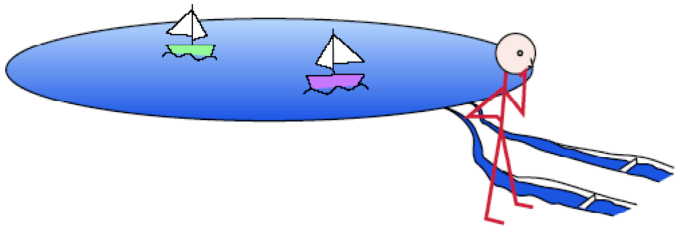
Chapitre 9

L'analyse de la scène auditive

9.1 Intégration et ségrégation de l'information acoustique

Le système auditif traite l'information acoustique pour déterminer la présence, la position et la nature des sources sonores de l'environnement, afin de pouvoir comprendre leur comportement ou les messages qu'elles émettent. Tout cela implique l'organisation perceptive d'un environnement composé de sources multiples, processus que le psychologue Albert Bregman appelle *l'analyse des scènes auditives* (ASA) dans son livre *Auditory scene analysis : The perceptual organization of sound* (MIT Press/Bradford Books, 1990).

Lorsque nous sommes entourés de signaux sonores provenant de différentes sources – par exemple à un concert : un violon qui joue une mélodie, une personne qui tousse, une autre qui déballe une pastille pour la gorge, et des voisins qui chuchotent – ce qui entre dans l'oreille est un patron de vibrations complexes où toutes les sources sont entremêlées. Le rôle du système auditif est alors de déterminer ce qui appartient aux sources sonores potentielles qui nous entourent, et de bâtir une « image » cohérente du monde sonore. Ainsi, au concert, les chuchotements du voisin ne seront jamais entendus comme faisant partie de la mélodie jouée par le violon.

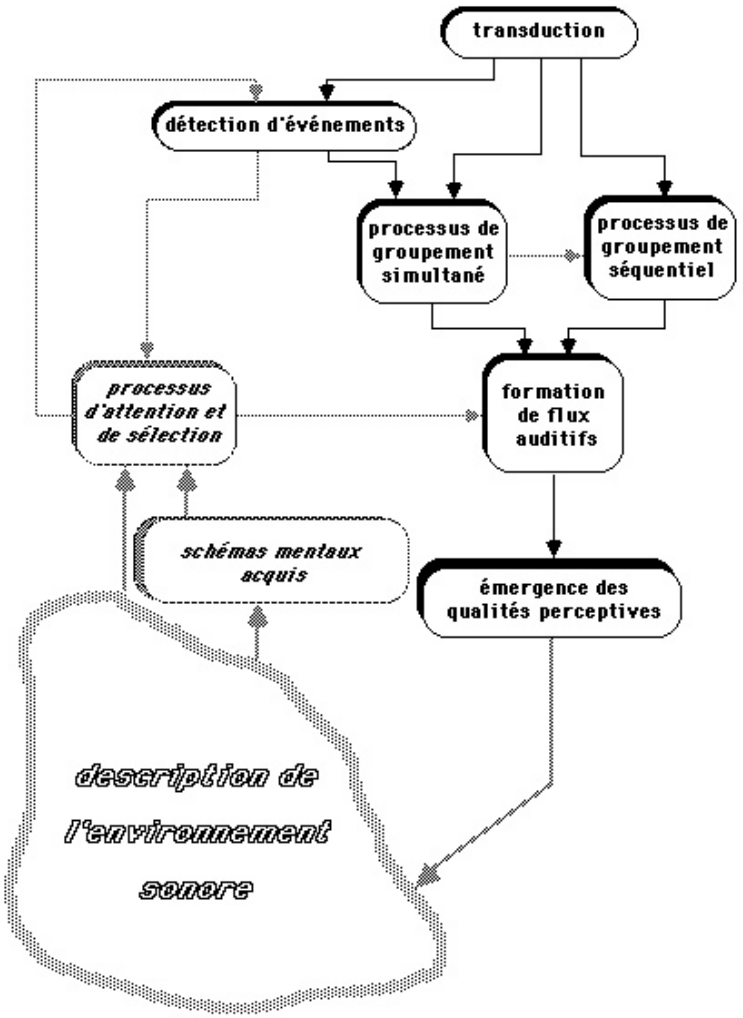


En apparence simple, ce processus de séparation de l'information acoustique en plusieurs flux sonores est extrêmement complexe. Albert Bregman exprimait cette complexité par l'analogie suivante : *Imaginons deux canaux étroits creusés à l'extrémité d'un lac, avec deux petits mouchoirs de poche (représentant les tympans) étirés transversalement à la surface de l'eau sur chacun de ces canaux. En regardant seulement le mouvement des mouchoirs, il faudrait pouvoir répondre à des questions telles que : combien de bateaux y a-t-il sur le lac ? où se trouvent-ils ?* (d'après Bregman, 1990).

On remarquera que l'étude de ces processus n'a suscité l'intérêt des chercheurs en audition qu'à partir des années 1980, alors que le problème semblable avait été identifié depuis longtemps dans le domaine de la vision.

9.2 Les processus d'organisation auditive

L'analyse auditive d'un environnement sonore peut être modélisée comme un processus de traitement d'informations ayant pour but de construire un modèle ou schéma du monde acoustique, tel que l'illustre le modèle proposé par Stephen McAdams[1] :



Ce processus de construction est basé sur :
– des indices d'**assemblage** d'éléments acoustiques **simultanés** en **événements**,
– des indices de **connexion** des éléments acoustiques **séquentiels** en **flux d'événements**,

[1] L'organisation perceptive de l'environnement sonore, Stephen McAdams, Rencontres IPSEN en ORL, 1997.

– la dérivation de qualités perceptives à partir des propriétés des groupes assemblés,
– et peut-être même des schémas stockés reflétant le comportement cohérent du monde physique et la structure de configurations d'événements (ou formes) familières.

La notion d' « **objet auditif** » est importante pour la compréhension des processus d'organisation perceptive dans la modalité auditive. Ce terme se réfère à une représentation mentale d'un groupe d'éléments qui possèdent une **cohérence interne** dans leur comportement et qui sont ainsi interprétés comme provenant de (ou, dans le vocabulaire des psychologues gestaltistes, "appartenant à") la même source sonore.

Ce processus de représentation doit nécessairement permettre non seulement le **groupement d'éléments acoustiques en images sonores simples**, tel un groupe de fréquences rassemblées en une note de clarinette, mais également le **groupement de plusieurs sources sonores physiques en images complexes**, tel que les textures ou timbres composés que l'on trouve dans la musique pour orchestre, ou le groupement d'événements émis à travers le temps par une source sonore, telle une phrase parlée ou une mélodie.

Cette tendance à rassembler les éléments ayant une cohérence structurale en une unité psychologique permet à l'auditeur d'organiser l'environnement sonore en sources qui sont très complexes acoustiquement. Par exemple, des chocs entre morceaux de métal, entre caoutchouc et pierre, et une série périodique d'explosions peuvent être unis en l'image d'une voiture roulant sur les pavés.

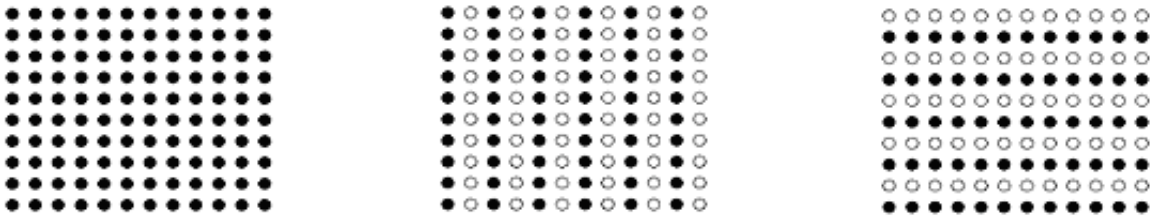
9.3 Les principes de la théorie de la Gestalt

La théorie de la Gestalt propose des « lois » pour rendre compte de la formation des groupements et des configurations. Dans le domaine visuel, le terme de groupement désigne, d'un point de vue phénoménal, le fait que des éléments picturaux paraissent aller ensemble, appartenir à une même unité perceptivement séparée d'autres unités[2].

Parmi les principes de groupement, on trouve : la *proximité* des éléments, leur *similarité*, leur *bonne continuation*, et leur *connexité*.

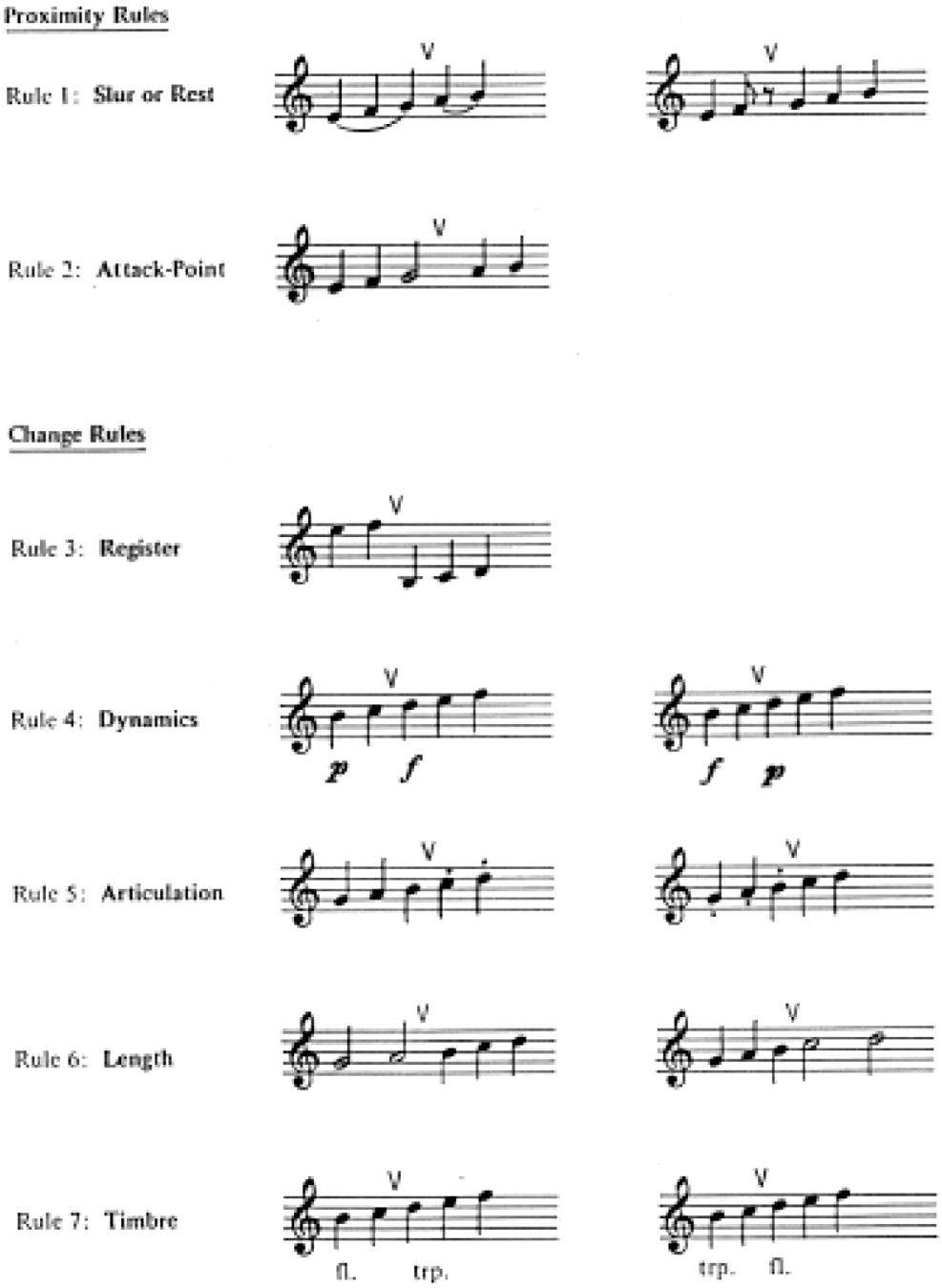
Par exemple, la série de points consécutifs suivante sera perçue comme étant un enchaînement de série de deux points. C'est la loi de groupement par proximité qui s'applique dans ce cas.

Les figures suivantes illustrent, quand à elles, la loi de groupement par similarité.



[2]Traité de psychologie cognitive, Bonnet, 1989.

Certaines de ces lois ont d'ailleurs été appliquées à l'étude des principes de segmentation des phrases musicales dans la musique tonale, par Fred Lerdhal et Ray Jackendoff dans leur ouvrage *A Generative Theory of Tonal Music* (1983), ouvrage qui eut un impact important dans le domaine de la psychologie de la musique et de la théorie musicale.



9.4 Indices physiques exploités pour l'ASA

Les indices physiques que le système auditif (dans son entier, et en particulier les centres plus élevés) prend en compte pour analyser une scène auditive sont répertoriés et classés par Yost (dans *Fundamentals of Hearing : An Introduction*, 2000) de la manière suivante :

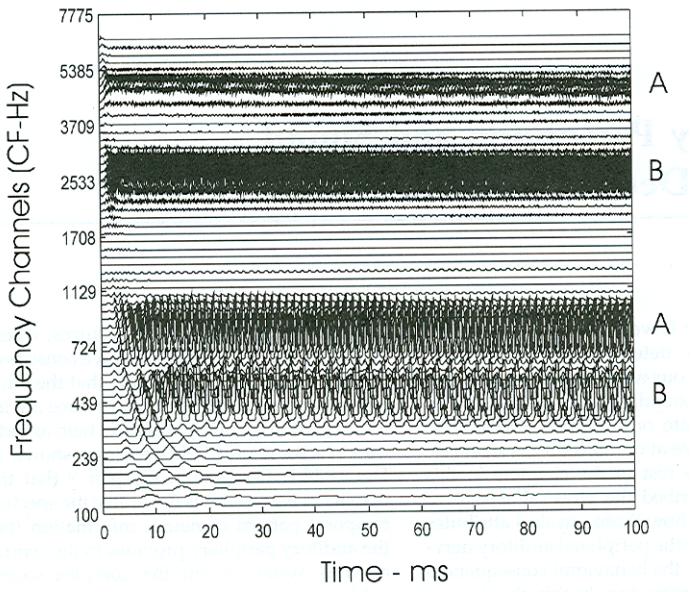
1. la séparation spectrale
2. le profil spectral
3. l'harmonicité
4. la séparation spatiale (position commune dans l'espace)
5. la séparation temporelle
6. la synchronisation des attaques et des chutes
7. les modulations (d'amplitude et de fréquence)

9.4.1 La séparation spectrale

La capacité de filtrer du système auditif («découpé» en bandes critiques) aide à distinguer une source sonore d'une autre, lorsque les sources sont différentes spectralement. Par contre, lorsque des bandes de fréquences se chevauchent, la séparation spectrale n'est pas suffisante pour différencier deux sources. Sur la figure ci-dessous est représentée la sortie des 64 filtres auditifs du modèle de Patterson qui simule le traitement périphérique d'un son complexe composé de 4 sons purs appartenant à deux sources différentes :

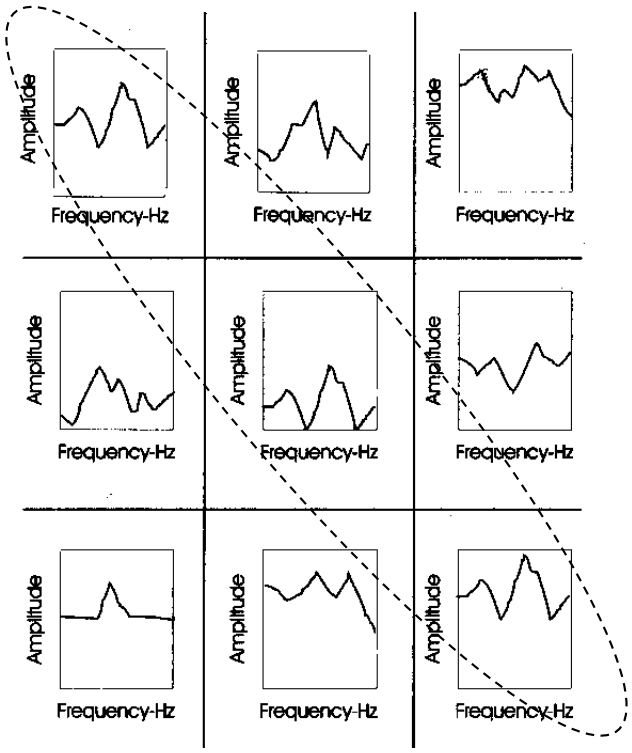
SOURCE A : 725 Hz et 5 000 Hz SOURCE B : 450 Hz et 2 500 Hz

Les composantes des deux sources se chevauchent. On remarque l'étalement de l'activité neuronale autour du site de déplacement maximal. Même si l'information temporelle et spectrale est préservée au niveau périphérique, il y a peu d'information sur cette figure qui indique à quelles sources appartiennent les composantes.



9.4.2 Le profil spectral ou enveloppe spectrale

La figure ci-dessous présente le spectre de sons complexes provenant, a priori, de différentes sources.



On peut cependant facilement repérer que les spectres le long de la diagonale du coin gauche supérieur au coin droit inférieur appartiennent à la même source, puisque globalement, l'enveloppe spectrale est préservée. Ces trois spectres diffèrent seulement sur le plan de l'amplitude globale. Les autres spectres appartiennent bien à des sources différentes.

Pour regrouper les composantes d'une même source, le système auditif va suivre les différences relatives d'amplitude des composantes spectrales au cours d'un changement de niveau d'intensité global. Si les différences relatives d'amplitude sont maintenues constantes, le système auditif aura tendance à regrouper ces composantes en un objet auditif unique.

9.4.3 L'harmonicité

De façon générale, le fait que les composantes spectrales du son soient des multiples entiers d'une fondamentale (et ont donc une fondamentale commune) aide à percevoir ces fréquences comme faisant partie d'une même source sonore. La perception de la hauteur en est un bel exemple : toutes les harmoniques fusionnent en une seule hauteur. Si une composante s'écarte de la grille harmonique (et n'est donc plus en relation de multiples entiers), ce partiel a de fortes chances d'être perçu comme un son distinct par rapport au son complexe harmonique.

Afin d'illustrer l'effet de cet indice, nous allons décrire une démonstration présentée sur le CD réalisé par Albert S. Bregman et Pierre A. Ahad : *Demonstrations of Auditory Scene Analysis.*

The Perceptual Organization of Sound (disponible à la bibliothèque).

Démonstration : A. Bregman n° 18

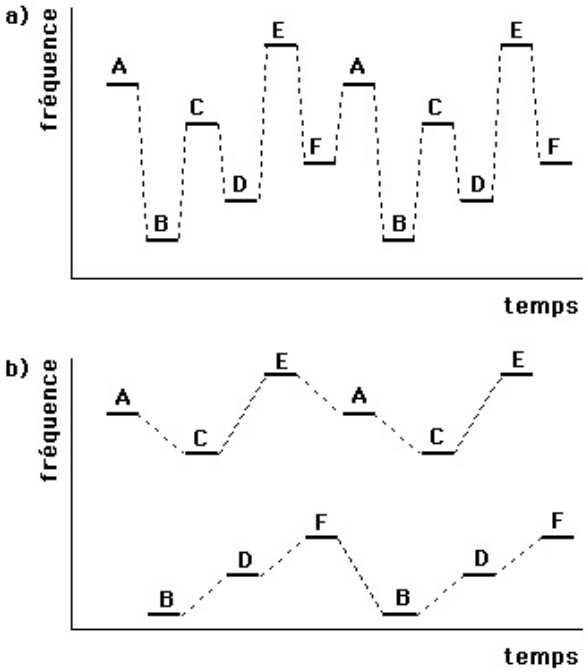
Isolation d'une composante fréquentielle (3^{e} harmonique) due à un désaccordage (*mistuning*) par rapport aux harmoniques d'un son complexe.

9.4.4 La séparation spatiale et application à l'effet du *cocktail party*

L'effet de *cocktail party* réfère à l'habileté du système auditif de déterminer les sources sonores lorsqu'elles sont localisées à différents endroits dans l'espace. Une situation typique serait une fête où nous serions entourés d'un grand nombre de personnes en train de discuter. En focalisant notre attention sur une conversation, nous sommes capables de l'isoler du mélange complexe résultant de la superposition de toutes les voix en présence. La séparation spatiale peut aider à effectuer cette tâche de ségrégation des flux mais précisons qu'elle est un indice faible en soi pour séparer des sources sonores réelles. D'autres indices sont plus efficaces, comme la synchronisation des partiels par exemple.

9.4.5 La séparation temporelle

La fusion et la ségrégation/formation/séparation des flux (*fusion and stream segregation*) a été illustrée par Albert Bregman à l'aide de plusieurs démonstrations sonores. Par exemple, dans certaines conditions, deux sons qui alternent (un de haute fréquence et un de basse fréquence) vont créer la perception de la présence de deux sources — c'est la *ségrégation des flux auditifs* — et dans d'autres conditions, une seule source sera perçue — c'est la *fusion des flux auditifs.*



Il faut bien comprendre que ces situations peuvent être créées de toutes pièces, c'est-à-dire que l'on peut tromper le système auditif pour imiter les conditions acoustiques réelles dans lesquelles les auditeurs déduiront la présence d'une source ou de deux sources.

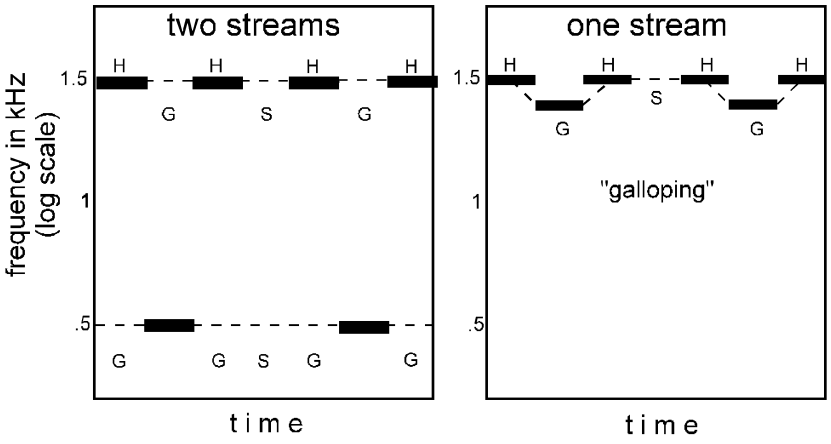
Par exemple, deux sons qui diffèrent par le contenu spectral, le patron temporel de modulation, le niveau, la localisation spatiale, seront plus souvent perçus comme deux sources sonores simultanées mais distinctes. Le contenu spectral est un indice déterminant pour produire une telle ségrégation.

Démonstration : A. Bregman n° 1

Dans cette démonstration, c'est l'augmentation de la vitesse de présentation des stimuli qui cause la ségrégation.

Démonstration : A. Bregman n° 3

Cette démonstration, similaire à la précédente, illustre la *perte de l'information rythmique en conséquence de la ségrégation des flux*. Des séquences de trois sons groupés pour former un rythme « galopant » (son aigu – son grave – son aigu) sont présentés de plus en plus rapidement.



À basse vitesse, le galop est clairement perçu (121 – 121 – 121 – 121 – ...). Mais à partir d'une certaine vitesse, le galop se sépare en deux flux distincts : un premier flux constitué des sons aigus (1 – 1 – 1 – 1 – ...) et un second flux constitué des sons graves (2 – 2 – 2 – 2 – ...). Si on diminue la distance fréquentielle entre les deux sons, la ségrégation se produit encore mais à une vitesse plus importante. On en conclut qu'*une plus grande séparation fréquentielle facilite la ségrégation en des flux distincts*. Ce principe de ségrégation des flux est exploité en musique instrumentale, comme par exemple, dans certaines sonates de Telemann jouées à la flûte donnant l'impression que le flûtiste joue plusieurs voix à la fois (polyphonie virtuelle).

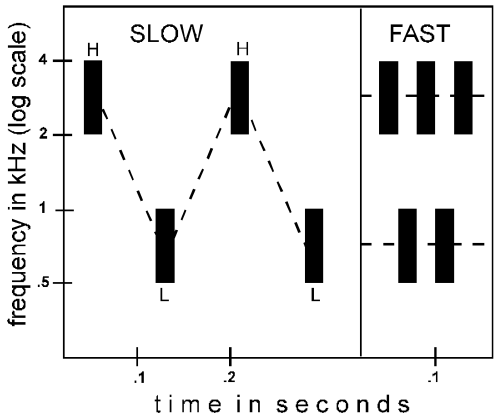
Démonstration : A. Bregman n° 5

Dans cette démonstration, les sons d'une mélodie alternent d'abord avec des sons de hauteurs aléatoires issues du même registre. Ensuite, en transposant la mélodie, on permet à celle-ci

d'émerger des sons interférants pour constituer un flux distinct. La mélodie peut alors être reconnue. Il observe ainsi de la *ségrégation de sons d'une mélodie et de sons interférants* et l'influence sur cette ségrégation de la séparation fréquentielle entre la mélodie et les sons interférants.

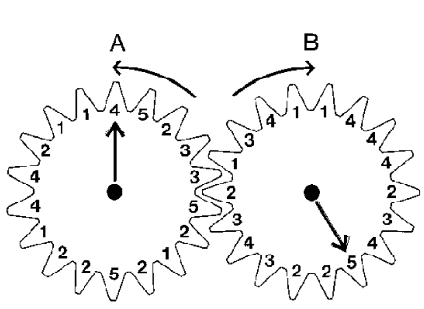
Démonstration : A. Bregman n° 14

Dans cette démonstration, on présente en alternance deux bandes de bruit dans les basses et les hautes de fréquences respectivement. Si on augmente la vitesse (en diminuant la durée des silences entre les sons), on facilite ici encore la ségrégation en deux flux distincts.

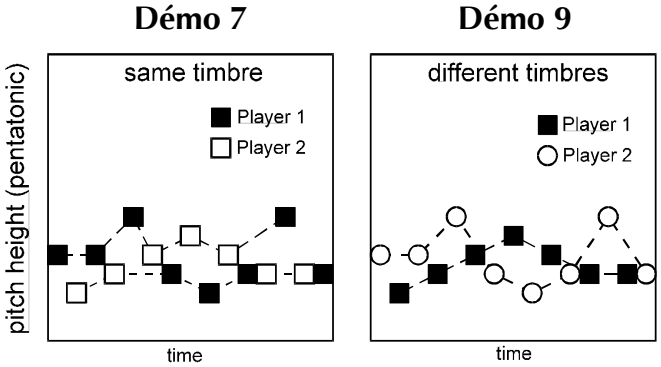


Démonstration : A. Bregman no 7, 8, 9

Dans ces démonstrations, on entend une musique africaine jouée au xylophone par deux percussionnistes produisant des motifs mélodiques qui s'entrelacent dans le temps. Quand les timbres sont identiques, les deux voix fusionnent en une seule. Quand les timbres sont différents, les deux voix sont bien distinctes (démonstration no 9). Quand les deux voix sont séparées d'une octave (séparation fréquentielle), la ségrégation en deux flux distincts est également facilitée (démonstration no 8).



Les notes alternent d'un motif rythmique à l'autre



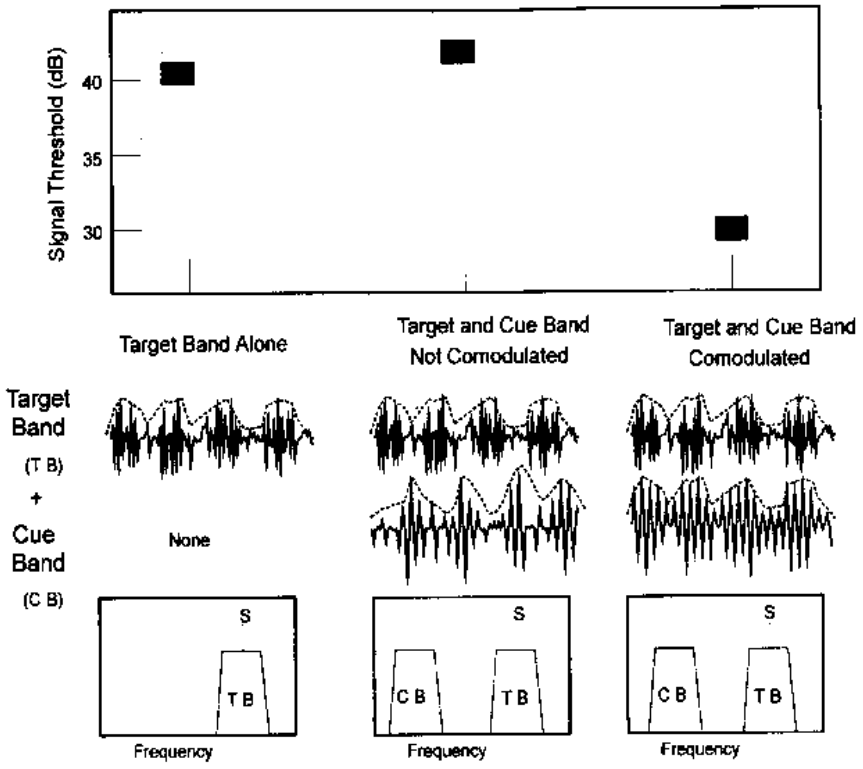
9.4.6 La synchronisation des attaques (*onsets*) et chutes (*offsets*)

Helmholtz a remarqué qu'en général, les composantes spectrales d'un événement émanant d'une même source tendent à commencer en même temps, à varier ensemble et à chuter simultanément. Ainsi, si deux sons ne démarrent pas ni ne s'arrêtent en même temps, il y a plus de chances qu'ils soient perçus comme provenant de sources sonores différentes. Mais on peut aussi tromper le système auditif, en orchestration, où on peut donner l'illusion d'un instrument hybride résultant de la superposition rigoureusement synchrone de deux sons instrumentaux différents.

Un décalage d'attaque de l'ordre de 30 ms entre deux sons simultanés est accompagné d'une *baisse de la fusion perceptive* des deux sons et d'une augmentation de l'identification de chacun. Un décalage de 40 à 80 ms suffit pour produire une *impression de dédoublement* de la source et pour briser une perception catégorielle résultant du groupement des deux éléments décalés. Dans le cas de la perception des voyelles, les asynchronismes d'attaque et de chute d'un seul harmonique, critique pour l'évaluation du premier formant, affectent la frontière catégorielle entre deux voyelles lorsque ces asynchronismes dépassent 30 ms.

9.4.7 Les modulations

La plupart des sons de la nature sont modulés lentement en amplitude et en fréquence. Ainsi, le système auditif fusionne les composantes qui varient (sont modulées) au même rythme, les distinguant du reste. Ce phénomène est illustré par le *relâchement du masquage par comodulation* (comodulation masking release) dont un exemple est représenté sur la figure suivante.



Ce phénomène se produit quand on masque un signal sinusoïdal avec deux bandes de bruits étroites situées dans des régions de fréquences différentes (par exemple 100 Hz et 6 000 Hz) et qui sont modulées en amplitude soit au même rythme (*cohérence*), soit à des rythmes différents (*incohérence*).

La figure ci-dessus donne le seuil de détection d'un signal en présence de deux bandes de bruit masquantes : la bande cible (centrée sur le signal) – Target band (TB) et la bande indice (située dans les basses fréquences) – Cue Band (CB).

- En présence de la bande cible TB, le masquage est important et le seuil est à 41 dB.
- En présence des deux bandes TB et CB non comodulées, le seuil augmente légèrement.
- En présence des deux bandes TB et CB comodulées (les enveloppes temporelles sont alors cohérentes), on arrive à un résultat étonnant : le seuil du signal baisse d'environ 12 dB !

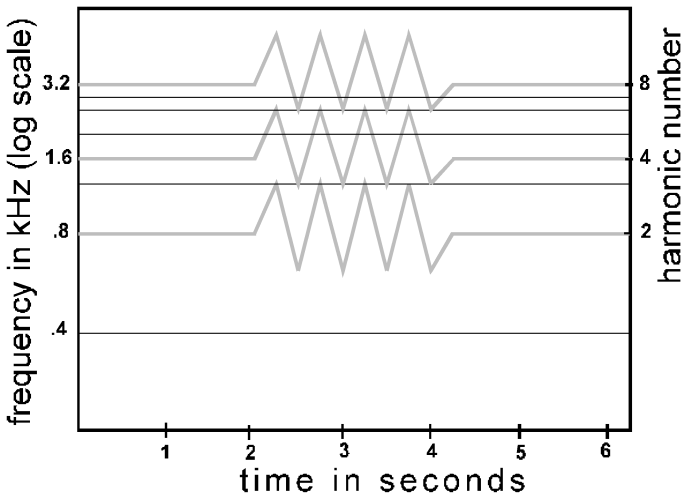
On en conclut que la *comodulation des masques facilite la détection du signal*, ce qui se traduit par un abaissement du seuil de masquage.

Modulations de fréquence

Démonstration : A. Bregman n° 19

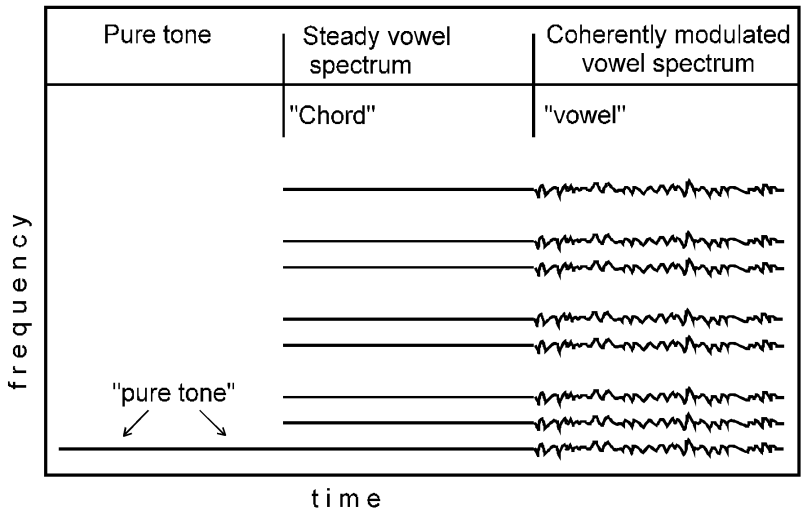
Ceci est un exemple de groupement des composantes fréquentielles dû à des changements parallèles de fréquence.

Dans un premier temps, toutes les harmoniques sont jouées avec des fréquences stables. On entend alors un son unique. Ensuite, les fréquences des harmoniques 1, 3, 5, 6 et 7 sont maintenues constantes tandis que les harmoniques 2, 4 et 8 sont modulées en fréquence, montant et descendant quatre fois. À ce moment-là, les deux ensembles d'harmoniques sont entendus séparément. On peut avoir l'impression que le son stable a changé de timbre alors qu'un autre son modulé vient s'ajouter au premier. Finalement, quand les partiels sont tous à nouveau stables, on réentend le timbre unique de départ.



Démonstration : A. Bregman n° 24

Cette démonstration illustre le rôle des micromodulations de fréquence dans la perception de la voix. D'abord, on entend un son pur correspondant à la fondamentale. Puis, d'autres sons sinusoïdaux s'ajoutent formant en principe, avec le premier son présenté, un son complexe évoquant un son vocal. Cependant, ce n'est qu'à partir du moment où des micromodulations cohérentes sont appliquées à ces composantes que la fusion perceptive s'opère.



9.5 Les processus de groupement – Synthèse

Dans le but de faire la synthèse des différentes processus de groupement que nous avons explorés, nous rappellerons tout d'abord qu'il existe deux types de processus de groupement : les processus de groupement **simultané** et les processus de groupement **séquentiel**.

9.5.1 Les processus de groupement simultané

Le groupement simultané sert à rassembler les informations concourantes analysées par le système auditif périphérique qui proviennent de la même source sonore, et à séparer les informations provenant de sources distinctes. Ce groupement semble s'effectuer sur la base d'un petit ensemble d'indices de la cohérence de comportement d'un événement, qui sont pour les plus importants :

- le synchronisme des attaques et des chutes,
- la cohérence de la modulation d'amplitude,
- l'harmonicité commune et la séparation des fréquences fondamentales,
- la position commune dans l'espace.

9.5.2 Les processus de groupement séquentiel

Le groupement séquentiel sert à affecter les événements successifs qui présentent une certaine cohérence entre eux à des représentations mentales du comportement temporel des sources so-

nores. Ce groupement semble s'effectuer sur la base de la continuité des événements au sein d'un flux sur le plan du contenu spectral et de l'intensité, ce qui représente la cohérence de comportement d'une source sonore en termes d'une certaine inertie dans le changement de ses propriétés acoustiques au cours du temps. Les indices de groupement séquentiel sont entre autres :

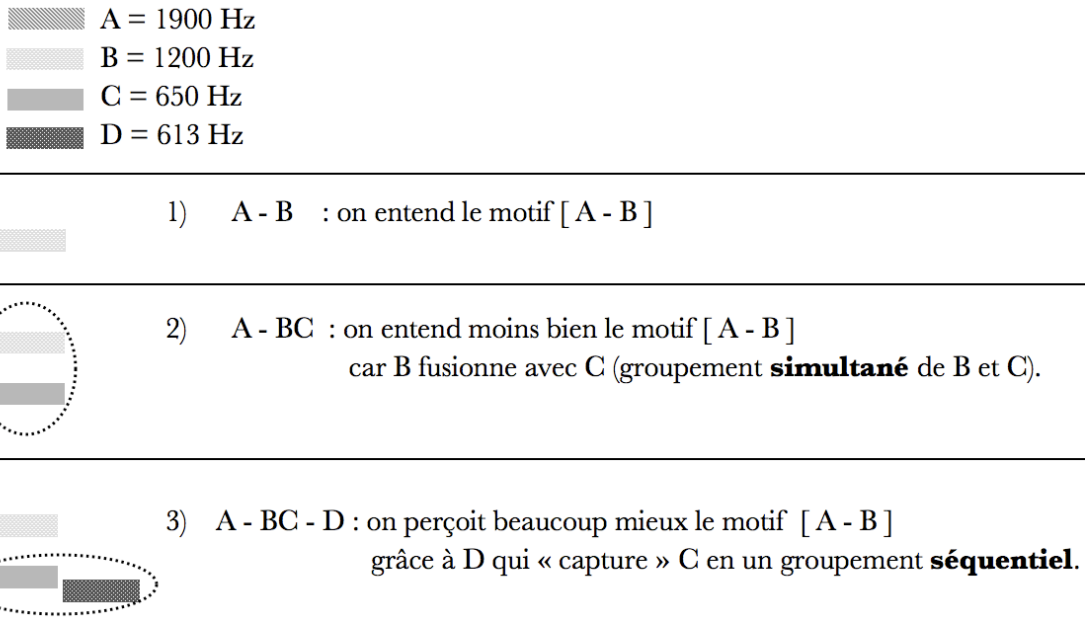
- Le contenu fréquentiel
- L'enveloppe spectrale
- L'intensité

9.5.3 Compétition des groupements séquentiels et simultanés

Dans certains cas, les processus des groupements séquentiel et simultané peuvent être en compétition. Ceci est bien illustré par la démonstration suivante.

Démonstration : A. Bregman n° 27

Cette démonstration montre qu'un changement dans la force d'un groupement en une position spectro-temporelle donnée peut altérer la force de groupement en d'autres positions. Quatre sons sinusoïdaux (de fréquence 1900, 1200, 650 et 613 Hz respectivement) sont groupés différemment en fonction de leur arrangement simultané et séquentiel.



Pour conclure ce chapitre, nous vous recommandons de consulter l'excellent article de David Huron intitulé *Tone and Voice : A Derivation of the Rules of Voice-Leading from Perceptual Principles* (Music Perception, Vol. 19, No. 1, pp. 1-64, 2001) dans lequel l'auteur dérive les règles de l'harmonie et de la conduite des voix dans la musique tonale sur la base de principes perceptifs. L'article est disponible en version html à l'adresse : http ://music-cog.ohio-state.edu/Huron/Publications/huron.voice.leading.html.