé avec le trajet de f_0 original** (avecVibrato.wav)
 - trajet de l'énergie << plat >> (trémolo dû au vibrato)
 - énergie à zéro pendant la reprise de souffle
- ▶ **son resynthétisé avec le trajet de f_0 une fois le vibrato supprimé** (sansVibrato.wav)
 - trajet de f_0 sans vibrato
 - trajet de l'énergie << plat >>
 - énergie à zéro pendant la reprise de souffle

Vibrato/Trémolo – Pourquoi ? Comment ?

Le vibrato a trois buts psychoacoustiques :

- ▶ Rendre la voix vivante/le son vivant
- ▶ Rendre pour le chanteur le passage d'une note à la suivante plus aisée
- ▶ Cacher le fait que le chanteur/l'instrumentiste ne parvient pas toujours à accrocher parfaitement la note qu'il veut accrocher

Notes :

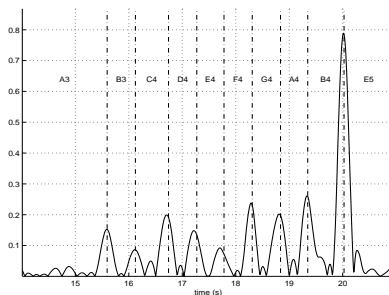
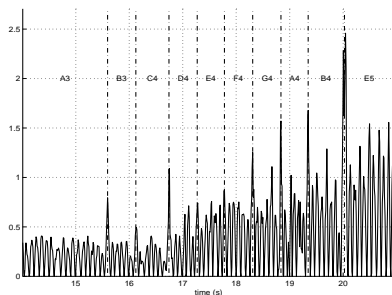
- ▶ Le vibrato, dans la musique occidentale, est le plus souvent sinusoïdal
- ▶ Ce n'est pas le cas pour le trémolo, à cause du mélange des deux types de trémolos (naturel – sinusoïdal – et dû au vibrato – hyper complexe –)

Perception du timbre – Un exemple : remarques

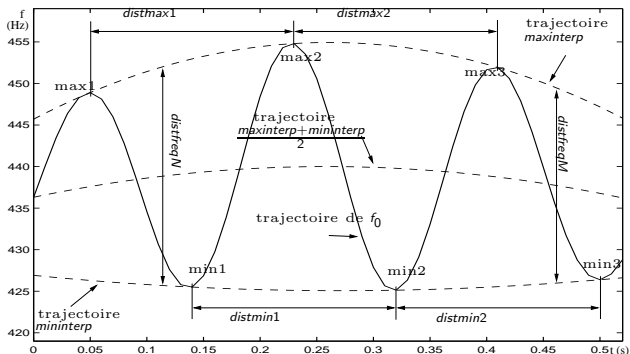
- ▶ On constate qu'on récupère bien les dix notes présentes
- ▶ D'autres modifications auraient pu être apportées : changer de voix, par exemple, ou changer de f_0
- ▶ Mais souvent pour que les changements de timbre soient de qualité à la resynthèse, il faut avoir segmenté avec précision le son, c'est-à-dire avoir donné tous les instants de transition entre 2 notes pour la musique ou entre 2 phonèmes pour la parole
- ▶ La première idée est de tracer f_0 dans le temps et de déterminer les sauts brusques de f_0
- ▶ Seulement, l'amplitude en fréquence d'un vibrato peut être supérieur à l'écart en fréquence entre 2 notes, comme c'est le cas ici, d'où la nécessité de s'affranchir du vibrato : de le supprimer du trajet de f_0
- ▶ Le vibrato, pour un tas de raisons, complique le travail du segmenteur!

Perception du timbre – Un exemple : et la segmentation ?

Gauche : dérivée de f_0 originale ; **droite** : dérivée de f_0 sans vibrato



Perception du timbre – Supprimer le vibrato

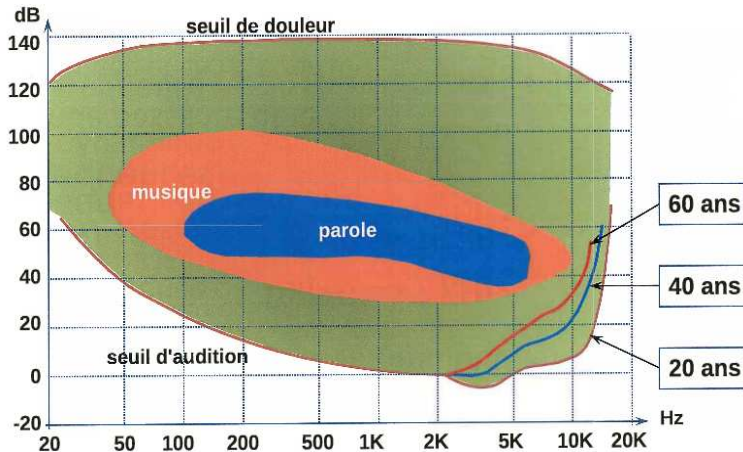


Perception du timbre – Détecter le vibrato

	Vmax	Vmin	Pmax	Pmin	Mfreq
voix chantée (vibrato fort)	0,0010	0,0012	85 %	84 %	0,087
flûte (vibrato quasi inexistant)	0,0090	0,0100	50 %	46 %	0,032

- ▶ V_{\max}/V_{\min} : variance des $\text{dist}_{\min}/\text{dist}_{\max}$
- ▶ P_{\max}/P_{\min} : proportion de $\text{dist}_{\min}/\text{dist}_{\max}$ compris entre 3 et 6 Hz
- ▶ M_{freq} : moyenne des $\text{dist}_{\text{freq}}$

Parole versus Musique



Domaine de la parole

- ▶ Zone concernée :
 - ▶ en fréquence : environ de 100 *Hz* à 7 *kHz*
 - ▶ en dynamique : à partir de 35 *dB* et jusqu'à 75 *dB*
- ▶ Conséquences :
 - ▶ bande passante utile limitée en basses fréquences et en hautes fréquences
 - ▶ niveau de quantification (numérisation du son) : 8 bits ($q^2/12 = 53$ *dB*) suffisent pour que le bruit de quantification soit inaudible ou au moins acceptable

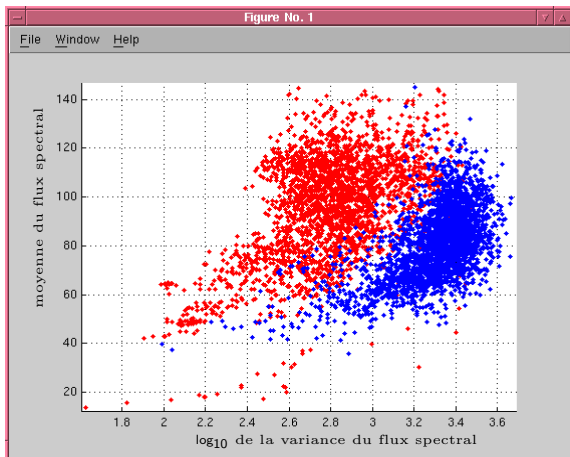
Domaine de la musique

- ▶ Zone concernée :
 - ▶ en fréquence : environ de 30 *Hz* à 10 *kHz* ou même 16 *kHz*
 - ▶ en dynamique : à partir de 30 *dB* et jusqu'à 100 *dB*
- ▶ Conséquences :
 - ▶ la bande passante utile couvre la bande audio complète
 - ▶ niveau de quantification (numérisation du son) : 16 bits (environ $q^2/12 = 101$ *dB*) sont nécessaires pour assurer la dynamique du signal

quantification2.m

Perception du timbre – Parole/Musique

Ce qu'on peut déjà faire grâce au flux spectral seul :



Perception du timbre – Parole/Musique – Comment ça marche ?

Objectif : il s'agit de discriminer

- les **segments** (temporels) de **musique**

⇒ musique = musique instrumentale et/ou voix chantée

- des **segments** (temporels) de **voix parlée**

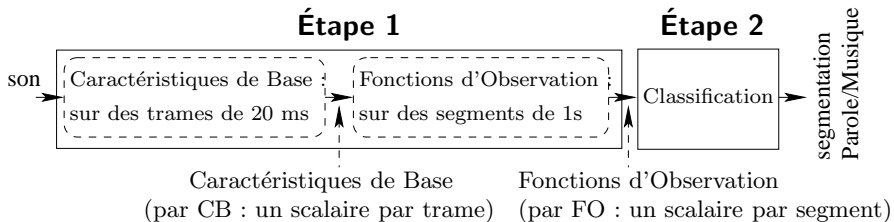
Il faut mettre en évidence :

- **musique** : succession de zones de stabilité relative
- **voix parlée** : rapide succession de zones de stabilité relative
et de zones de bruit

⇒ parole et musique enregistrées à la radio

Perception du timbre – Parole/Musique – Stratégie

Procédure :



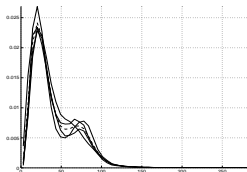
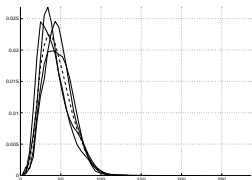
Perception du timbre – Parole/Musique – Caractéristiques de base

Caractéristiques de base mises en place :

- Le flux spectral : c'est la somme échantillon fréquentiel par échantillon fréquentiel de la valeur absolue de la différence entre deux spectres d'amplitude ; le 1er est calculé sur une portion pondérée de largeur T du signal (T vaut quelques dizaines de millisecondes), puis le second est calculé sur une portion pondérée du signal décalée de quelques millisecondes par rapport à la première ; les spectres d'amplitude sont calculés ici par FFT (méthode non-paramétrique), ou par modélisation auto-régressive (méthode paramétrique)
- Le centroïde (hauteur spectrale)
- Le taux de passage par 0 (appartient au domaine temporel)
- etc.

Perception du timbre – Parole/Musique – Caractéristiques de base – Densités de probabilité pour le flux spectral

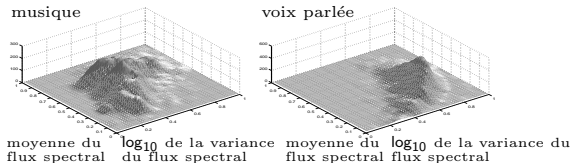
À gauche : **musique** ; à droite : **voix parlée**



- **musique** : un seul mode → les caractéristiques de base restent relativement constantes sur un segment
- **voix parlée** : deux modes (le second plus rare que le premier) → les caractéristiques de base varient beaucoup sur un segment

Perception du timbre – Parole/Musique – Fonctions d'observation

- **Espérance** par segment pour chaque Caractéristique de Base
- \log_{10} **variance** par segment pour chaque Caractéristique de Base



- **Nombre de trames au-dessous de la moyenne** par segment pour chaque Caractéristique de Base
- **Position \times du max de l'histogramme** par segment pour chaque Caractéristique de Base
- **Produit \times par \log_{10} de la variance du mode correspondant** par segment pour chaque Caractéristique de base

Perception de l'espace

Page intentionnellement laissée blanche

Perception de l'espace – Définition

Il s'agit d'obtenir des indices sur la position de la ou des sources présentes dans la salle où l'on se trouve. Il existe quatre sortes de critères :

- Les critères monauraux
- Les critères binauraux
- Les effets de salle
- Les critères de mouvement

Perception de l'espace

1. **Critères monauraux** : ils regroupent les indices qui impliquent 1 seule oreille :
 - Les circonvolutions du pavillon de l'oreille entraînent des atténuations différentes pour les ondes sonores en fonction de leur direction de provenance
 - Certaines personnes ont encore l'usage/la maîtrise des muscles qui permettent, instinctivement ou non, de dresser les oreilles (bruit fort, bruit << inquiétant >>...); bien sûr, c'est pour mieux percevoir (observer les chiens et les chevaux)
 - La perception de l'élévation ne se fait que de manière monaurale (elle est bien moins précise que la localisation horizontale)

Perception de l'espace

2. Critères binauraux : ils regroupent tous les indices qui impliquent les 2 oreilles :

- Différence d'intensité entre les deux oreilles : c'est ce qui est utilisé en stéréophonie en HIFI
 - ▶ très variable selon les individus (environ 15 dB max)
- Déphasage des signaux parvenant aux 2 oreilles. Une impulsion générée à ma droite arrive d'abord à mon oreille droite, puis à mon oreille gauche. Pour les sons stables et périodiques, cela induit un déphasage entre la voix gauche et la voix droite (se rappeler que la longueur d'onde à 1 *kHz* est de 34 *cm*).
 - ▶ reproductible pour $f < 1500$ *Hz* (comme $\lambda = c/f$, on obtient $\lambda \simeq 22$ *cm*, qui correspond à la largeur de la tête)

Perception de l'espace

2. **Critères binauraux (suite)** : ils regroupent tous les indices qui impliquent les 2 oreilles :
- Différence de temps
 - ▶ Une impulsion générée à ma droite arrive d'abord à mon oreille droite, puis à mon oreille gauche.
 - ▶ avec un diamètre de la tête de l'ordre de 21 *cm* et une vitesse du son de $c_s = 343 \text{ m/s}$, on a au maximum : 0.6 *ms*
 - On bouge la tête pour améliorer la détection (action importante)

Perception de l'espace

3. **Les effets de salle** : les critères précédents ne sont pas suffisants. En fait, dans une salle, on arrive à distinguer :

- le son direct
- les premiers échos
- les réflexions tardives

Les durées et les amplitudes respectives de tous ces sons sont des critères qui nous aident à juger de la proximité ou de l'éloignement de la source. Et nous obtenons aussi des informations sur la salle.

4. **Les critères de mouvement** :

- le mouvement d'une source sonore ou du récepteur entraîne une signature acoustique caractéristique : l'effet Doppler.

Conclusion à la psychoacoustique

Page intentionnellement laissée blanche

Conclusion à la psychoacoustique

- Il faudrait tenir compte de l'organisation que le cerveau fait de toutes ces informations : l'ouïe n'est pas seule ; en réalité on n'est jamais hors contexte, comme les expériences décrites ci-dessus voudraient que l'on soit : *l'observateur modifie ce qu'il observe*. Ceci a deux sens :

1. premier sens (incertitude) :

plus on observe finement la physique, moins on sait ce qui se passe en contexte ;

plus on s'intéresse au comportement en contexte, moins on sait ce qui se passe physiquement.

Conclusion

2. second sens :

un auditeur ayant été observé une fois ne peut plus, ou presque, être observé : il apprend... ainsi, vous-mêmes... vous êtes plutôt prêts pour observer, à présent...

⇒ tests spécifiques pour tromper les musiciens ; pour vous tromper vous (pour nous tromper nous, physiciens et psychoacousticiens)

Conclusion

- **Exemples :**

1. conversation dans la rue, avec bruit : on utilise le contexte (lecture sur les lèvres, posture, etc.), et on filtre (???) adaptativement
2. le morse : le rapport signal sur bruit pouvait être excessivement bas : rien n'est audible pour un profane... et pourtant... : importance de l'apprentissage

- **Échantillonnage** : a priori, 44.1 kHz est suffisant, puisque l'oreille n'entend pas au-delà de 20 kHz. Cependant, il est question de passer à 96 kHz. Des phénomènes autres que temporels/fréquentiels sont en jeu, notamment pour les attaques. Et les attaques sont importantes pour la reconnaissance du timbre.

Conclusion

- Emploi ambigu d'« Acoustique Physique » ? Deux sens à « Physique » :
 - ▶ relatif aux propriétés générales de la matière
 - ▶ relatif au corps humain
- La psychoacoustique (la psychophysique en général) est un domaine extrêmement vaste :
 - ⇒ seuls quelques résultats sont disponibles
 - ⇒ seuls quelques-uns d'entre eux sont présentés ici
- Liens entre la mécanique des solides et des fluides, la physique des matériaux, la chimie, l'électronique, l'architecture, la parole, la musique, la psychologie/cognition, la médecine, les sciences de la Terre et de l'atmosphère, etc.

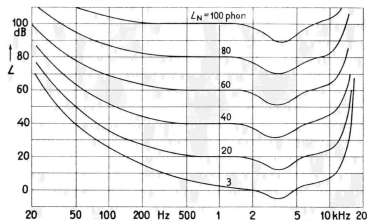
Conclusion – Critères de qualité auditive

- ▶ restitution du niveau/dynamique respectée
 - ▶ de 0 à 100 dB
- ▶ restitution des fréquences
 - ▶ de 16 Hz à 16 kHz (au moins)
 - ▶ et sans << coloration >>, pour respecter le timbre
- ▶ localisation spatiale
 - ▶ systèmes multi-haut-parleurs/multi-capteurs
 - ▶ systèmes holophoniques

- ▶ Les stimulus : le son
- ▶ Les capteurs humains (entre autres)
- ▶ Les sensations auditives
- ▶ Modèles de perception

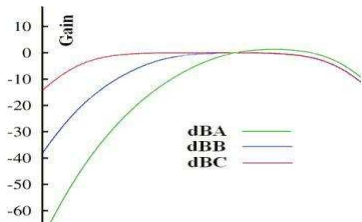
Modèles de sonie

- ▶ les courbes d'isononie ne sont pas simplement translattées
- ▶ de ce fait, un son enregistré ou transmis au niveau L dB ne sera pas perçu de la même façon s'il est rejoué au niveau $L - x$ dB ou au niveau $L + y$ dB
- ▶ l'écart entre deux courbes d'isononie est plus faible aux basses fréquences et aux fréquences élevées qu'aux fréquences intermédiaires (là où l'oreille est le plus sensible)



Modèles de sonie

- ▶ Il faut donc essayer de compenser ça en utilisant des correcteurs différents pour chaque niveau de restitution :
 - ▶ pondération A : faibles niveaux (de 25 *dB* à 55 *dB*)
 - ▶ pondération B : niveaux de 55 *dB* à 85 *dB* (parole et musique standard)
 - ▶ pondération C : niveaux supérieurs à 85 *dB*



Utilisation des modèles de sonie

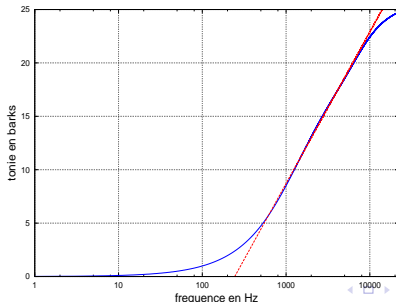
- ▶ correcteur physiologique ; appelé << loudness >> (attention : c'est l'inverse des pondérations décrites ci-dessus) ; présent sur quelques amplificateurs HiFi
 - ▶ surtout utile pour des restitutions à bas niveau
- ▶ législation imposant les niveaux max
 - ▶ nouvelles constructions (note : importance des simulations)
 - ▶ locaux déjà construits (modification/classification)
 - ▶ niveaux utilisés par les radios/télévisions (par exemple, au moment des pubs)
- ▶ mesures des bruits (dBA, dBB, dBC décrits ci-dessus ; et il y en a d'autres, spécifiques aux corps de métiers)

Modèles de tonie

Zwicker et Terhardt (1980) :

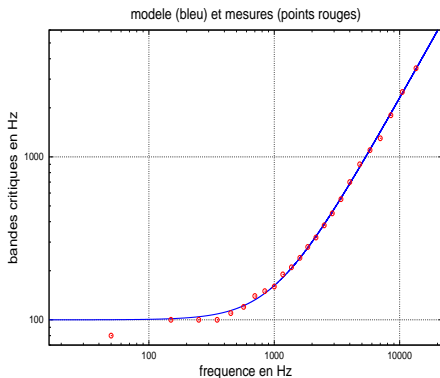
$$B = 13 \arctan \left(\frac{0.76f}{1000} \right) + 3.5 \arctan \left(\left(\frac{f}{7500} \right)^2 \right)$$

$$B = 8.7 + 14.2 \log_{10} \left(\frac{f}{1000} \right)$$

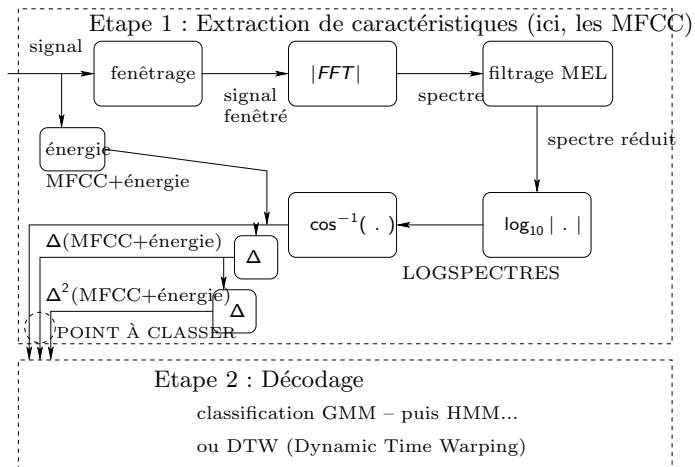


Modèles de bandes critiques

On utilise :
$$\Delta f = 25 + 75 \left(1 + 1.4 \left(\frac{f}{1000} \right)^2 \right)^{0.69}$$



Traitement de la parole et Mels – 1/3



Traitement de la parole et Mels – 2/3

- ▶ formule de conversion F (domaine MEL) vers f (domaine fréquentiel) :

$$F = 2595 \log_{10} \left(\frac{f}{700} + 1 \right)$$

- ▶ note : formule inverse aisée à trouver
- ▶ domaine MEL : filtres linéairement espacés et de largeurs égales ; bornes d'un filtre = fréquences centrales des 2 filtres adjacents
- ▶ domaine fréquentiel : filtres de forme triangulaire
- ▶ soit N_f le nombre de filtres considérés, F_m (16 Hz) la plus petite fréquence MEL considérée et F_M (16 kHz) la plus grande ; largeur de chaque filtre dans le domaine MEL : $W = 2 \frac{F_M - F_m}{N_f + 1}$

Traitement de la parole et Mels – 3/3

- ▶ alors, les trois fréquences importantes, c'est-à-dire celles des deux bornes et la fréquence centrale, sont égales à (filtre de numéro d'ordre i) :

$$F_c^{(i)} = F_m + i \frac{W}{2}$$

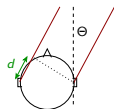
$$F_m^{(i)} = F_m + (i - 1) \frac{W}{2}$$

$$F_M^{(i)} = F_m + (i - 1) \frac{W}{2} + W$$

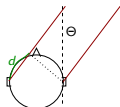
- ▶ et il est aisé de trouver les $f_m^{(i)}$, $f_c^{(i)}$ et $f_M^{(i)}$ correspondant

Modèle binaural – Calcul de la différence de marche

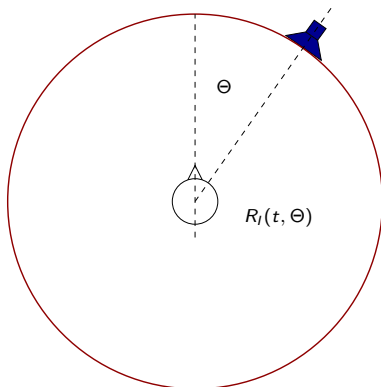
- ▶ modèle simple : les oreilles sont chacune un point, séparées de D ; la différence de marche d est alors $d = D \sin(\Theta)$ ($\Theta = \pi/2$ donne $d = D$)



- ▶ modèle élaboré : la tête est une sphère et les ondes sont diffractées par elle (elles tournent autour de la tête) ; la différence de marche est un peu allongée : $d \simeq D\Theta$ ($\Theta = \pi/2$ donne $d = D\pi/2$)



Modèle binaural – Head Relative Transfer Functions



Modèle binaural – Head Relative Transfer Functions

- ▶ mesure des HRTF/HRIR
 - ▶ sur des têtes acoustiques de mesure
 - ▶ qualité/correction du haut-parleur et des micros
 - ▶ les impulsions ne peuvent pas être utilisées :
 - ▶ difficultés physiques
 - ▶ dommages causés à l'oreille
 - ▶ alors des rampes en fréquence, ou des MLS (Maximum Length Sequences ; séquences pseudo-aléatoires au spectre plat, comme l'impulsion) les remplacent

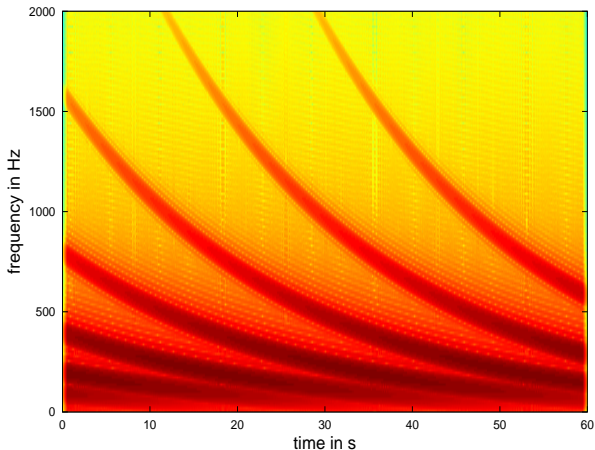
Modèle binaural – Head Relative Transfer Functions

- ▶ traduction en fonctions de transfert
 - ▶ pour chaque valeur de Θ
 - ▶ combien de Θ utiliser ? \Rightarrow interpolation possible
 - ▶ on obtient les filtres à appliquer pour donner l'illusion à un auditeur qu'une source vient de tel ou tel endroit
- ▶ problème : dans l'absolu, chaque personne est différente
 - ▶ interpolation possible des facteurs morphologiques d'une tête à l'autre
- ▶ problème : la taille de la base de données

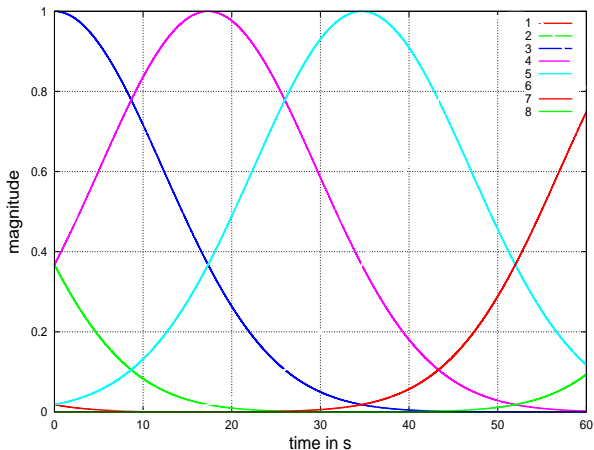
Pour en finir avec le son et sa perception

- ▶ L'IRCAM – Centre Pompidou ; Institut de Recherche et Coordination Acoustique/Musique, Paris (<http://www.ircam.fr>)
 - ▶ Scientifiques + Assistants musicaux + Compositeurs
 - ▶ Équipes scientifiques de l'IRCAM (R et D) : 1. Analyse et synthèse des sons ; 2. Acoustique instrumentale ; 3. Acoustique des salles ; 4. Perception et cognition musicales ; 5. Représentations musicales ; 6. Studio en ligne ; 7. Systèmes temps réel ; 8. Informatique
 - ▶ ⇒ Nombreuses possibilités de stages !!!
 - ▶ Master ATIAM (Acoustique, Traitement de Signal et Informatique Appliqués à la Musique)
 - ▶ Tous les ans (ou presque), des étudiants de Supélec vont à l'IRCAM : majeure plutôt bien considérée
- ▶ Parrot : anciens SIR/SIF ; stages (quasi-)tous les ans

Pour en finir avec le son – Qu'en est-il du son de Risset ?



Pour en finir avec le son – Qu'en est-il du son de Risset ?



Modèles de production – Sons

Stéphane Rossignol

2015 – 2016

Modèles de production – le son

Systeme de phonation

Composition d'un haut-parleur

Modèles physiques : système de phonation

Modèles considérés

- ▶ Fonctionnement du système de phonation
- ▶ Composition d'un haut-parleur
- ▶ Modèles physiques : système de phonation

Émission des sons de la parole

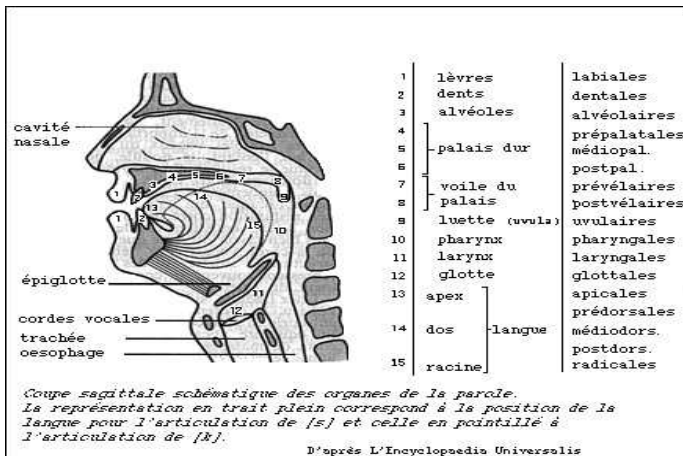
- ▶ sons voisés
 - ▶ périodicité \Leftrightarrow hauteur f_0 ; qualitativement, on a :
 - ▶ basse : f_0 autour de 100 Hz ; voix d'homme
 - ▶ haute : f_0 autour de 250 Hz ; voix de femme
 - ▶ très haute : f_0 entre 400 Hz et 600 Hz ; voix d'enfant
 - ▶ voyelles, certaines consonnes
- ▶ sons non voisés
 - ▶ sifflements, chuintements, ... : /s/, /ch/, /f/
- ▶ plosives ou occlusives
 - ▶ /p/, /t/, /k/, /b/, /d/, /g/

Les sons voisés

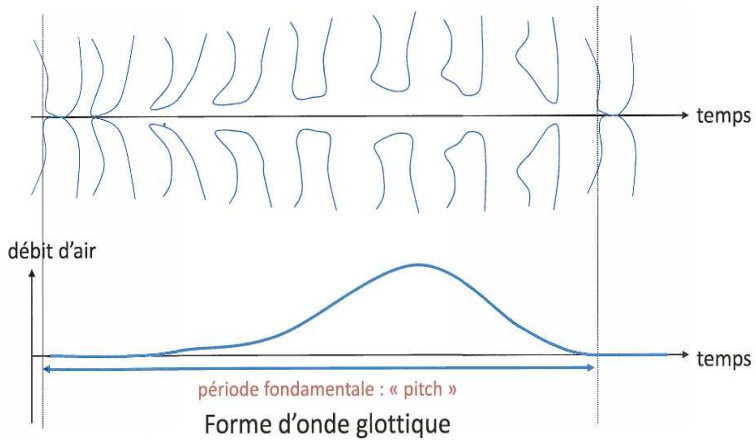
Dans ce cas, on a mise en vibration des cordes vocales

- ▶ périodicité \Leftrightarrow hauteur \Leftrightarrow fréquence fondamentale $f_0 \Leftrightarrow$ le PITCH
- ▶ mélodie (de la voix parlée !) : variation du pitch
- ▶ les sons concernés sont :
 - ▶ voyelles /a/, /e/, /i/...; y compris les voyelles nasales : /an/, /on/ (ayant des anti-formants)...
 - ▶ consonnes /v/, /z/, /j/ : voisement + bruit
 - ▶ consonnes /m/, /n/ : couplage acoustique nasal

Le conduit vocal

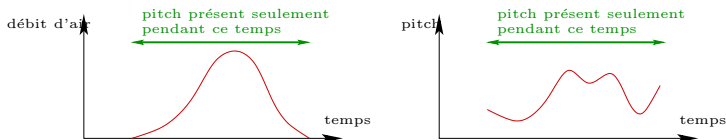


Cordes vocales



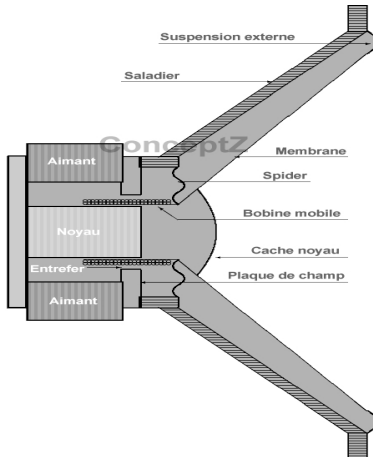
L'onde glottique

- ▶ variabilité de la << période >> et de la forme d'onde : c'est ce qui donne son aspect naturel à la voix

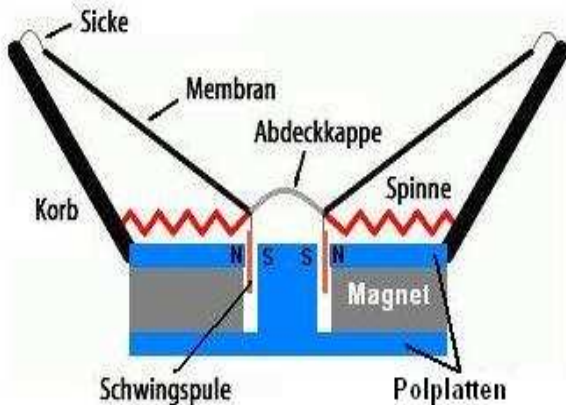


- ▶ du coup, cette variabilité pose forcément des problèmes quand on veut détecter le fondamental : la pseudo-stationnarité du signal est encore plus forte qu'on pouvait s'y attendre
- ▶ modélisation de l'excitation du conduit vocal : il existe des dictionnaires de formes d'onde

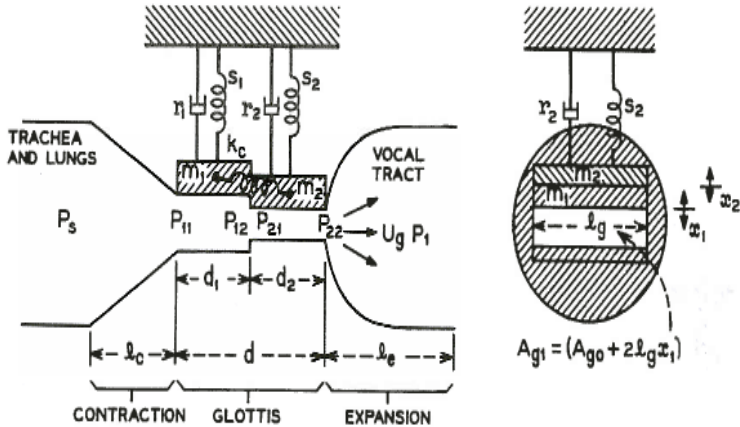
Composition d'un haut-parleur



Composition d'un haut-parleur



Modèle physique du système de phonation



Modèle physique du système de phonation

- ▶ source : Synthesis of Voiced Sounds From a Two-Mass Model of the Vocal Cords, Ishizaka and Flanagan
- ▶ après, il faut écrire toutes les équations différentielles qui régissent ce système
- ▶ puis intégrer numériquement, avec des conditions initiales différentes pour chaque son voisé
- ▶ puis expliquer le pourquoi
- ▶ *note : une autre approche consiste à faire plutôt des modèles de signaux*

Modèles

Pour produire/reproduire, il y a 3 types de modèles :

- ▶ modèles physiques (on colle à la physique du producteur naturel : les petits ressorts)
- ▶ modèles producteurs (on fait de systèmes qui s'arrangent pour reproduire des stimulus qui collent aux stimulus naturels : les haut-parleurs)
- ▶ modèles de signaux (on colle aux signaux produits par un producteur naturel : les spectres)
 - ▶ bienvenue à la modélisation spectrale
 - ▶ note : ça ne veut pas dire qu'on n'essaie pas de quand même expliquer la physique sous-jacente

Examen

N'oubliez pas l'examen, commun avec celui sur les interfaces innovantes et immersives, en janvier prochain

- ▶ est-ce que vous avez réussi à récupérer les slides ?
- ▶ je préfère ne pas les faire imprimer (même en noir et blanc, 4 slides par page)

Ce cours est un pré-requis pour mon cours de traitement de la parole

fin