CHAPITRE 4. INTELLIGIBILITE

4.1. Introduction

Un des paramètres fonctionnels principaux des langues sifflées est leur intelligibilité. Elles répondent au besoin d'assurer vite et bien une communication à distance dans des conditions où le bruit de fond de la nature n'est pas négligeable. Tous les témoins des langues sifflés ont été très étonnés de l'efficacité des systèmes qu'ils observaient. Ainsi Cowan, dans son premier papier de linguistique sur ce thème, relate: *« Un jour, Eusebio Martinez, debout devant sa maison, siffla en direction d'un homme qui était à une distance considérable. Il passait sur le chemin en contrebas pour aller y vendre des paquets de feuilles de maïs qu'il portait. L'homme répondit à Eusebio par un sifflement. L'échange fut répété plusieurs fois avec différents sifflements. Finalement l'homme fit demi-tour, revint sur ses pas et monta sur le sentier qui venait à la maison d'Eusebio. Sans dire un mot, il laissa tomber sa charge au sol. Eusebio verifia la charge, rentra dans sa maison, ressortit avec de l'argent et le paya. Celui ci fit demi-tour et s'en alla. Pas un mot n'avait été dit. Ils avaient parlé, marchandé le prix et étaient arrivés à un accord satisfaisant pour les deux en utilisant seulement le sifflement comme mode de communication*[79] *»* (Cowan 1948, p.280, traduction libre). Le succès d'une telle communication tient à un grand nombre de paramètres: la motivation de l'émetteur et du récepteur, l'efficacité de l'encodage linguistique, la propagation du signal dans l'air, la manière dont le signal sera noyé dans le bruit de fond à son arrivée au niveau de l'oreille du récepteur, les capacités perceptives de l'auditeur et bien sûr, de manière cruciale, la connaissance du vocabulaire et du système de règles de la langue employée.

Si les langues sifflées surprennent ce sont indirectement les capacités du cerveau humain et de l'oreille qui étonnent. La principale originalité tient à l'usage d'un signal sifflé qui exploite la redondance de la voix parlée pour sélectionner un *squelette informatif* en fonction de la structure de la langue. Celui ci est opté pour répondre aux contraintes liées aux conditions d'usage qui entraînent une réduction du rapport signal sur bruit et une augmentation de la réverbération. Les mesures que nous avons faites auparavant montrent que la parole sifflée relève d'une stratégie qui est la continuité directe de celles de la voix parlée puis criée.

Comme le sifflement s'appuie sur des éléments phonétiques et phonologiques de la langue locale, les processus cognitifs qui permettent à un siffleur de comprendre le message sifflé sont donc, à notre avis, non seulement similaires à ceux qui interviennent lors de l'écoute d'une langue mais ils impliquent fortement ceux en jeu lors de l'écoute dans des conditions difficiles liées au milieu ambiant.

---

[79] Discussion rapportée dans une publication faite en espagnol par Cowan: A: Qu'as tu donc acheté? B: C'est une charge de maïs? A: Et où l'amène tu? B: Je l'amène à Tenango. A: Tu vas la vendre? B: Je vais la vendre. A: Combien en veux tu? Vends la moi ici. B: Ce sera deux pesos cinquante centavos pour la charge? A: Tu ne veux pas deux pesos vingt cinq centavos? C'est ce que je peux te donner. B: La où je vais la vendre, ils m'en donneront trois pesos. A: Oui mais c'est loin, que décides tu ? B: Je vais laisser ma cargaison ici. A: A la bonne heure, tu demandais beaucoup. (Cowan 1976 in Sebeok et Umiker-Sebeok p.1396, traduction libre).

La parole humaine étudiée dans des conditions idéales d'écoute ne permet pas toujours de rendre compte de ces derniers aspects. Pourtant la voix parlée est quotidiennement utilisée dans le bruit et parfois même à des distances non négligeables pour le signal phonétique qui parvient à l'auditeur. De nombreuses études sur la parole dans le bruit ont montré que le rapport signal sur bruit et la réverbération sont les principaux facteurs qui affectent les performances de reconstuction cognitive de la parole. Compte tenu de ces remarques, l'analyse des langues sifflées est susceptible de fournir un matériau d'étude très riche du fonctionnement cognitif langagier.

La performance de reconnaissance des mots par un auditeur est parfois appelée l'*«intelligibilité de la parole»*. Le but de toute communication linguistique humaine est d'atteindre une intelligibilité maximale avec un effort adapté au contexte, si possible optimisé. Etudier l'intelligibilité au sens général implique l'analyse de l'influence des différents paramètres affectant le taux de reconnaissance des mots d'une phrase. Une telle approche concerne à la fois un locuteur, principalement à travers le signal qu'il emet et sa localisation dans l'espace, un milieu de transmission acoustique appelé canal (pour les langues sifflées, le canal est le plein air dans 99% des cas[80]) et un récepteur qui fait un prétraitement acoustique au niveau de l'oreille puis une reconnaissance et une interprétation au niveau du système central (Figure 77).

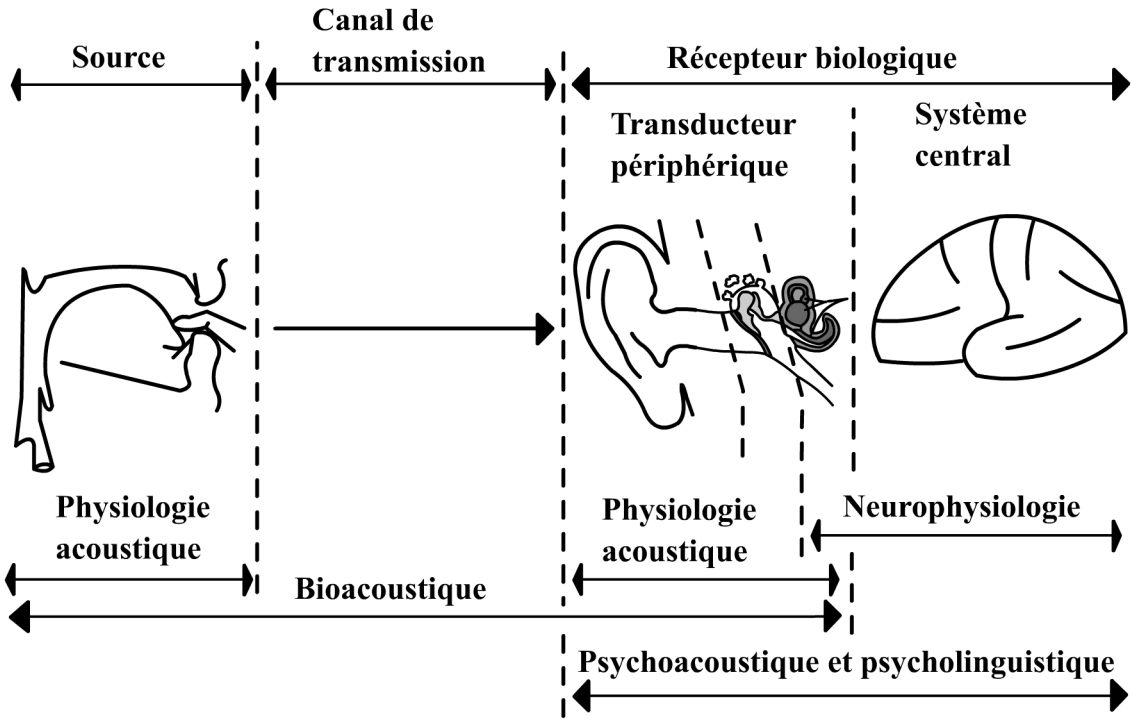

Figure 77 : Schéma des étapes de la transmission et de la reconnaissance d'un message linguistique

[80] Un de nos informateurs Gomero appelle souvent au téléphone son frère résidant sur l'île de Tenerife en utilisant le sifflement par boutade.

Les scientifiques travaillant sur les systèmes de télécommunication ont été les premiers à attirer l'attention sur ce problème de l'intelligibilité (Campbell, 1910, Fletcher & Steinberg 1929, Chavasse 1962). Leurs travaux sur les caractéristiques du message parlé et sur l'adaptation des moyens de transmission aux possibilités de la perception périphérique ont mis en valeur la nécessité de mesurer l'identification correcte des sons de la parole. Ils ont créé la théorie metrologique dont l'usage, entre autres, pour la mesure de la surdité est d'un considérable intérêt. Ils ont perçu la notion différentielle qui doit séparer netteté, d'intelligibilité. L' *« intelligibilité »* est la compréhension des idées, qui se mesure avec des phrases et des mots significatifs d'une langue donnée. Elle diffère de la *« netteté »* qui se réfère, elle, à des associations conventionnelles de sons ou de mots dépourvus de sens (*« logatomes* » et *« non-mots »).*

Mais, pour les ingénieurs, l'intelligibilité est exclusivement relative au courant d'information et à ses variations au travers des systèmes de transformation et de transmission (par exemple en téléphonie: par le microphone, la ligne, l'écouteur) dont la mesure doit être considérée comme constante, bien qu'elle se fasse au travers d'un récepteur humain qui ne devrait théoriquement que « mesurer » ce « courant d'informations » sans pouvoir l'influencer. Pourtant, les propriétés psychophysiologiques du récepteur sont à la base de toute mesure d'intelligibilité. Comme il n'est pas pensable de s'affranchir complètement des variations dues aux jugements subjectifs des individus et aux conditions variables de l'expérience, des protocoles particuliers ont été mis au point afin de normaliser les résultats obtenus dans les études technologiques[81] (pour un bilan sur le sujet voir Cartier 1989).

Dans notre cas, notre but n'est pas d'améliorer l'efficacité d'un canal de transmission (puisque cela a déjà été fait par l'évolution naturelle) mais de comprendre l'influence des différents paramètres intervenant dans la compréhension de l'information linguistique encodée dans les phrases telles qu'elles parviennent à l'oreille du siffleur. La notion d'intelligibilité au travers du recepteur humain et de son système cognitif telle que nous l'envisageons est donc, par essence, différente de celle qui se rapporte au courant d'information mesuré suivant des normes technologiques telle que définie par les ingénieurs de téléphonie.

Dans cette partie, nous verrons dans un premier temps que notre connaissance actuelle du système périphérique auditif définit un cadre dans lequel le sifflement se positionne avantageusement. L'audition testée en psychoacoustique par des sons simples éclaire également certains des choix faits par les siffleurs. A partir du cadre définit par ces résultats expérimentaux préliminaires, nous pourrons aborder les recherches qui permettent de comprendre comment l'être humain organise les sons complexes qu'il recoit. Par exemple,

[81] Les protocoles d'évaluation de l'intelligibilité existant aujourd'hui se distinguent en méthodes directes et indirectes. Les méthodes indirectes font appel au jugement des sujets. Les méthodes directes ont souvent été mises en place grâce à des méthodes indirectes, ce sont des méthodes de calcul des taux de reconnaissance à partir de normes (comme les normes RASTI), d'Indices d'Articulation des mots (ou I.A.) ou de modèles mathématiques. Les normes RASTI qui sont utilisées dans l'industrie restent mal adaptées à certains milieux (Tisseyre, 1998). Les indices d'articulation se basent sur l'acoustique de la voix, ils ne sont donc pas adaptés pour les langues sifflées. Les modèles de probabilité sont nombreux et les plus récents sont bien adaptés à l'écoute dans le bruit (Bronkhorst et al, 1993).

l' « *analyse de la scène auditive »* (Bregman, 1990) explique comment l'oreille humaine réalise une véritable enquête sur l'origine des sons (Risset, 1994). Un certain nombre de regroupements des attributs de la perception en résultent, ils expliquent de nombreux aspects de la perception des sons complexes comme ceux de la musique ou du langage. Ils suggèrent également d'analyser les processus de reconnaissance de la parole à plusieurs échelles. De telles démarches, dont nous signalerons quelques résultats, ont été initiées pour l'étude des langues. *L'analyse de la prosodie* est un des domaines qui a cherché à répondre à cette exigence. Nous verrons pourquoi l'on peut considérer que les langues sifflées redéfinissent la prosodie en proposant une approche tenant compte de la perception.

Après ces considérations générales, nous ferons un bilan progressif de la dynamique de l'intelligibilité des langues sifflées en partant de la perception des voyelles jusqu'à l'intelligibilité des phrases. Certaines étapes de cet exposé apporteront des éléments expérimentaux nouveaux, en particulier sur les voyelles sifflées et sur la dégradation de la structure phonétique avec la distance. Le cadre de notre étude rapellera en quoi elle peut être instructive pour l'ensemble des langues et ne doit pas être confinée à la curiosité exotique.

4.2. Adaptation étonnante du sifflement à l'audition humaine

Les bioacousticiens ont observé sur de nombreuses espèces l'adaptation des systèmes de communications aux possibilités de perception acoustique. Comme chaque espèce animale, nous sommes immergés dans un monde sonore dont l'oreille ne capte qu'une partie de façon très inégale selon les fréquences, les niveaux d'amplitude.

4.2.1. Système périphérique

Le récepteur humain est formé d'un ensemble: celui ci comprend un premier étage de transducteurs déjà très complexe mais à partir desquels les chercheurs ont pu faire des mesures relativement reproductibles. Le système auditif périphérique effectue le codage du signal acoustique en influx nerveux. C'est un transducteur non linéaire de sons qui comporte trois niveaux physiologiques distincts: l'oreille externe (transmission dans l'air jusqu'au tympan), l'oreille moyenne (transmission mécanique grâce aux osselets et à un différentiel de surface), l'oreille interne (transmission en milieu aqueux dans la cochlée et activation des cellules cilliées). Sur la Figure 77, nous avons appelé ce stade le canal de la physiologie acoustique. Il est relié à l'étage central par le nerf auditif. Le traitement réalisé par cette étape de la perception fait subir des transformations au signal sonore en fonction de ses propriétés. Celles-ci le préparent à l'analyse réalisée par le système central. Cela influence l'intelligibilité car la résolution du signal au niveau de l'amplitude, de la fréquence et du temps est modifiée en même temps que le rapport signal sur bruit.

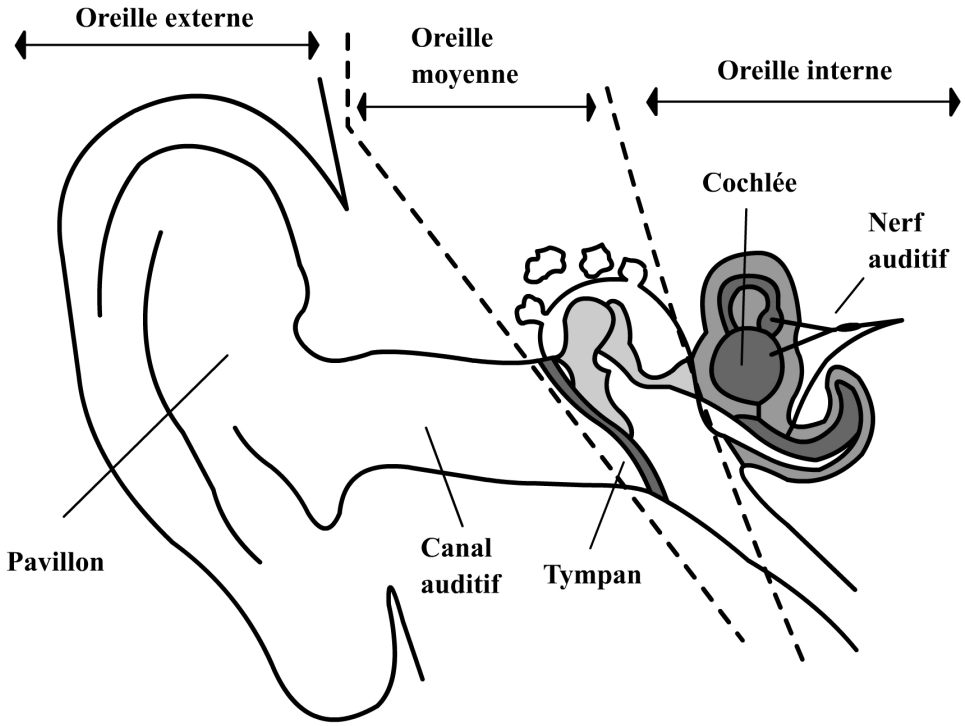

Figure 78 : Schéma du système auditif périphérique

4.2.1.1. Le prétraitement réalisé par l'oreille externe: La physiologie auditive

A ce niveau de la perception auditive, la réception est facilitée par la conformation du pavillon et du canal auditif. Cet ensemble joue un rôle important dans l'audition. Des mesures réalisées en placant un microphone au niveau du tympan ont permis de conclure qu'au niveau acoustique ces structures physiologiques forment un filtre dont la réponse en intensité en fonction de la fréquence n'est pas linéaire (Batteau 1967, Busnel 1976 in Busnel et Classe 1976).

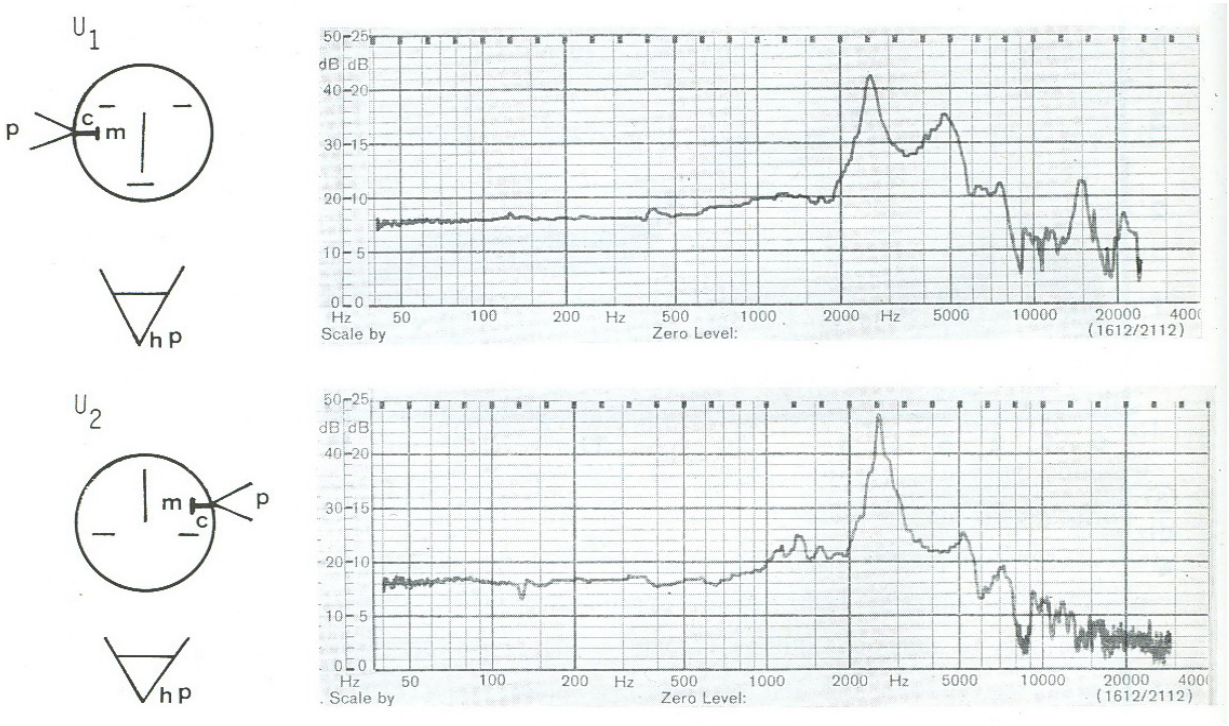

Figure 79 : Réponse du filtre de l'oreille externe (Busnel et Classe 1976, p. 42)

Outre les propriétés utiles pour la localisation auditive et la représentation spatiale des stimuli sonores du pavillon (Batteau 1967, Canevet 1989), l'effet du pavillon améliore le rapport signal sur bruit. Pour l'ensemble pavillon-canal, certaines bandes de fréquences sont favorisées. Ainsi, comme on peut l'observer sur Figure 79, le domaine de fréquence 1000-5000Hz est le plus réactif. Le canal auditif résonne à une fréquence d'environ 3800 Hz et crée une amplification de niveau des sons de 15 décibels.

A l'intérieur du domaine fréquentiel ainsi favorisé, l'augmentation maximale d'intensité est de 20 à 25 dB pour la bande de fréquences 1800-3500 Hz. C'est précisément le domaine de sensibilité maximum de l'audition humaine prise dans son ensemble et mesurée couramment avec la technique des audiogrammes. Ainsi, l'oreille externe prépare le signal acoustique à son traitement par les autres niveaux de l'oreille (intermédiaire et interne) et par le système central.

4.2.1.1.1. Conséquences pour le langage sifflé

Le domaine de fréquences favorisé par l'oreille externe correspond à celui exploité pour encoder le langage dans les langues sifflées. L'application du filtre du pavillon et du canal auditif dans une écoute de face[82] à un signal sifflé en turc, enregistré à 150m, montre que la fréquence fondamentale est renforcée (rapport signal sur bruit amélioré de 7% en moyenne[83]). Par contre cela ne change rien pour les harmoniques qui n'émergent pas mieux du bruit ambiant (Figure 80).

[82] Condition d'écoute normale d'un siffleur.

[83] Calcul effectué sur les pics d'énergie les plus intenses de chaque voyelle.

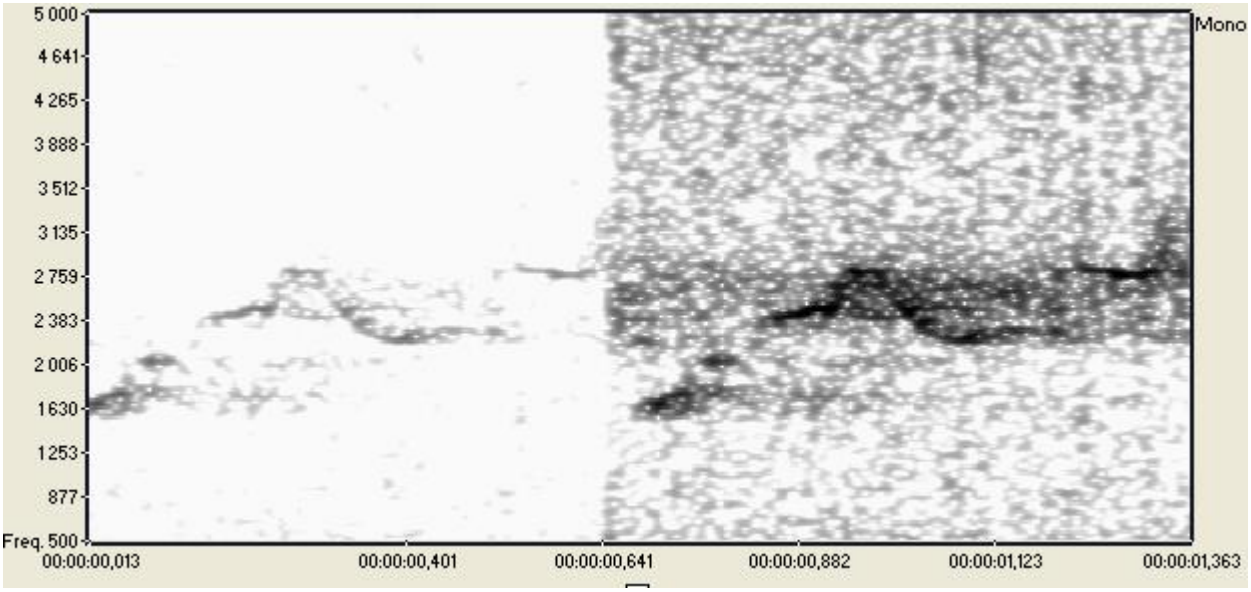

Figure 80 : Effet du filtre de l'oreille interne modélisé à partir de l'application d'un filtre artificiel à partir des valeurs de la Figure 79, position face à l'émission.

4.2.1.2. Aspects importants des autres niveaux physiologiques de l'oreille

4.2.1.2.1. L'oreille moyenne

La fonction de l'oreille moyenne est de permettre de dépasser l'effet de réflexion d'une grande partie de l'onde acoustique qui a habituellement lieu lors du tranfert d'energie d'un milieu à impédance faible (air) à un milieu à impédance moyenne (fluide). Un tel transfert entrainerait une perte de 30 dB en amplitude (Glattke, 1973) mais ce problème est résolu par l'oreille moyenne qui "*...augmente la pression d'environ 30 dB*" (Borden & Harris, 1980, p.165). Un tel résultat est obtenu par l'action combinée des osselets de l'oreille et du rapport d'environ 17:1 entre la surface du tympan et celle de la fenêtre ovale (Bekesy, 1960). La fonction de transfert du gain en pression de l'oreille moyenne n'est pas uniforme mais présente un pic à 1000 Hz, relativement plat (moins de -10 dB par rapport au maximum dans le domaine des fréquences de 300 Hz à 7000 Hz) et qui décroit à plus de 20 dB d'atténuation en deçà de 100 Hz et au delà de 10000 Hz (Nedzelnitsky, 1980). Cette étape de la transduction auditive est adaptée aux domaines de fréquences à la fois de la voix parlée et de la parole sifflée.

4.2.1.2.2. L'oreille interne

Description générale

Le point le plus important à comprendre est que l'oreille interne realise une analyse fréquentielle du signal acoustique grâce aux bancs de filtres créés par l'ensemble des filtres nerveuses. Quelles que soient les interprétations des multiples analyses scientifiques réalisées ces dernières années concernant les possibilités de codage du domaine fréquentiel perçu -soit codage tonotopique de 200 Hz à 20000 Hz, soit codage temporel 20 Hz à 4000 Hz soit les deux - le domaine de fréquences couvert par la voix humaine et par les sifflements permettent aux deux principes d'être appliqués. Dans l'hypothèse où ils seraient tous les deux

impliqués dans l'analyse fréquentielle ceci signifierait que la bande de fréquence de 200 Hz à 4 kHz serait une zone privilégiée, car elle permettrait un traitement fréquentiel suivant deux modalités ayant des propriétés différentes[84].

Les fibres nerveuses de l'oreille interne

Les fibres nerveuses de la cochlée répondent sélectivement au son dont la fréquence est proche de leur fréquence de décharge (ou fréquence caractéristique). Elles se comportent chacune comme un filtre passe bande. Un point important tient au fait qu'en plus de coder la fréquence d'une composante du signal, elles préservent dans une certaine mesure leur information temporelle en produisant une décharge à un instant précis de l'onde qui les stimule. Ce phénomène est appelé *phase locking*.

La structure même des fibres est complexe, elle pourrait expliquer certaines des non-linéarités de la perception auditive. En effet, différents modes d'activation ont été observés:

(i) certaines fibres ne déclenchent leur décharge que pour des variations spécifiques du signal (répétition, ou temps de montée de l'enveloppe) (Koch and Piper, 1979).

(ii) Beaucoup de cellules nerveuses du noyau cochléaire répondent peu -ou pas du tout- à un bruit filtré dont la largeur de bande est supérieure à une valeur donnée. Cependant, lorsque la fréquence centrale d'une telle bande de fréquence varie rapidement, ces cellules nerveuses répondent vigoureusement et ont une sélectivité fréquentielle prononcée (Miller, 1979).

(iii) D'autre part, LePage (1987) a fournit la preuve que la réponse fréquentielle cochléaire ainsi que sa précision dépendaient de l'intensité du stimulus.

(iv) De plus, des intéractions entre groupes de fibres introduisent des composantes fréquentielles qui sont absentes du signal acoustique ou suppriment partiellement des composantes comprises dans le signal.

Conclusion

Les découvertes réalisées sur l'oreille interne démontrent que le système auditif périphérique n'est pas un transducteur passif, comme l'est un microphone. Nous retiendrons également qu'il existe différents modes d'activations des fibres neuronales suivant les propriétés dynamiques du signal et que des processus de correction ou d'augmentation de la précision de la réponse en fréquence en fonction de l'intensité du signal sont en jeu. A ce jour, aucune conclusion définitive ne peut être tirée à propos du mode de traitement des fréquences par la cochlée ou de son contrôle par des mécanismes centraux. Enfin, il apparaît clairement que les sifflements et la voix sont dans les zones les plus sensibles de cette partie dynamique de l'oreille.

[84] Dans ce cas, la bande de fréquence couverte par le téléphone (400 Hz à 4000 Hz) et celle couverte par la parole sifflée sont particulièrement favorisées.

4.2.1.2.3. Conclusion pour l'oreille périphérique

A chaque étape de filtrage réalisé par le « canal physiologique » le signal de parole est transformé suivant de nombreux processus pour la plupart non linéaires qui améliorent le rapport signal sur bruit général en privilégiant certains domaines de fréquences. Une dynamique particulière des signaux est également favorisée. Le domaine de fréquences mis en valeur par l'effet cumulatif des trois niveaux de l'oreille périphérique (externe, moyenne et interne) va de 1000 à 4000 Hz. Cette bande de fréquence est efficace car elle se situe largement au dela de la majorité des bruits de fond de notre environnement. D'autre part, jusqu'à 4000 Hz l'oreille est en mesure de réaliser une analyse temporelle précise de la dynamique des signaux. Etant donné que cette bande de fréquence permet d'obtenir une intelligibilité de plus de 90% avec un signal de voix (Moore, 1982[85]) et contient l'ensemble des signaux sifflés, nous pouvons raisonablement penser qu'elle représente une zone clef de la perception de la parole.

4.2.2. Psychoacoustique: une première approche du système central

La psychoacoustique a pour objet l'étude expérimentale des relations quantitatives entre des stimuli acoustiques mesurables physiquement et les réponses de l'ensemble du système auditif de l'être humain. Elle permet donc de confirmer les hypothèses que nous avons émises au vu des résultats de physiologie auditive, tant sur les limites de décodage[86] des sons simples que de certaines dynamiques. Elle permet également de tester la résolution de la perception de la hauteur, de l'intensité et de la durée.

4.2.2.1. Le domaine d'audibilité: résolution de la perception auditive humaine en fonction de la fréquence et de l'intensité

4.2.2.1.1. Analyse synthétique

La courbe de "réponse" de la sensation subjective auditive d'égales intensité (lignes isosoniques: Fletcher & Munson 1933) en fonction de la fréquence donne une idée générale de la perception auditive humaine (Figure 81). Cette courbe est statistiquement normale à des excitations de fréquences pures variables pour un ensemble d'intensités différentes. Contrairement aux travaux sur l'oreille externe, il n'est plus possible d'obtenir des mesures physiques directes mais c'est le jugement de sujets testés suivant un protocole précis qui permet de tirer des conclusions sur la perception acoustique. Les tests qui ont permis d'établir les résultats de la Figure 81 ont été réalisés à partir de sons purs, c'est à dire des sons de type sinusoïdaux ayant une fréquence et une intensité donnée à chaque présentation. Les sensations de hauteur du sujet sont appelées « hauteur tonale » (HT). *« Nous dirons que la HT d'un son pur est l'attribut perceptif du son sur la base*

[85] Moore obtient 90% d'intelligibilité des phrases avec un signal de parole filtré dans la bande 1000-2000 Hz.

[86] Dans ce travail, nous considérons que le signal est encodé par l'émetteur et décodé par le récepteur. Certains chercheurs utilisent le terme encodage de manière différente puisqu'ils l'utilisent pour le passage de l'acoustique à l'influx nerveux réalisé dans l'oreille.

duquel il est possible de lui apparier, par ajustement de fréquence, un autre son pur différant par le niveau d'intensité. Et nous dirons aussi qu'après appariement de leur HT, les deux sons ne diffèrent que par la « sonie ». Ce qui signifie que deux sons purs quelconques ne peuvent différer que par la HT et/ou la sonie » (Demany 1989, p. 42). Ces approches psychoacoustiques ont permis de remarquer que l'audition humaine ne peut détecter un changement que si la variation de l'excitation dépasse une certaine quantité ou *seuil différentiel*.

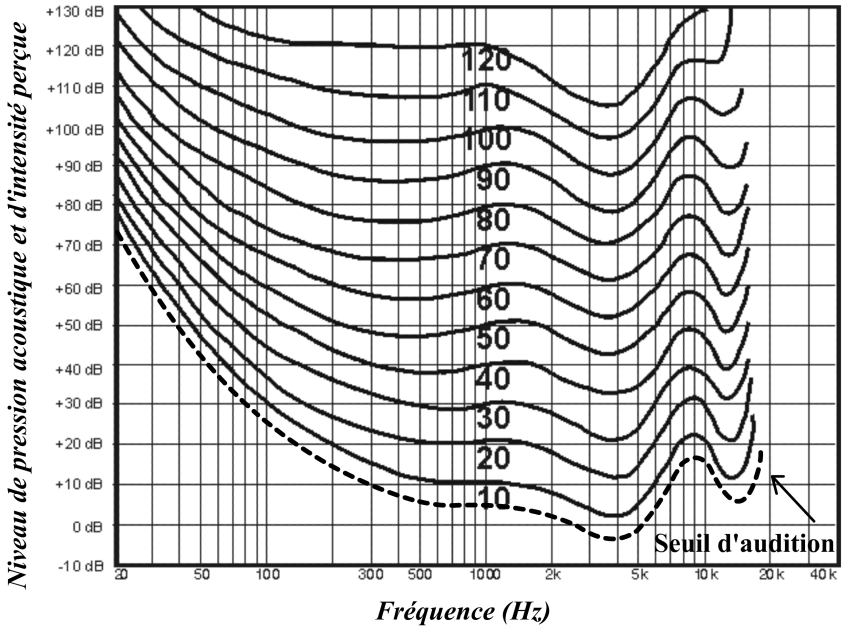

Figure 81 : Lignes isotoniques et niveau d'audition

Cette représentation graphique nous fournit plusieurs renseignements :

-Le domaine d'audibilité humain s'étend entre les deux courbes des seuils d'audibilité et de douleur.

-Une zone de fréquences prédomine s'étendant autour de 3000Hz (de 2000 à 5000 Hz) où l'être humain parvient à distinguer les plus faibles intensités.

-De part et d'autre de cette zone, la sensibilité diminue. Ainsi à 30 Hz, il faut augmenter l'intensité de 60 dB (c'est à dire multiplier l'amplitude du son par 1 million) pour que ce son de basse fréquence soit tout juste audible, comme l'était celui de 3000Hz. La sensibilité décroit également du côté des hautes fréquences et au delà de 20000 Hz, la sensation auditive ne permet plus de percevoir les sons.

Comparaison avec quelques analyses spectrales de parole sifflée

Les données présentées ici sont issues d'un sifflement émis à 105 dB émergeant à 15 m (distance d'enregistrement) de 40-50 dB d'un bruit de fond de 30-40 dB(Figure 82 et Figure 83).

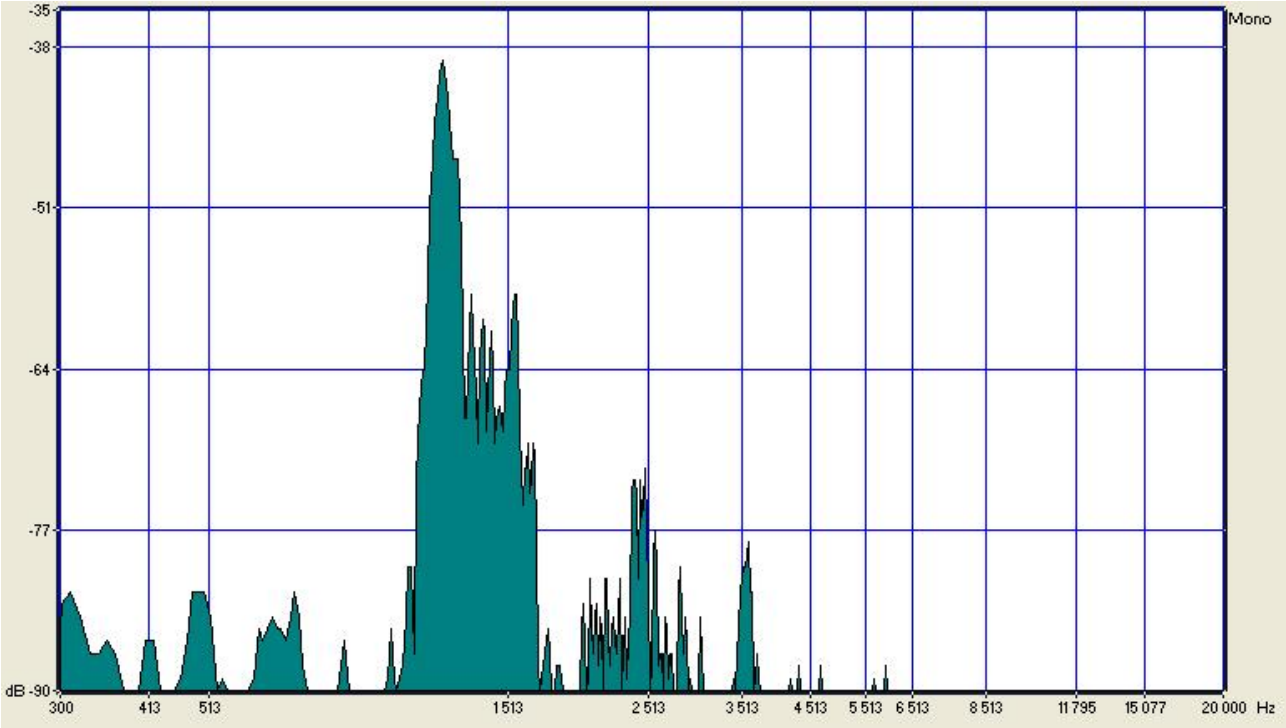

Figure 82 : Analyse fréquentielle de la voyelle «a» de la dernière syllabe du mot espagnol « montañeta »

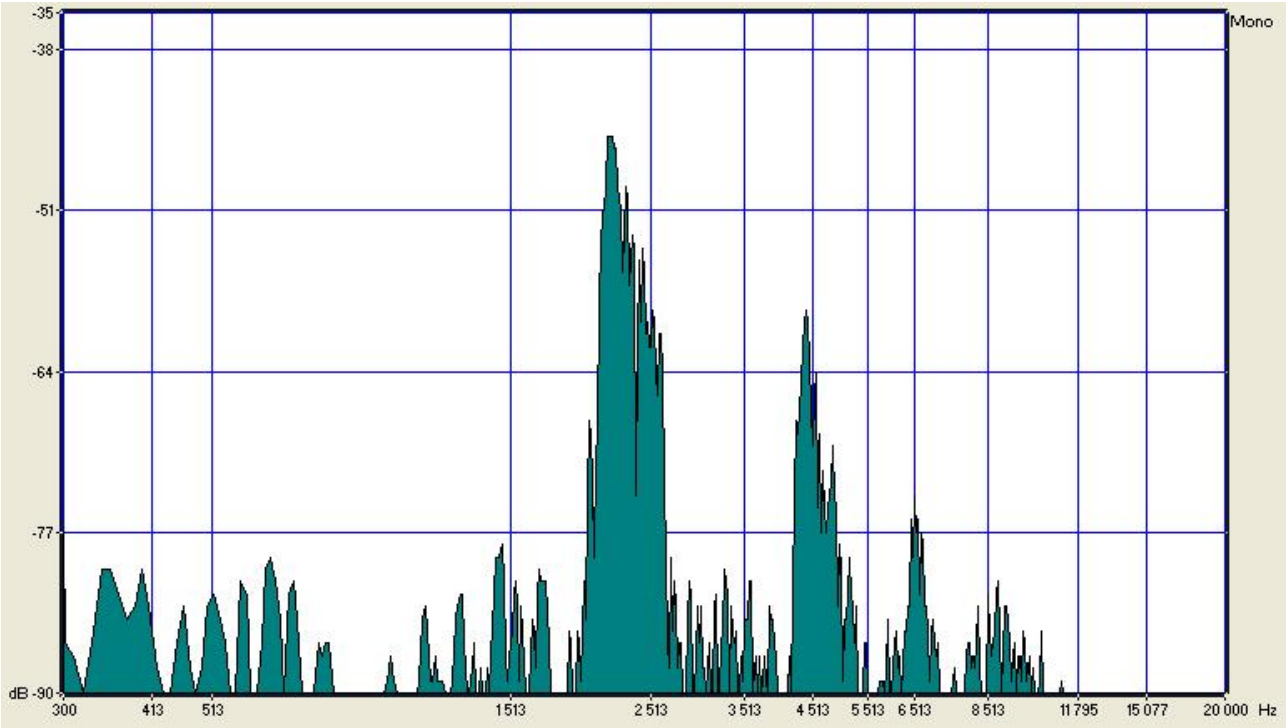

Figure 83 : Analyse spectrale de la fréquence d'un point de la modulation de la consonne «ñ» du mot espagnol « montañeta »

Remarques :

-Les analyses spectrales de la fréquence présentées sur les Figure 82 et Figure 83 témoignent du fait que le sifflement se situe dans la zone privilégiée par la perception en fréquence (Figure 86).

-La fréquence ayant l'amplitude la plus élevée sur ces deux figures est la fréquence qui ne sera pas dégradée à moyenne et longue distance, c'est donc elle qui porte l'essentiel de l'information alors que ses harmoniques sont vite dégradées.

-A courte distance le sifflement est également dans la zone la plus précise de la perception de l'intensité.

4.2.2.1.2. Discrimination en fréquence ou résolution fréquentielle

Le seuil differentiel en fréquence, plus petit intervalle perceptible entre 2 sons purs, varie beaucoup avec la fréquence. Il présente un minimum entre 1500 et 4000Hz et croît autour de ces valeurs. Le minimum est de l'ordre de 1/300e d'octave à 2000 Hz et le maximum est de l'ordre de 1/6eme d'octave à 30 Hz. Pour une même fréquence, le seuil différentiel diminue quand l'intensité augmente. Ce qui confirme que la hauteur perçue est déterminée en première approximation par sa fréquence. Au-dessus de 1000 Hz la fréquence doit être plus que doublée pour obtenir une sensation de hauteur double. Ces résultats ont été obtenus par une méthode qui consiste à rechercher la plus petite différence de fréquence que doivent présenter deux sons purs stables et successifs pour qu'ils soient juste discriminés. L'intervalle de silence entre les deux sons n'est pas critique, i.e. trop long pour empêcher de les comparer, tant qu'il est compris entre 100 et 1000 ms (Harris 1952).

4.2.2.1.3. Echelles de discrimination ou de sélectivité en fréquence

Cadre de l'analyse

L'échelle de HT peut être graduée en *mel* (définition de Carlson et al, 1970). Un écart constant de mel correspond à un écart constant de hauteur perçue. Une autre échelle psychoacoustique perceptive des fréquences a été mesurée à partir de la largeur de la bande critique pour laquelle deux fréquences peuvent être discriminées (sélectivité fréquentielle). Cette bande critique variable en fonction de la fréquence est appelée *bark*. Un écart constant en bark va correspondre à un écart de perception entre des fréquences simultanées. Ces échelles, construites comme des approximations mathématiques de la perception fréquentielle humaine, à partir de données expérimentales, témoignent du fait que le mode de perception des fréquences par le système cognitif humain se rapproche d'une forme logarythmique[87]. Ces deux échelles ne sont pas acoustiquement équivalentes malgré certaines approximations (ou redéfinitions) faites dans des travaux de recherche de reférence (Peterson et Barney 1952, Zwicker & Fastl, 1990). Les études psychoacoustiques de perception de la fréquence réalisées par Zwicker et Fastl rendent équivalentees la sélectivité fréquentielle (bark), la discrimination de fréquentielle (jugements de la fréquence : *« frequency-jnds »*) et la perception de la hauteur (mels) dans un souci d'uniformisation. Cependant cette simplification ne rend pas compte des différences réèlles entre les différents modes d'observation de la perception de la fréquence. Cette conclusion est imposée par la considération de résultats plus récents sur les largeurs de bande des filtres auditifs perceptuels : ceux-ci ont montré que les largeurs de bandes critiques sont en général

[87] Nous verrons un peu plus loin que la réalité est plus tridimentionelle.

surévaluées en particulier à basse fréquence et que la sélectivité fréquentielle du système auditif est mieux représentée par l'échelle ERB, échelle utilisant une méthode de détermination des largeurs de bandes fréquentielles mise au point par Patterson (1976) et qui est aujourd'hui préconisée par Rosner & Pickering (Rosner & Pickering, 1994) pour étudier la perception des voyelles.

Masquage et rapport Signal sur Bruit des langues sifflées

L'échelle ERB est importante pour expliquer certains aspects perceptifs de l'émergence du signal sifflé dans le bruit. En effet d'après les données de Patterson (1976) et de Moore et Glasberg (1983) les filtres auditifs seraient des filtres à pentes exponentielles et convexes au sommet dont la largeur peut être estimée par le calcul d'un Equivalent Rectangular Bandwidth (ERB) :

$$ERB = 6{,}23f^2 + 93{,}39f + 28{,}52$$

Avec f en kHz et ERB en Hz.

Ainsi pour un sifflement centré sur f=2 kHz, ERB=250 Hz. Donc de 1 à 4kHz on obtient une bande ERB qui varie entre 120 et 500 Hz. Nous avons mesuré qu'à courte distance (15 m) le signal émerge du bruit de fond avec une largeur de bande de 450 Hz au maximum (Figure 82 et Figure 83). Il activera au maximum 4 filtres perceptifs de l'oreille à un instant donné. A plus grande distance, la largeur de la bande de fréquence qui émerge est plus réduite. Cependant nous avons mesuré qu'à 550m un signal qui est noyé dans le bruit de fond reste intelligible. Cela n'est possible que parce que la sélectivité de l'oreille a pour conséquence que, lorsqu'un bruit à large bande masque un son relativement pur, seules les composantes fréquentielles du bruit proches de la fréquence du signal sont effectivement masquantes pour ce dernier[88]. Par conséquent, il est pertinent d'observer l'émergence d'une phrase en langue turque comprenant toute l'information sifflée dans une bande de fréquence de 1500 à 3000 Hz en ne considérant que le bruit dans ce domaine (le résultat du filtrage est présenté sur les Figure 84 et Figure 85). A la limite d'intelligibilité le son émerge encore de 20 dB de cette bande avec une largeur de bande de 150 Hz. On peut donc dire que la bonne sélectivité auditive combinée aux caractéristiques du sifflement permet d'expliquer à la fois la précision de la perception et la bonne émergence du bruit.

[88] Le masquage réalisé par un son de fréquence donnée sur des sons de fréquences différentes traduit l'amplitude et l'étendue de l'excitation du système auditif (Zwiker 1970). Zwicker a su adapter l'explication des phénomènes de masquage à des sons dont le niveau varie rapidement au cours du temps comme ceux de la parole. Il a mis en évidence qu'à un niveau donné, lorsqu'un bruit vient interférer avec le signal, le masquage qu'il exerce sera d'autant plus important que sa fréquence est inférieure et son intensité supérieure au signal de parole. Dans une bande de fréquence de 250 Hz (bande d'émergence typique d'un sifflement), très peu de signaux seront concernés.

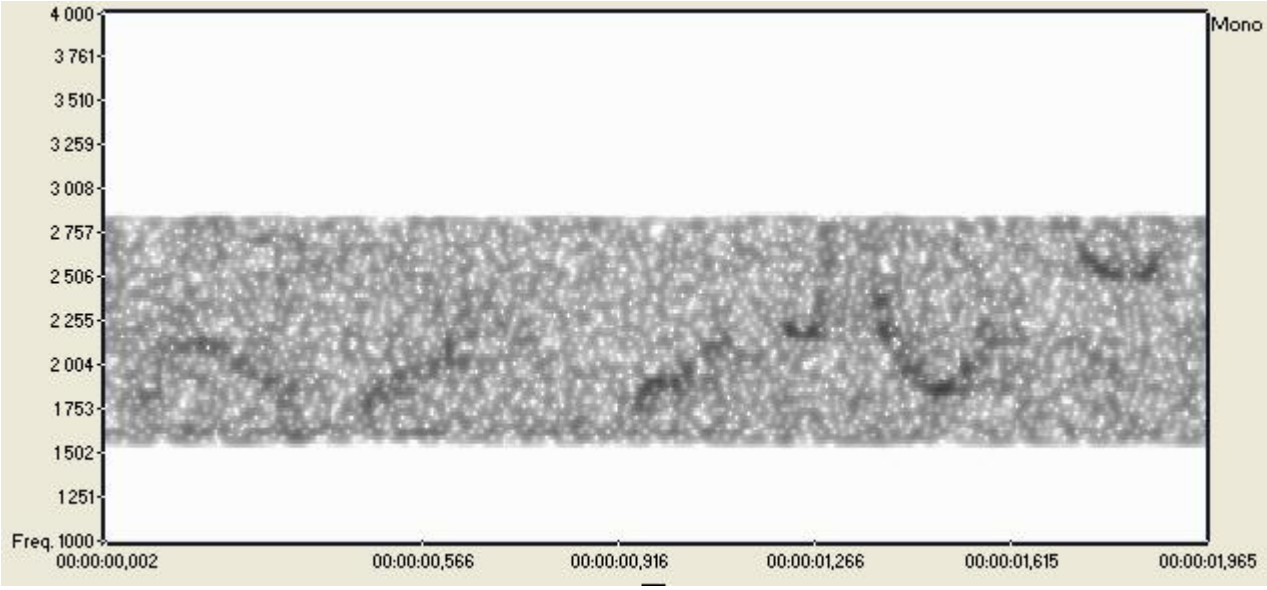

Figure 84 : Sonogramme de la phrase turque « Mehmet okulagit » à 550m élimination des bandes de fréquence en dessous de 1500 Hz et au dessus de 3000 Hz

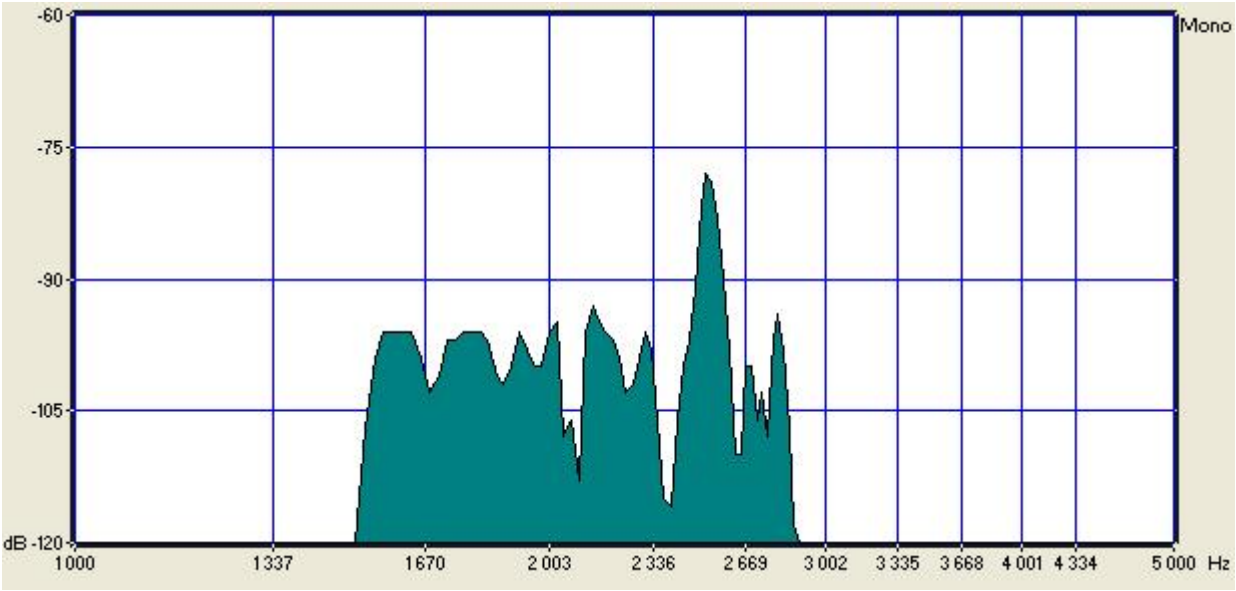

Figure 85 : Emergence en amplitude du sifflement de la phrase turque à 550m après filtrage (1500-3000 Hz)

4.2.2.1.4. Discrimination en intensité

Le seuil différentiel d'intensité suit des variations similaires à la fréquence, sa valeur minimale témoignant d'une sélectivité optimale de l'oreille étant de 0,02 dB pour un son de 4000 Hz à 90 dB (Stevens et Davis, 1938 p.140). La « sonie » (intensité perçue d'un son pur) est déterminée, en première approximation par son niveau de pression acoustique. Mais, à niveau de pression acoustique égale, les sons à fréquence basse ou très élevée ont une sonie inférieure aux sons de fréquence moyenne. L'échelle des décibels qui est la plus communément utilisée pour rendre compte de la sensation d'intensité est une des plus représentatives de la perception. Elle a été établie à l'origine par les premiers ingénieurs travaillant sur le téléphone qui avaient

trouvé en elle une échelle appropriée pour relier l'amplitude sonore à la perception humaine suivant une fonction non linéaire de l'intensité ou de la pression acoustique.

Dans le but d'étudier les éléments de la phonétique, il est légitime de se demander dans quelle mesure l'échelle des decibels est appropriée pour rendre compte de la netteté de la perception des voyelles des consonnes, des tons et des contours de tons. En effet, la perception humaine de l'intensité est également complexe car elle est dépendante de la fréquence et de la durée du stimulus. Il y a trois échelles principales qui peuvent être considérées pour paramètrer l'amplitude du signal de parole. Ce sont l'échelle des pressions (Pascal), l'échelle dérivée des jugements d'intensité (intensity-jnd-rate), l'échelle de sonie et l'échelle des décibels (loi de Fechner). Dans une étude comparative réalisée par Mannell (1994), les échelles jugées comme les plus représentatives de la position des pics (maxima) d'amplitude est l'échelle des Pascals, alors que la plus adaptée à rendre compte des évolutions de l'amplitude le long de la forme spectrale est l'échelle des logarythmes de sonie et l'échelle des décibels. Un des résultats les plus intéressants de cette étude concerne le fait que ces deux dernières échelles sont les plus efficaces pour les consonnes faisant un stop (occlusives), car leur perception nécessite la prédiction de l'amplitude basse et de l'amplitude haute de ces composants phonétiques.

4.2.2.1.5. Diagramme synthètique

A partir des données précédentes, on peut calculer le nombre total de sons élémentaires que le système auditif peut distinguer et étudier leur répartition dans l'aire auditive en dressant une carte de la finesse de résolution de l'oreille en fréquences et en intensité (Figure 86). Le coeur de ce schéma est une zone privilégiée située entre 70 et 120 dB dans l'intervalle de frequence 1500-4000Hz correspondant à de meilleurs seuils d'audibilité et à une plus grande sélectivité de l'oreille (Stevens et Davis 1938).

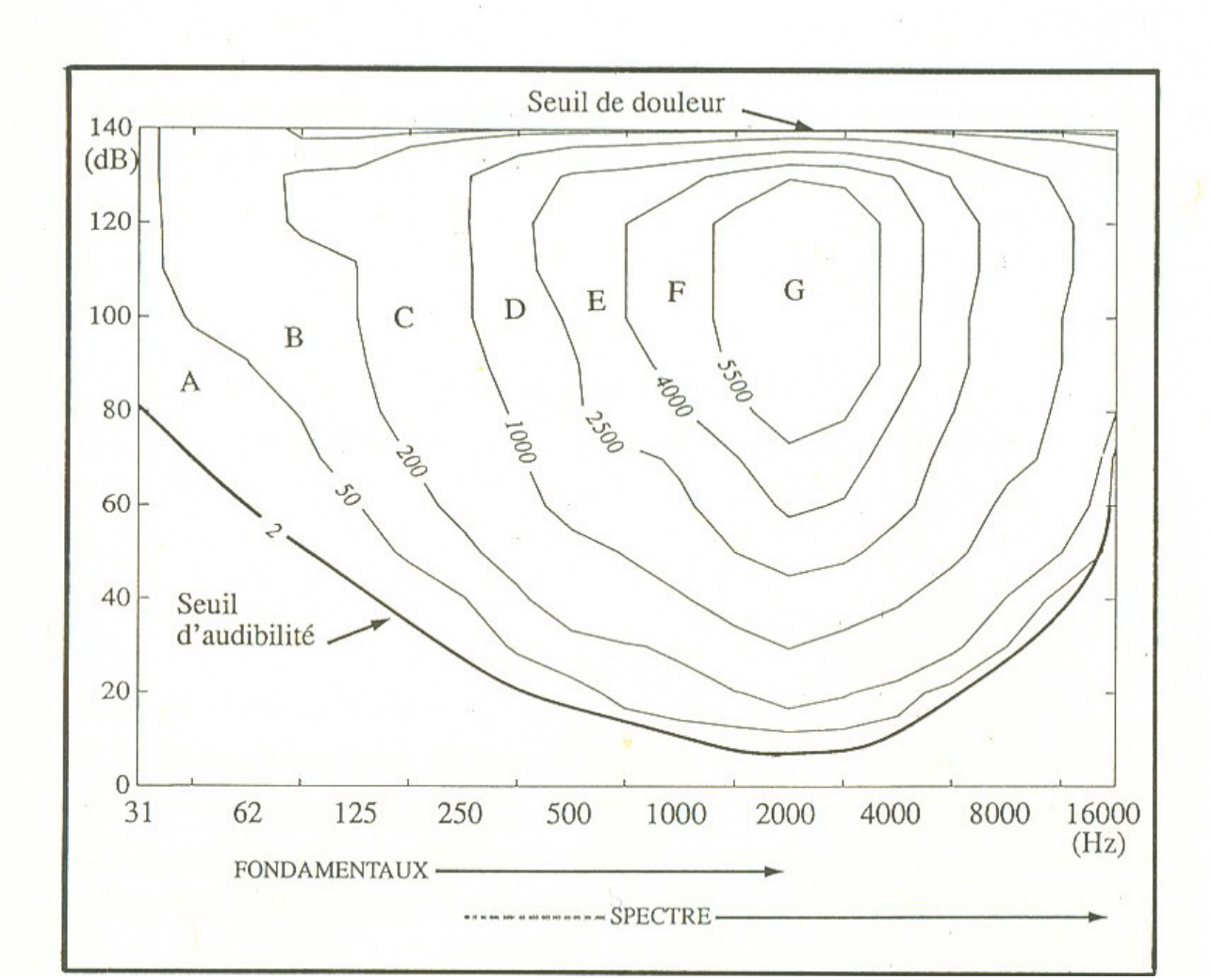

Figure 86 : Variation de la finesse de résolution de l'oreille humaine

Variation de la finesse de résolution de l'oreille humaine « à partir de l'estimation du nombre de sons élémentaires perceptibles par unité de surface (10 dB * ½ octave, les courbes délimitent les zones de finesses croissantes (de A à G) dont les seuils inférieurs sont respectivement 2, 50, 200, 1000, 2500, 4000 et 5500, le maximum (au cœur) étant de l'ordre de 7000 unités l'intensité (d'après Stevens et Davis, 1938 p.140 in Zenatti, 1994, p.89).

Certaines productions sonores humaines sont partiellement adaptées à cette zone de perception. Tout d'abord la parole classique, première et plus importante source d'éducation de l'oreille. Par exemple, les variations spectrales qui permettraient de distinguer les voyelles et les principales transitions phonétiques, donc de reconnaître les mots, évoluent entre 300 Hz et 4000 Hz. D'autre part, de nombreux instruments traditionnels exploitent ces domaines (flûtes, violons, ...).

En ce qui concerne les sifflements, les conditions d'émission du signal évoluent typiquement entre 80 et 110 dB. Donc, en production, les langues sifflées se situent dans les zones de meilleure perception à la fois en intensité et en fréquence (zones F et G sur la Figure 86). En raison de l'usage des langues sifflées à longue

distance, l'amplitude de perception n'a de sens qu'en terme de contrôle de la dynamique par l'émetteur[89] ou d'écoute à courte distance (par exemple par dessus un torrent comme en Turquie, dans une taverne comme en Grèce ou dans un marché comme au Mexique). Dans ces cas, c'est un point d'explication supplémentaire à la bonne émergence du sifflement dans le bruit.

4.2.2.2. Autres aspects de la perception mis en valeur par la psychoacoustique

4.2.2.2.1. Composition de la Hauteur Tonale des sons complexes

Parmi tous les sons dont le système auditif est capable d'extraire une sensation de hauteur tonale, la grande majorité est constituée de sons écologiques complexes et périodiques (ou quasi-périodiques). C'est le cas des voyelles de la voix et des sons de très nombreux instruments de musique. Ces sons complexes périodiques sont harmoniques[90].

En général, un auditeur à qui l'on confie la tâche d'apparier un son pur à un tel son complexe, ajustera la fréquence du son pur à la fréquence du premier de tous les harmoniques (Davis et coll 1951). Cette fréquence est dite *fréquence fondamentale*. Dans la voix humaine c'est la fréquence de phonation ou de vibration des cordes vocales. La différence de qualité des timbres du son pur et du son complexe rend cette tâche plus difficile qu'entre deux sons purs.

Le timbre est un attribut perceptif lié au spectre de fréquences d'un son à tel point que deux sons complexes ayant la même fréquence fondamentale seront tout de même discriminés en fonction du contenu du spectre fréquentiel. Si l'énergie de l'un est plus grande et compacte aux fréquences élevées on dira qu'il est plus brillant et plus *« aigu »* (Demany 1989, com pers Gautheron 2005). Par conséquent, la HT d'un son complexe périodique a deux qualités distinctes: la qualité de timbre, également appelée qualité de *Hauteur Brute* (HB), (Risset, 1968) et la qualité de *Hauteur Fondamentale* (HF).

Comme la HF manifeste des ambiguités d'octaves, cette qualité est jugée par certains auteurs comme parente de la qualité de *chroma*. La HF est en effet plus liée à la périodicité qu'à la fréquence (Plomp 1967). Le chroma serait une qualité selon laquelle deux sons purs dont le rapport de fréquence est proportionel à 2 et qui forment donc un multiple d'intervalle d'octave sont perçus comme similaires ou identiques (Bachem 1950).La qualité de HB est quant à elle associée à la qualité de tonie.

Demany (1989) précise que ces deux qualités peuvent être contrôlées séparément en production : *« La HB et la HF peuvent être variées de façon complètement indépendantes, en modifiant les caractéristiques physiques distinctes de ce son complexe periodique ».* Il explique aussi qu'elles sont perçues sans

[89] Ce facteur peut être particulièrement important à l'acquisition de la technique de la langue sifflée après la période d'apprentissage passif du bébé. En effet, les siffleurs de Turquie, de Grèce, de la Gomera ou du pays mazatèque expliquent avoir appris le sifflement simultanément à la version parlée, vers 2 ans.

[90] Leurs composantes spectrales ont des fréquences qui sont des multiples entiers successifs d'une même fréquence f ; tout autre son complexe est dit « inharmonique » comme les percussions des tambours par exemple.

apprentissage particulier par tout être humain *: « Il est permis de dire que HB et HF sont deux qualités de HT qu'évoquent spontanément les sons complexes périodiques. En général, pour la plupart des sons complexes périodiques émis dans notre environnement courant, ces deux qualités s'imposent à la conscience sans qu'un effort d'attention soit nécessaire pour les appréhender. Elles sont le produit de ce que Helmholtz appelait l'écoute « synthétique » »* (Demany 1989 p.55). Dans les sons complexes périodiques de la parole classique humaine, ces deux qualités sont exploitées pour encoder les différents éléments porteurs de sens, suivant des principes généraux mais également des règles propres à chaque langue. Par exemple, l'auditeur perçoit les tons ou l'intonation au niveau de la HF alors que le timbre des voyelles est perçu au niveau de la HB. D'une manière générale, suivant la structure des langues les deux qualités de HB et de HF sont combinées différemment pour marquer la phonologie de la langue. La grande variabilité des combinaisons possibles rend nécessaire un approfondissement de la compréhension de la perception de ces deux qualités.

Le spectre complexe de la Hauteur Brute (HB) dans les langues

Pour toutes les langues, la présence d'un son complexe périodique indique l'existence du phénomène de « voisement » qui caractérise les voyelles et certaines consonnes (« consonnes voisées » utilisant les vibrations des cordes vocales). Deux voyelles ou deux consonnes voisées identiques peuvent porter des tons ou des intonations différents tout en gardant leur identité perceptuelle. De même plusieurs types de voix ayant des HF différentes encodent les mêmes voyelles ou les mêmes consonnes. Dans tous ces cas, ce sont les similarités de la répartition spectrale des harmoniques et de l'intensité qu'elles contiennent qui rapprochent perceptivement différentes réalisations. Plusieurs auteurs ont montrés qu'un auditeur parvient en effet à utiliser des indices dans les harmoniques (Hartmann et al 1986, Plomp 1965).

Approches phonétiques et formants

De nombreux phonéticiens se sont intéressés à cet aspect de l'acoustique des voyelles et des consonnes voisées. Ils ont observé que les zones de densité spectrale les plus intenses sur les harmoniques forment des « régions fréquentielles de concentration d'énergie » inter-harmoniques appelés formants (Potter et Steinberg 1950). Plus la voix est grave, plus ses harmoniques sont rapprochées et donc plus les formants seront facilement visibles sur un spectrogramme. C'est pourquoi les voix d'hommes posent moins de problèmes à l'identification des formants que les voix de femmes ou d'enfants dont la HF (ou Fo) est plus haute. D'autre part, plus l'articulation est réalisée clairement syllabe par syllabe, plus les concentrations fréquentielles des formants sont graphiquement identifiables et stables.

Malgré cette différence de résolution graphique, l'identité perçue des voyelles ne décroît pas avec la hauteur du Fo et, dans certaines limites, la perception de la parole n'est pas affectée par la rapidité de production (Gay 1977)[91]. D'autre part, en parole continue, les formants vocaliques sont régulièrement déplacés de leur

[91] Gay (1977) a montré par exemple que sur de la parole rapide, les transitions de formants des consonnes changeaient de plus de 60 Hz pour le /p/ et de 25 Hz pour le /b/ sans affecter la reconnaissance des syllabes. Il semble donc que les modulations pertinentes pour la reconnaissance ou la détection d'éléments de la parole varient en fonction du type de parole : rapide/lente, statique/naturelle entre autres.

fréquence habituelle, ce qui n'affecte pas l'intelligibilité des phrases. De nombreux expérimentateurs y ont vu l'effet d'une assimilation contextuelle (Lindblom 1963, Steven et House 1963). L'existence d'un grand nombre d'irrégularités formantiques et la difficulté méthodologique de leur estimation est encore une réalité aujourd'hui.

Hauteur Fréquentielle (HF), tons et intonation:

Parallèlement à l'existence de formants dans la qualité HB, les phonéticiens ont remarqué que les tons phonologiquement distinctifs caractéristiques des langues tonales sont exprimés à travers la HF. En raison de l'indépendance des qualités de HB et de HF, une même voyelle (HB) peut porter tous les tons (HF) d'une langue tonale (comme nous l'avons vu pour le mazatèque dans la partie de la typologie). Cette souplesse est également exploitée à la fois dans les langues à tons et les langues sans ton pour transmettre des informations liées à l'intonation.

Hauteur Brute, Hauteur Fondamentale et langues sifflées

Les langues sifflées qui n'encodent l'information linguistique qu'à travers une seule bande de fréquence privilégient l'une des deux hauteurs mais comme nous l'avons vu, ceci ne veut pas dire que l'autre n'est pas prise en compte. Ainsi la transposition des langues sifflées est le résultat d'un compromis entre HB et HF qui permet de rendre compte des éléments les plus phonologiquement distinctifs de chaque langue afin d'optimiser l'intelligibilité du message.

Particularités de la HF et perception des langues sifflées

Risset (2000) souligne que les effets d'ambigüité d'octave cachent parfois que notre perception de la HF est en fait mieux représentée sous forme d'une spirale (Figure 87). Il a obtenu une illusion sonore en acoustique musicale qui illustre ce phénomène : un son qui semble descendre en hauteur peut conduire à une hauteur perçue finale bien plus élevée que celle du départ. *« La notation en spirale rend compte de la similarité des notes à intervalles d'octave. Les gammes, descentes ou montées indéfinies correspondent à la dégénerescence de la spirale en un cercle lorsqu'il y a extrême ambiguïté d'octave »* (com. pers. Risset 2005).

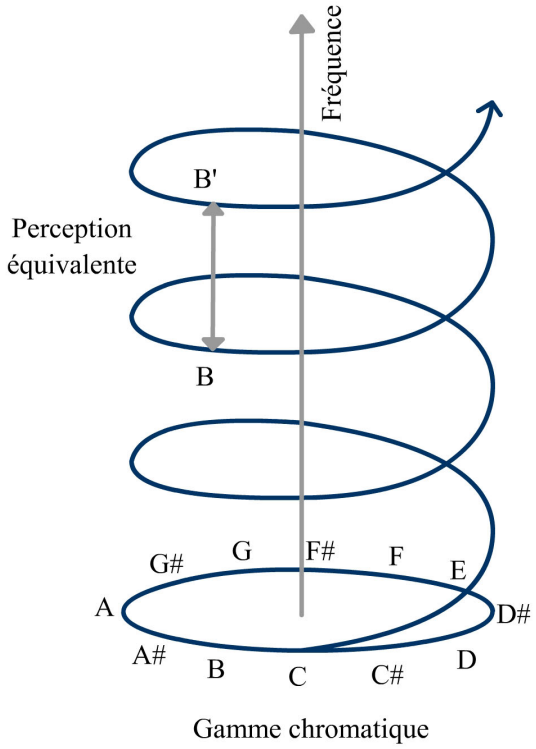

Figure 87 : Perception de la hauteur fondamentale (schéma inspiré de Shepard, 1965)

Par apprentissage et imprégnation culturelle la spirale peut être affaissée en un cercle, comme dans le cas de la musique occidentale tonale. Deux notes de fréquences différentes mais de position équivalente sur le cercle chromatique seront perçue similaires (Par exemple B et B' toutes deux perçues comme un « si »).

Il se peut que l'affaissement de la spirale par imprégnation culturelle puisse expliquer pourquoi les transpositions sifflées de langues tonales évoluent dans une bande de fréquence limitée à un octave pour une même phrase (la même remarque est valable pour les formes parlées). Une rupture de cette limite entraînerait des possibilités de confusion entre les hauteurs perçues des tons comme dans le cas de l'illusion provoquée par Risset. Cette remarque n'est pas limitée aux langues tonales puisqu'en turc et en grec, la même limite de un octave semble s'appliquer également. Par contre, comme nous l'avons vu, ce facteur n'a apparement pas la même importance lors de la perception des voyelles sifflées du silbo car ces dernières évoluent couramment sur plus d'un octave dans une même phrase. Il se peut que les Gomero soient culturellement moins marqués par l'octave. Ce serait un aspect interessant à tester avec ces siffleurs.

4.2.2.2.2. Perception temporelle

L'analyse des indices acoustiques de la parole, en particulier liés aux consonnes, permet d'observer que la durée est un paramètre important à la fois pour la différenciation d'évènements proches et pour la détection de la fréquence des indices phonétiques. Green (1985) a distingué deux classes principales de phénomènes auditifs temporels, l'intégration temporelle et l'acuité temporelle.

Intégration temporelle

Les études sur l'intégration temporelle cherchent à trouver la longueur des intervalles pour lesquels le système auditif intègre l'information acoustique de différents éléments. En effet, le phénomène d'intégration temporelle, qui est valable sur des durées allant jusqu'à 200 ms est à la fois lié à la possibilité de détecter des silences (*gap detection*) et au phénomène de masquage non simultané (ou *masquage temporel*). Pour attribuer une hauteur précise à un son pur il faut non seulement qu'il ait un certain niveau de pression acoustique appelé seuil de perception tonale mais également une durée d'au moins 10 ms. En dessous de cette limite, il tend à être perçu comme une transitoire ou un clic auquel une hauteur peut difficilement être attribuée.

Acuité temporelle

Les études sur l'acuité temporelle analysent la rapidité de réponse du système auditif à des évènements acoustiques brefs. Les durées minima de séparation de deux évènements successifs sont mesurées. Ainsi les seuils de détection de silence (*gap detection thresholds*) ont des valeurs uniformes de 6 à 8ms sur toute la bande de fréquence de perception humaine. Par contre les seuils différentiels de durée de deux sons varient: ils sont de 50 ms pour des sons purs de 1 s (sons ayant des caractéristiques temporelles proches des voyelles) et de 2 à 3 ms pour des sons purs de 5 à 20 ms (sons ayant des caractéristiques proches des consonnes).

Précisions sur le masquage temporel

D'autre part, en ce qui concerne le phénomène de masquage temporel, une différence entre l'effet de masquage « antérieur » (*rétroactif*) et celui de masquage « postérieur » (*proactif*) montre que la réponse temporelle du système auditif est asymétrique avec une réaction plus rapide aux attaques de stimuli (*stimulus onsets*) (perception des variations de 20 dB en 10 ms) qu'aux relâchements (*offsets*) (perception des variations de 20 dB en 20 ms). Le rôle fondamental de l'entité *Consonne-Voyelle (CV)* observé par les linguistes dans la plupart des langues est ici en partie justifié.

La résolution temporelle moyenne liée au masquage est de 20 à 30 ms. Avec l'apprentissage, ces valeurs sont améliorées. Elles varient également en fonction du contexte sonore (Figure 88).

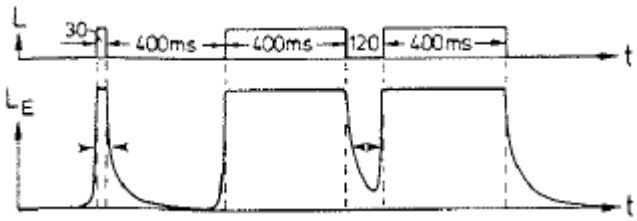

Figure 88 : 3 formes d'excitations temporelles provoquées par des sons purs de durées différentes (Zwicker 1982)

Pour chaque événement on observe que l'attaque engendre un masquage rétroactif plus court que le masquage proactif. C'est pourquoi un silence de 120 ms pris entre 2 excitations de 400 ms aura la même durée perceptive qu'un son de 30ms.

Durée et perception de l'intensité

Pour des durées sonores inférieures à 1 seconde, qui caractérisent la plupart des éléments de la parole, la sonie (intensité subjective) augmente avec la durée (Botte 1989). D'après Pedersen et al (1977), la relation entre la sonie et la durée est une fonction exponentielle dont la constante de temps est 80 ms et qui s'approche de sa valeur limite à 180 ms.

Conclusion pour la perception temporelle et conséquences pour l'analyse des langues sifflées

Si l'on tient compte de tous ces facteurs perceptifs temporels on conclut que la résolution temporelle phonétiquement significative sera entre 10 ms et 30 ms et sera la même pour à peu près toutes les fréquences. L'acuité temporelle varie en fonction de la durée des sons impliqués. Jusqu'à ce jour, les adaptations non linéaires de la réponse du nerf auditif et l'amélioration des performances par apprentissage ont compliqué les tentatives de détermination d'un modèle temporel des réalisations phonétiquement pertinentes.

D'autre part, le temps de résolution temporelle du système auditif est de l'ordre de 6 à 8 ms pour des signaux sinusoïdaux dont le signal sifflé est un parent proche. Cette précision accrue justifie l'intérêt de faire une étude des régularités temporelles encodées dans ces systèmes linguistiques particuliers. Dans cette perspective, les résultats que nous avons présentés dans les paragraphes précédents permettront de préciser la validité des mesures réalisées.

4.2.2.3. Conclusion sur la psychoacoustique et sifflements

Le système auditif réagit comme un banc de filtres passe-bande se chevauchant et dont la largeur de bande croît avec la fréquence. Une telle organisation favorise la perception relative des amplitudes et des fréquences. Les sifflements se situent dans une zone où la largeur du filtre auditif varie entre 120 et 500 Hz. Ils n'activent que peu de filtres auditifs à un instant donné, car d'après nos mesures, la largeur de bande sonore qui émerge du bruit est d'environ 400 Hz à courte distance et 150 Hz à 550m. Cette précision a pour conséquence de réduire les possibilités de masquage et c'est une raison pour laquelle un sifflement à une bonne résistance au bruit.

De plus l'acuité temporelle de l'audition est sensiblement la même à toutes les fréquences, mais elle varie en fonction de la durée des évènements sonores. Elle est plus précise pour les sons sinusoïdaux et peur être améliorée par apprentissage. Enfin, la zone pour laquelle l'audibilité et la sélectivité de l'audition humaine sont les plus performantes se situe entre 1 et 4 kHz et 75-120 dB. Le sifflement se situe là encore, dans les zones de perception fréquentielle les plus efficaces de l'oreille. Si l'on considère le signal sifflé au niveau de l'émetteur, la zone de perception la plus précise en termes d'intensité est également concernée.

On peut donc dire que l'observation de la réaction du système auditif à des sons purs ou des éléments simples extraits de la parole est un premier pas permettant déjà de préciser le protocole d'analyse du signal de parole et de faire émerger l'adaptation spectaculaire des langues sifflées à l'audition.

4.2.3. Structuration perceptive du flux de parole

Nous avons vu que la perception des sons est organisée suivant plusieurs dimensions ou attributs de la perception (intensité, hauteur, durée principalement). Cependant le signal de parole naturelle n'est pas perçu comme un ensemble d'entités distinctes à chaque instant, mais plutôt comme un flux organisé à plusieurs échelles (segments-syllabe-mots-phrases) et entrecoupé par des discontinuités. C'est pourquoi l'analyse d'éléments segmentaux ou tonaux considérés comme isolés risque d'exclure un grand nombre de mécanismes mis en oeuvre au quotidien lors de la production et du traitement perceptif de la parole.

Deux domaines de la recherche actuelle, a priori différents, adoptent une perspective globale multiéchelle sur le problème de la perception du continuum sonore d'une phrase : il s'agit d'une part l'*« analyse de la scène auditive »* (Bregman, 1990) qui a permis d'expliquer comment notre cerveau scanne l'espace auditif et en tire des informations sur l'origine des sons, d'autre part de *l'analyse de la prosodie* développée en linguistique pour comprendre la mélodie et le rythme du langage. L'usage à distance des langues sifflées a dirigé notre étude vers le premier domaine et l'usage d'une bande unique de fréquence pour la parole nous a suggéré un rapprochement avec la prosodie. Il s'est avéré que les deux domaines éclairent de façon complémentaire les relations perceptives fondamentales qui forment la base de l'intelligibilité des langues sifflées.

4.2.3.1. Attributs de la perception et « Analyse de la scène auditive »

4.2.3.1.1. Organisation de la scène auditive: origine

A l'écoute d'une phrase dans un environnement bruyant, notre experience est bien différente de la réception d'un ensemble de sons dont les fréquences et les amplitudes varient dans le temps. Nous structurons plutôt le monde sonore en entités cohérentes que l'on peut (i) détecter, (ii) séparer, (iii) localiser, et si possible (iv) identifier. La notion d'intelligibilité d'une forme sonore s'appuie sur l'existence d'aptitudes perceptives auditives permettant de réaliser ces tâches vitales au quotidien. *« L'oreille a développé ses capacités dans un monde où les sons sont presque tous d'origine mécanique: elle procède à des enquêtes, elle essaye de débusquer le mode de production physique qui est -ou pourrait être - à l'origine du son [...] elle effectue des inférences complexes à partir d'indices subtils pour évaluer la direction de la source sonore, sa distance et l'intensité émise à sa source. L'audition organise le complexe sonore qui lui parvient, elle sépare ou regroupe les éléments constituants, elle y distingue des « voix » ou des « images » des sources sonores différentes »* (Risset 1994, p. 103).

Les modes de groupements sont d'autant plus importants pour la description des langues et de tout phénomène sonore qu'ils soulignent que ce ne sont pas directement les paramètres physiques du signal qui importent pour le récepteur mais la relation qu'ils vont avoir lorsqu'il les perçoit. Cet aspect fondamental est exploité par les langues humaines et sans lui les langues sifflées ne pourraient pas transposer une langue.

4.2.3.1.2. Lois de regroupements et images acoustiques

Introduction

Dès lors, un des défis majeurs est de comprendre comment les caractères physiques se manifestent sous la forme de relations perceptuelles. Selon Helmholtz (1821-1894), l'observateur met en oeuvre des lois perceptives qui donnent naissance à l'interprétation la plus efficace de l'environnement. Ce type de raisonnement peut expliquer pourquoi les philosophes de la Gestalt ont élaboré des lois fondamentales d'organisation de la perception (en particulier Wertheimer). Ce sont les lois dites (a) de proximité, (b) de similitude, (c) de continuité et (d) de clôture, qui sont présentées sous leur forme appliquée à la vision sur la Figure 89.

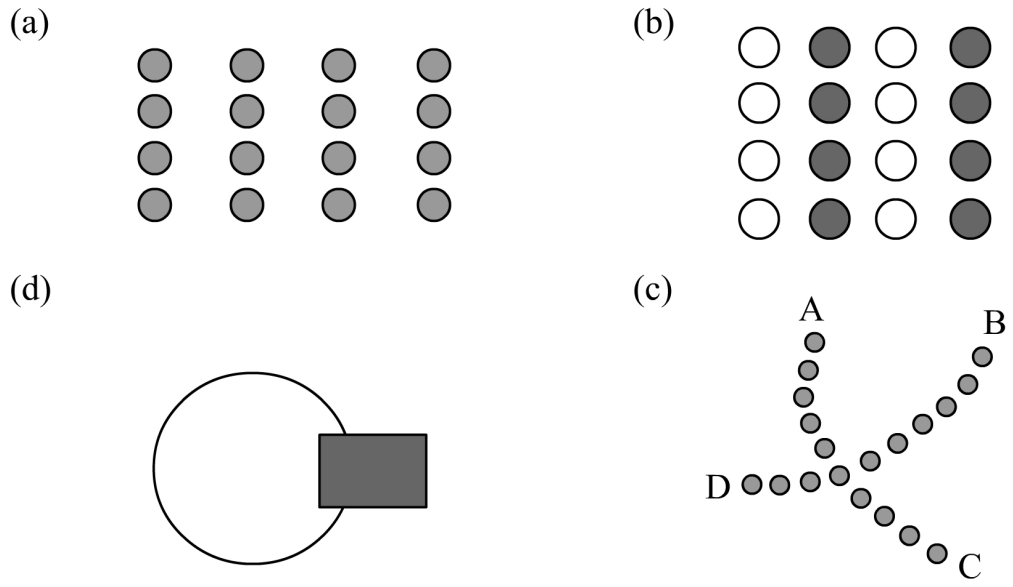

Figure 89: Illustration des lois d'organisation perceptive énoncées par les psychologues gesthaltistes (a) proximité, (b) similitude, (c) continuité, (d) clôture (reproduction de Deutsch 1982)

Ces lois s'appliquent s'appliquent également à l'audition : (a), (b) et (c) favorisent les groupements, (d) la segmentation (Deutsch 1982).

D'après Bregman (1990), les séparations et les regroupements sont les processus élémentaires à la base de la faculté d'« analyse de la scène auditive », dont la fonction est de faciliter l'extraction d'informations sur l'origine des signaux dans un monde sonore complexe. Cette tâche est nécessaire parce que des sons indépendants issus de nombreuses sources arrivent à chaque instant à l'oreille. L'auditeur n'a accès qu'à un mélange continu d'ondes. Comme Helmholtz le faisait remarquer, il y a 150 ans, notre oreille est presque dans la même situation qu'un oeil qui regarderait un point précis de la surface de l'eau, et observant l'ensemble des perturbations à sa surface, devrait analyser l'origine des ondes qui la composent.

Or, comme l'emplacement et les caractéristiques physiques des objets changent relativement lentement par rapport à la célérité des vibrations des sons, deux évènements sonores à la fois proches dans le temps[92] et dont d'autres dimensions perceptives sont communes (hauteur, intensité[93], timbre...) ont de bonnes raisons d'avoir été produits par la même source (Bregman 1990). La localisation de la position de la source est également souvent un critère important[94].

Image acoustique et flux

D'après Bregman et McAdams (1979), l'analyse de scène réalisée permet de dresser une *image acoustique* stable de chaque source sonore de l'environnement.

Dimension verticale

Suivant une dimension « verticale » de cette image, les éléments spectraux, jugés comme issus de la même source sonore sont fusionnés en un événement sonore complexe (phénomène de *groupement simultané* comme par exemple pour le timbre de la voix). Cette catégorie de regroupements peut être issue de plusieurs types de relations:

(1) Coordination par synchronisation spectrale : le timbre en est l'illustration grâce à la cohérence temporelle de son attaque, qui participe à son établissement (Bregman & Pinker 1978). La structure consonne-voyelle exploite ce phénomène en variant les attaques mais en établissant toujours au final une cohérence temporelle. Il n'est donc pas étonnant que les langues sifflées sans tons traduisent ce type d'évènements par une continuité de modulation entre la consonne et la voyelle[95].

(2) Coordination par microvariations communes : Chowning a mis en évidence que l'audition s'appuie sur le destin commun de composants fréquentiels différents: elle associe les composants qui ont des microvariations synchrones (par exemple des composantes affectées de « vibrato » similaires, ou alors des composantes se déplacant ensemble dans l'espace). Dans ces cas, l'enveloppe spectrale relate les changements de dynamique de sa source sans être affectée par la propagation et l'atténuation des sons. La perception humaine de la perspective auditive exploite cette particularité (Chowning 2000).

[92] À l'échelle de la perception de l'oreille.

[93] Évolution de l'enveloppe spectrale entre autres.

[94] Elle est mesurable par le système auditif, d'une part grâce aux différences temporelles (ITD) et d'intensité (IID) qui existent pour chaque onde qui parvient successivement à une oreille puis à l'autre, d'autre part au niveau du pavillon entre les ondes directes et les ondes réfléchies sur les différentes cavités de l'oreille externe (Batteau, 1967).

[95] Principalement dans les langues utilisant une transposition s'appuyant sur ces segments. Notre typologie montre que lorsque leur rôle informatif décroît, cette modulation perd progressivement de l'importance en sifflements.

Dimension horizontale

Au niveau « horizontal » de la métaphore de l'image auditive, ce sont des flux temporels continus qui sont formés. La continuité peut émerger de plusieurs indicateurs perceptuels:

(1) la proximité fréquentielle et la proximité temporelle ont été les plus étudiées (Van Noorden 1975),

(2) la similitude de la qualité spectrale des sons permet également des regroupements en flux (Wessel 1979). Ce dernier phénomène est particulièrement important puisqu'il souligne la possibilité d'associer des évènements sonores en fonction de leur timbre: il est à la base de l'intelligibilité des voyelles d'une langue.

Conséquences pour la fission d'évènements

La création de flux perceptifs a des conséquences importantes au niveau des relations entre les évènements acoustiques qui les constituent: par exemple les coordinations temporelles entre deux evènements faisant partie de deux flux différents sont impossibles (Bregman et Campbell, 1971). Cette dernière propriété est constament utilisée dans l'élocution d'une langue.

Par opposition à la fusion en flux, il existe donc des phénomènes de ségrégation de flux ou fission. Ils apparaissent surtout quand les éléments à l'origine de la séquence perçue sont notablement différents en fréquence, en spectre, en temps de montée de l'enveloppe ou en localisation spatiale.

4.2.3.1.3. Conclusion

Les mécanismes de groupement interagissent entre eux de manière complexe: chaque image auditive possède des attributs perceptifs qui ne sont pas le résultat simple de la somme des évènements perceptifs qui la constituent. A chaque instant, l'image possède une dynamique qui dépend du contexte, ce qui entraîne des évolutions où des transferts sont effectués entre attributs perceptifs (Bregman et Pinker, 1978). Les caractéristiques de cette dynamique vont permettre d'encoder des formes perceptives. Pour cela, un facteur important est la possibilité d'encodage, suivant une dimension donnée, en présence de changements suivant d'autres dimensions (indépendance par rapport aux autres dimensions). Le facteur temporel a donc une grande robustesse pour l'encodage de formes. On dira également qu'une dimension donnée est un bon candidat pour véhiculer des formes perceptives auditives si un grand nombre de possibilités de configurations peuvent être codées suivant cette dimension. Par exemple, nos capacités d'encodage et de perception des fréquences relatives vont permettre le développement de lignes mélodiques variées suivant la dimension fréquentielle. Les langues sifflées exploitent tous ces aspects.

A travers de telles relations la perception auditive bouleverse parfois la structure acoustique d'origine. Ces phénomènes ont des implications directes sur la perception de la parole.

4.2.3.2. Prosodie et langues sifflées

4.2.3.2.1. Introduction : l'étude de la parole continue en linguistique

La notion de prosodie est très ancienne en linguistique, puisqu'elle fut utilisée dans la philologie classique puis trouva un nouvel essor dans la description comparative des langues indo-européènnes. Elle fut finalement intégrée dans les diverses tentatives de théorisation des écoles de linguistes du XX^{ème} siècle (Jakobson, 1931; Pike, 1954, ...). Son usage expérimental fait souvent l'objet d'une définition assez vague comparée aux autres concepts de la linguistique car les paramètres physiques étudiés cachent la méthodologie d'étude que sous-entend cette notion.

Sur ce point, l'étymologie du terme « prosodie » fait référence à un lien entre musique et parole: « ôdê », veut dire chant en ancien grec ou plus précisément chant accompagné d'une musique instrumentale. Il a ensuite été utilisé pour désigner la « *science de la versification »* ou *métrique*, qui gouverne la voix humaine lorsqu'elle est en train de lire de la poésie. D'après l'encyclopédie Universalis la définition du terme est : *« prosodia: accent, quantité dans la prononciation »* ou *« Règles concernant les rapports de quantité, d'intensité, entre les temps de la mesure et les syllabes de la musique vocale »*. L'analyse de la prosodie suppose donc un point de vue synthétique, elle invite à observer les paramètres physiques du signal et leurs attributs perceptifs à différentes échelles (microprosodie à l'échelle du phonème, macroprosodie à l'échelle de la phrase). Elle s'appuie à la fois sur les résultats de la phonétique et de la phonologie pour considérer des phénomènes relatifs qui ne sont pas liés à un phonème particulier mais plutôt à l'enchaînement des segments entre eux (éléments suprasegmentaux). C'est pourquoi les notions d'accentuation, de mélodie, de durée relative et de rythme sont étudiées en linguistique dans le cadre de l'analyse de la prosodie.

4.2.3.2.2. Les composantes de la prosodie

La méthode utilisée pour observer ces éléments consiste à considérer les variations d'intensité et de hauteur (fréquence et amplitude perçues) ainsi que les durées successives des segments syllabiques (Calliope 1989). La complexité des phénomènes observés explique la difficulté qu'ont eu les chercheurs à cerner les dimensions acoustiques les plus pertinentes à analyser. Suivant une approximation courante, ces éléments sont essentiellement abordés à travers l'étude conjointe des variations de la fréquence de phonation et de la courbe d'amplitude du signal de parole. C'est pourquoi ces deux composantes physiques du signal de parole sont les composantes principales de l'analyse de la prosodie :

-L'amplitude représente l'énergie sonore du signal à chaque instant,

-La fréquence fondamentale (F0) et correspond à la fréquence de vibration des cordes vocales (par exemple, en Français, le F0 augmente à la fin d'une phrase interrogative).

De nombreuses études considèrent que l'analyse du Fo seul est représentative de la prosodie d'une langue.

Trois autres composantes d'ordre prélexicales sont parfois utilisées (Ramus 1999):

-La structure syllabique dont l'importance varie en fonction de la phonologie de la langue,

-La structure Consonne/Voyelles (CV) qui est un facteur commun a toutes les langues ;

-La structure phonémique qui est liée au fait que certains types de sons n'existent pas dans toutes les langues : on rencontrera des sons glottalisés en turc ou en chepang, mais pas en grec. Par ailleurs des sons que deux langues ont en commun peuvent tout de même servir à les discriminer, s'ils ont des distributions très différentes dans les deux langues.

Ce regroupement en cinq composantes n'est rarement utilisé à l'heure actuelle en linguistique mais nous verrons que les langues sifflées soulignent le rôle important des composantes prélexicales.

Information portée par ces composantes prosodiques

L'information de ces composantes se situe à trois niveaux différents :

-L'intonation, ou l'enchaînement des tons dans les langues tonales, sont fournis par la fréquence fondamentale et l'amplitude,

-Le rythme: il provient de plusieurs paramètres comme la fréquence fondamentale, l'amplitude, la structure syllabique[96] et la structure CV.

-La phonotactique est donnée par l'enchaînement des phonèmes.

4.2.3.2.3. Langues sifflées : une approche pragmatique et naturelle de la prosodie

Un regard alternatif sur la notion de prosodie

L'analyse de la prosodie a donc développé une manière propre à la linguistique d'analyser le flux sonore dans une parole continue. Cependant, en raison de sa méthodologie reposant avant toute chose sur le Fo, une approche prosodique d'une langue sifflée fera émerger des différences avec le même type d'analyse réalisé sur la version parlée de la même langue. Ceci sera particulièrement visible sur les langues sans tons et apparaîtra également dans les langues à tons qui utilisent quelques éléments spectraux différents du Fo comme base du sifflement. En effet les langues sifflées considèrent naturellement que la HB peut porter des éléments à valeur prosodique. La *« prosodie interne »* des éléments spectraux des consonnes et des voyelles n'est donc pas à négliger, du point de vue de la perception et de la production des siffleurs.

Par conséquent, s'il est vrai que chaque langue sifflée propose une description prosodique d'une langue à partir d'une seule bande étroite de fréquence, ses moyens de l'obtenir et les critères que les siffleurs retiennent sont différents de l'approche développée par la tradition scientifique: alors que l'analyse de la prosodie s'appuie systématiquement sur la fréquence fondamentale, la description naturelle des langues sifflées, quant à elle, compose avec les différents éléments de la voix soulignant ainsi que les composantes « prosodiques » les plus pertinentes pour l'intelligibilité de chaque langue diffèrent en fonction de la structure de la langue.

[96] Les langues humaines ont une rythmicité plus ou moins accentuelle ou syllabique ce qui fait que suivant la langue la syllabe jouera un rôle différent dans le rythme.

4.2.3.2.4. Collusion entre l'analyse de la prosodie et l'analyse de la scène auditive

La prosodie est connue comme un facteur améliorant l'intelligibilité de la parole. Pour ce faire, elle s'appuie sur tout un ensemble de paramètres qui ont leur origine dans la faculté d'analyse de la scène auditive. Nous en détaillons ici les points les plus importants.

Traquage de la voix d'un locuteur

Les propriétés perceptives qui permettent de reconnaître les variations prosodiques sont soumises à l'identification préalable d'un flux de parole issu d'une seule et même source: le locuteur. Cette tâche est réalisée en grande partie par l'analyse de la continuité de Fo. La cohérence des variations du Fo permet à un auditeur de continuer à décoder une voix particulière, même pendant les périodes non voisées ou de silence lors desquelles le signal acoustique est souvent en compétition avec celui d'autres voix ayant des fréquences fondamentales différentes (Nooteboom et al, 1978).

Continuités spectrales intra mot

Tous les autres indices permettant d'identifier des sons cohérents viendront renforcer l'analyse de la prosodie. Ainsi, à l'échelle du mot ou de la phrase l'analyse des continuités de Fo est complétée par une analyse des continuités des concentrations energétiques portées par les harmoniques (Dorman et al, 1985). A ce niveau, le rôle de la synchronisation des attaques lors de la fusion des enveloppes spectrales a été étudiée spécifiquement pour les sons de la parole. Par exemple, Darwin (1981) a montré que des sujets ayant écouté de manière décalée dans le temps, d'abord les harmoniques constituant le premier formant d'une voyelle auxquelles ont été ensuite ajoutées les harmoniques du deuxième formant, ont souvent identifié individuellement ces deux entités, ce qui a modifié la perception de la voyelle d'origine. Une telle identification séparée de groupes d'harmoniques est bien plus difficile si les deux groupes débutent simultanément. Les consonnes servent donc en partie à synchroniser les éléments spectraux de la voyelle. Une conséquence importante de ce phénomène est de lier fortement la consonne et la voyelle en un même flux auditif. La réalité perceptive de la syllabe en est renforcée (d'autant plus que la perception de la synchronie n'implique pas une identification consciente)[97]. Cela se traduit dans les langues sifflées sans ton par une modulation continue et unique de la voyelle avec la consonne.

[97] Bien que les phonèmes soient parfois les plus petites unités distinctives permettant de distinguer deux mots, ils ne sont pas perçus séparément: les temps de réaction des sujets lors d'expériences de psycholinguistique testant la perception de phonèmes cibles le montrent : ils sont plus courts pour des syllabes ou des mots que pour des phonèmes ce qui suppose un traitement à plus haut niveau pour les phonèmes (Savin et Bever, 1970; Segui 1988)).

Continuité temporelle dans une phrase

Analyse spectrale: évolution temporelle de l'énergie sonore

D'après plusieurs auteurs, le système auditif serait capable d'estimer la cohérence de l'évolution de l'énergie au cours du temps dans un même flux sonore (Botte 1989). Ce facteur est un des éléments clef de la perception de la continuité temporelle de la prosodie d'une phrase. En effet, les variations d'intensité permettent de déterminer la position de l'accent lexical propre à de nombreuses langues, ce qui conditionne en partie le rythme d'une langue[98]. En général, l'enveloppe d'amplitude joue un rôle non négligeable de marqueur de contrastes distinctifs qui donnent une cohérence rythmique d'ensemble au signal. C'est pourquoi un bruit blanc modulé en amplitude par l'enveloppe d'un signal de parole peut être perçu comme un mot (Katz et Berry 1971). Or, Chowning a montré que des indices sonores ayant une intensité et une répartition spectrale cohérentes dans le temps avaient toutes les chances d'être issus de la même source et donc d'être groupés ensemble. Il a aussi observé que l'évaluation de la distance, grâce à la comparaison de l'intensité du signal direct par rapport aux signaux réverbérés, renforce l'estimation de cette cohérence et améliore donc la perception de l'évolution temporelle de l'enveloppe d'amplitude du signal[99]. Ces dimensions psychoacoustiques sont la base de la perception de la perspective auditive: *« auditory perspective is not a methaphor in relation to visual perspective, but rather a phenomenon that seems to follow general laws of spatial perception. It is dependent upon loudness (subjective !) whose physical correlates we have seen to include spectral information and distance cue, in addition to intensity »* (Chowning 2000, p.5).

Dans le cas des langues sifflées la tâche de détection de la cohérence temporelle par l'auditeur est simplifiée par l'usage d'une seule bande de fréquence dont la cohérence spectrale peut être évaluée grâce à sa largeur de bande dans la partie qui émerge du bruit.

Durées

Au niveau des phrases qui forment des séquences sonores longues, tous les phénomènes décrits précédemment favorisent la résistance de l'intelligibilité à la dégradation du signal. Huggins (1975) a réalisé des tests sur de la parole segmentée temporellement. Il a montré deux phénomènes révélateurs sur l'organisation temporelle de la phrase: (a) la parole interrompue par des silences de durée fixe (200 ms) atteint un taux d'intelligibilité de 90% si les intervalles de parole restants sont supérieurs à une durée de 175 ms. Ce qui correspond à la durée moyenne d'une syllabe. (b) D'autre part, si les intervalles de parole sont de 60 ms, l'intelligibilité atteint 90% pour des intervalles de silence inférieurs à 70 ms. Ce qui correspond à une durée tout juste supérieure à la durée moyenne d'une consonne de la voix parlée. Dans certaines situations de

[98] La rythmicité d'une langue est en partie portée par l'accent. Ce facteur se combine avec le rôle de la syllabe pour définir une rythmicité plus ou moins accentuelle ou syllabique en fonction de la phonologie de chaque langue (Ramus 1999).

[99] Les expériences réalisées par Chowning étaient à courte distance (50m) dans un milieu intérieur, la conservation de l'énergie issue des réverbérations est moins complète pour les langues sifflées en raison de la déperdition d'une grande partie de l'énergie sonore par reflexion.

dégradation intermédiaire, il a également montré que les sujets sont capables de discriminer certaines transitions ou certaines voyelles mais la parole reste inintelligible. Ceci confirme qu'une certaine continuité rythmique est nécessaire à l'intelligibilité de la parole et que les consonnes y jouent un rôle clef. Au niveau de la transition consonantique, cette continuité est rompue pour une durée de silence supérieure à 100 ms (Nooteboom, 1978).

4.2.3.3. Conclusion

L'analyse de la scène auditive et l'analyse de la prosodie nous ont permis de comprendre la base des relations qu'entretiennent les attributs de la perception dans le cadre d'un continuum de parole. Les points communs que nous avons dégagés entre ces deux domaines de recherche pour les besoins de l'analyse du sifflement linguistique s'appliquent à tout phénomène de langage. Une comparaison avec les résultats des travaux de recherche réalisés par ailleurs sur la parole dans le bruit est maintenant possible : ceux-ci concluent que les paramètres les plus utilisés par les personnes en condition d'écoute difficiles (rapport signal sur bruit faible, et/ou effet cocktail party, et/ou reverbération en salle) sont les suivants : (a) l'amplification sélective et l'amélioration du rapport signal sur bruit au niveau périphérique, (b) la séparation spatiale du locuteur et la perception de la perspective auditive (liées à la localisation des sons), (c) la dépendance temporelle du son ciblé (liée à la continuité rythmique), (d) l'identification et la poursuite des attributs de hauteur de la voix du locuteur (avec un suivi dynamique grâce à la perception de la continuité soit du Fo soit des concentrations spectrales d'énergie) (Bronkhorst 2000). On retrouve les éléments les plus saillants du signal de parole à l'échelle de la phrase qui sont mis en évidence par l'analyse de la prosodie et qui sont expliqués par l'analyse de la scène auditive. Toutes les recherches sur la parole considérée comme un flux sonore convergent donc vers les mêmes conclusions. Il n'est pas étonnant que les langues sifflées mettent particulièrement en valeur ceux de ces facteurs clefs de l'intelligibilité de la parole qui correspondent le mieux à la réalité acoustique d'un sifflement.

Un autre point a pu être souligné : les langues sifflées apportent des éléments de réflexion pratiques pour l'analyse de la prosodie puisqu'elles privilégient les éléments porteurs des relations les plus marquantes de la structure linguistique plutôt que seulement le Fo. Ce type d'approche est perceptivement justifié, en effet, chaque être humain est capable de détecter l'information la plus saillante dans le signal, selon la définition de Hombert & Maddieson (1998), à savoir les traits segmentaux et suprasegmentaux qui sont non seulement les plus identifiables du point de vue acoustique, mais aussi les plus discriminatoires des langues. En quelques sorte, les langues sifflées proposent une nouvelle *méthodologie naturelle d'analyse de la prosodie*, pertinente car elle repose sur plusieurs générations de siffleurs et donc de cerveaux entraînés depuis l'enfance à une pratique héritée d'un processus évolutif. Dans ces conditions, on peut penser que l'analyse du rôle prosodique du timbre a un grand avenir. En effet, en termes de hauteur le timbre a une réalité perceptive (Carlson et al 1970, Bladon et Fant 1978) dont les conséquences prosodiques valent le détour pour les langues non tonales (proéminence de l'accent par exemple). Si ce n'était pas le cas les siffleurs de Turquie, de Grèce et de la Gomera ne pourraient pas en faire un élément clef de leur transposition.

4.3. Intelligibilité de la parole sifflée: analyse progressive

Nous allons maintenant présenter l'ensemble des processus impliqués dans l'intelligibilité des langues sifflées tels qu'ils sont connus à l'heure actuelle et sans en éluder la complexité. Nous nous limitons aux langues sifflées sans ton car ce sont celles que nous avons analysées le plus en détail. Nous avons organisé cette troisième partie du chapitre en une série progressive de sections. Dans un premier temps nous présentons les résultats d'une analyse de psycholinguistique portant sur l'identification de voyelles de Silbo par des sujets Français qui ne connaissaient rien au sujet des langues sifflées. Ensuite nous rappelerons les performances de reconnaissance sur des non-mots des sujets testés expérimentalement en turquie dans les années 70. Puis nous passerons à l'analyse de la perception liée au contexte lexical, au niveau des mots et de leur contenu. Ce n'est qu'après ce stade que nous rentrerons plus avant dans la réalité rencontrée par les siffleurs avec l'analyse de l'intelligibilité des phrases. Enfin nous décrirons les résultats préliminaires d'une analyse des conséquences phonétiques de la dégradation naturelle de la parole sifflée à plusieurs distances.

4.3.1. Expérience de perception des voyelles sifflées par des sujets ignorant tout des langues sifflées

4.3.1.1. Principes généraux

Afin d'approfondir la compréhension de la perception des voyelles sifflées et leurs liens avec les voyelles articulées de la voix parlée, nous avons mis au point deux variantes d'une même expérience permettant de tester des sujets ne connaissant rien au phénomène des langues sifflées. Notre objectif était en effet de comprendre si ces personnes étaient en mesure de réaliser les mêmes catégorisations de voyelles que des siffleurs à la simple écoute de voyelles sifflées. Pour cela ils devaient effectuer une tâche simple et intuitive pour laquelle à aucun moment ils n'ont pu s'appuyer sur une quelconque correction, traduction ou explication de la répartition réèlle des fréquences des voyelles sifflées.

4.3.1.2. Méthode et corpus

4.3.1.2.1. Choix de la langue sifflée et choix de la langue maternelle des sujets testés.

Nous avons choisi de réaliser cette expérience sur des sujets de langue maternelle Française. Les voyelles sélectionnées étaient /i/, /e/, /a/, /o/ de la langue sifflée espagnole de la Gomera. En effet, ces voyelles existent également en français avec des réalisations similaires à l'espagnol. Une autre raison de ce choix tient au fait que ces quatre voyelles sifflées (ou des réalisations très proches) ont la même répartition fréquentielle à la fois en grec, en turc. Etant donnée la structure de la langue française, on peut raisonnablement penser que si une langue sifflée se développait, ces voyelles se répartiraient de la même manière.

4.3.1.2.2. Matériel sonore et stimuli

Le matériel sonore utilisé comprenait 84 voyelles de la langue sifflée espagnole de la Gomera (Silbo). Elles ont toutes été extraites de l'enregistrement de 20 longues phrases, sifflées relativement lentement en une seule session par une même personne dans les mêmes conditions (distance à atteindre, bruit, technique utilisée). Les 84 voyelles (21 /i/, 21 /e/, 21 /a/ et 21 /o/) ont été choisies suivant des critères statistiques basés sur notre analyse du Silbo. Ainsi, nous avons exclu les voyelles situées à la finale d'une phrase car elles sont souvent marquées d'un abaissement de la puissance du sifflement et nous avons décidé de ne retenir que des voyelles situées dans un intervalle de confiance à 5% de la moyenne des fréquences de chaque bande vocalique. De cette manière les bandes de fréquences des voyelles de l'expérience ne se chevauchaient pas, tout en restant relativement larges.

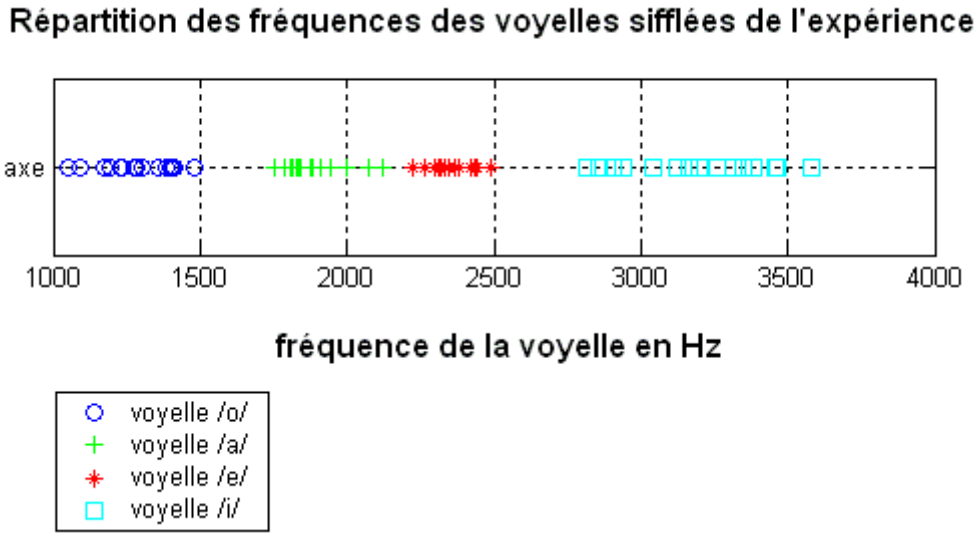

Figure 90 : Les voyelles jouées lors de l'expérience ont été choisies parmi celles de Luis (voir Figure 34)

D'autre part, nous avons décidé de ne pas traiter acoustiquement les sons par des filtrages de réduction de bruit afin de rester dans une situation proche des conditions d'écoute d'un siffleur.

Variante

Afin de tester l'effet du contexte acoustique de la phrase sifflée sur les sujets, un deuxième corpus a été établit comme suit: pour chaque voyelle du premier corpus nous avons conservé le contexte sifflé précédant son occurrence (2 à 3 secondes). Un corpus de 84 phrases sifflées se terminant par la voyelle dont nous voulions tester la perception a donc été constitué. Sur les 84 sons, 20 ont été réservés à la phase d'apprentissage et 64 à la phase de test.

4.3.1.2.3. Type de tâche proposée:

Pour chaque variante, la tâche principale de l'expérience était la suivante: après l'écoute de chaque son, le sujet désigne la voyelle qu'il estime la plus proche de la voyelle sifflée qu'il vient d'entendre en cliquant sur un des 4 boutons « a », « é », « i », « o ». La tâche était donc du type « choix forcé » parmi quatre solutions.

4.3.1.3. Réalisation de l'expérience:

4.3.1.3.1. Programmation

Le programme graphique de l'expérience a été réalisé grâce au logiciel *Flash 5* permettant l'utilisation du langage de programmation de dessin vectoriel *Actionscript*. De cette manière il a été possible de contrôler les évènements déclenchés par l'utilisateur[100], d'organiser les listes de présentation suivant un tirage aléatoire non récurrent et d'enregistrer les données dans des fichiers exploitables par un programme de traitement réalisé sous *Matlab*.

4.3.1.3.2. Sujets testés

Chaque variante de l'expérience a été passée par 20 personnes situées dans une tranche d'âge de 19 - 29 ans. Les personnes ayant passé la variante 1 (voyelle seules) étaient différentes de celles ayant passé la variante 2 (phrases). Un audiogramme a été effectué afin de vérifier qu'ils avaient une bonne audition. Les sujets ne devaient pas avoir été en contact avec une langue sifflée et n'avoir jamais entendu d'explication sur le système de répartition des voyelles des langues sifflées sans tons.

4.3.1.3.3. Déroulement de l'expérience:

Chaque variante comprenait trois étapes:

-Première étape: Questionnaire puis explication succinte de l'expérience suivie de l'écoute d'une phrase sifflée permettant au sujet de se familiariser avec les sonorités du sifflement articulé tout en comprenant qu'il s'agit bien d'une langue naturelle (mais la phrase n'a pas été traduite afin de ne pas donner d'indications sur les voyelles).

[100] Par exemple certains biais d'usage comme la répétition de clics successifs. Les boutons de contrôles non utiles dans l'interface pour la tâche en cours étaient également désactivés.

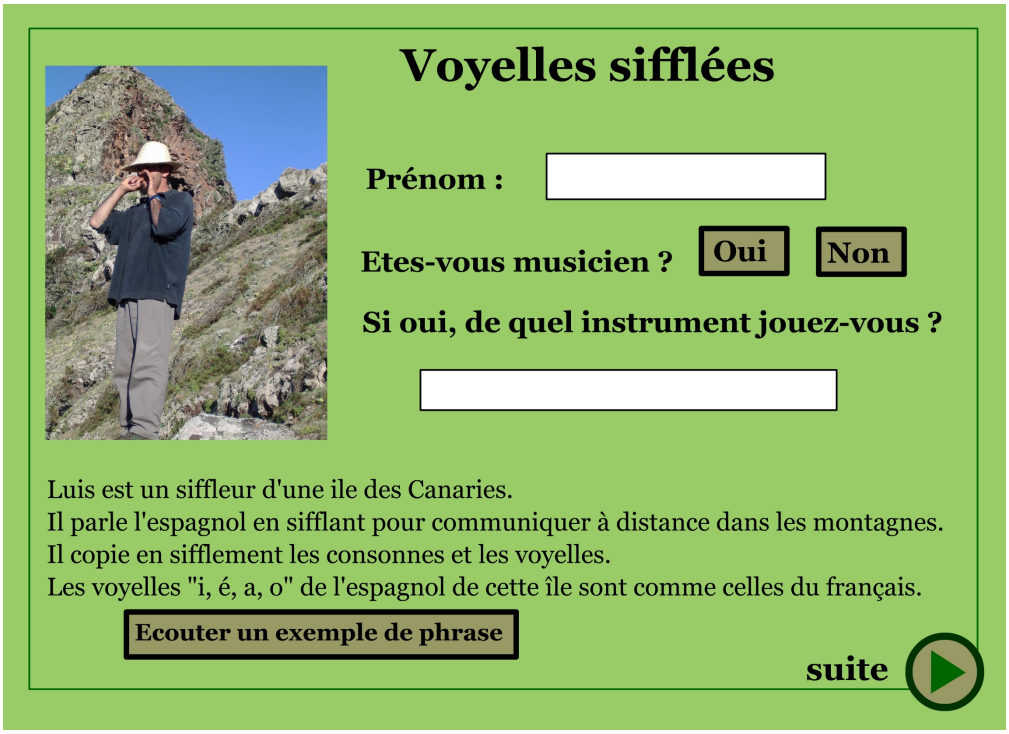

Figure 91 : Questionnaire et familiarisation avec l'expérience

-Deuxième étape: Apprentissage. Le principal objectif de cette étape était de permettre au sujet de se familiariser avec la tâche de test et avec le système vocalique sifflé.

La tâche à effectuer était donc identique à celle du test, seul le nombre de sons présentés changeait. L'étape d'apprentissage comportait 20 sons (5 /i/, 5 /e/, 5 /a/, 5 /o/) se succédant selon un *ordre préétablit* permettant au sujet d'entendre toutes les combinaisons possibles de voyelles succesives différentes.[101]

Lors de l'apprentissage de la première variante de l'expérience (voyelles seules) la session de 20 voyelles a été présentée 2 fois de suite de manière à rapprocher le taux d'exposition sonore avant le test de celui de la deuxième variante.

[101] Ce critère a été retenu car un tirage aléatoire sur le nombre réduit de sons de l'apprentissage présentait parfois tous les sons correspondant à une voyelle donnée dans la première partie de l'écoute. Le risque était que les sujets n'aient pas le temps de se familiariser avec le système vocalique sifflé dans son ensemble. Nous ne voulions pas que les sujets confrontés à ces listes aient l'impression de faire toujours la même réponse au début de l'expérience.

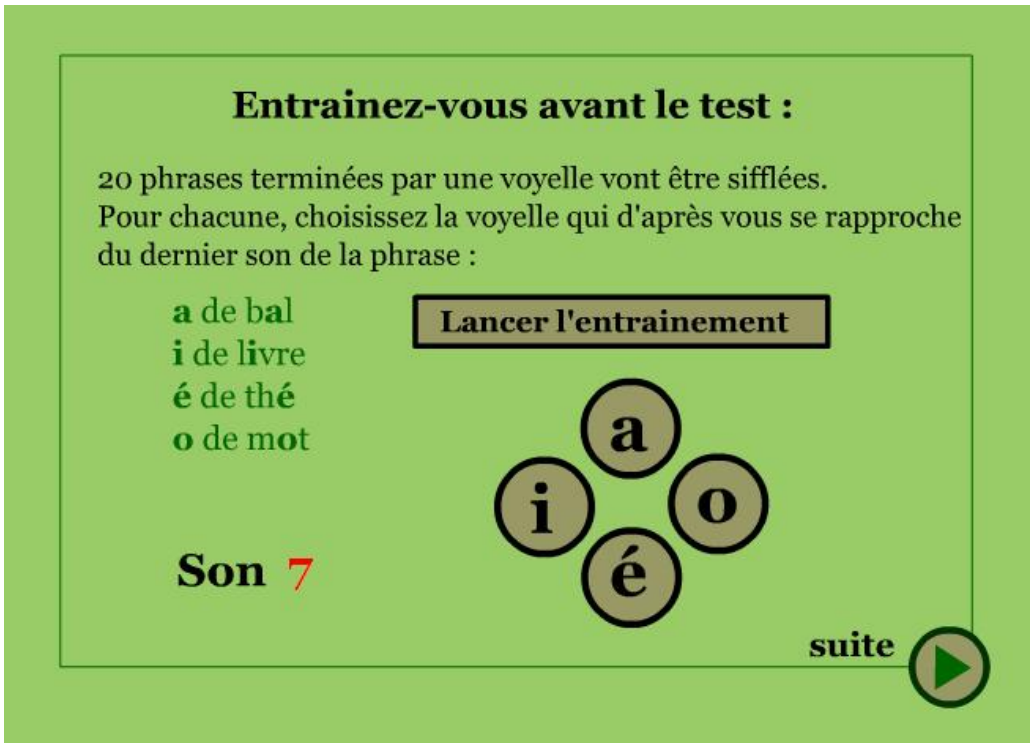

Figure 92 : Capture d'écran du déroulement d'une phase d'apprentissage de la variante 2 de l'expérience

-Troisième étape: Test

Le test proposait la même tâche que pour l'apprentissage, avec 64 voyelles sifflées (respectivement 64 phrases sifflées se terminant par des voyelles variante 2): 16 /i/, 16 /e/, 16 /a/, 16 /o/ se succèdant suivant une liste choisie par tirage aléatoire. Quand le test est terminé, le bouton « Quit » permet d'enregistrer le fichier résultat (Figure 93).

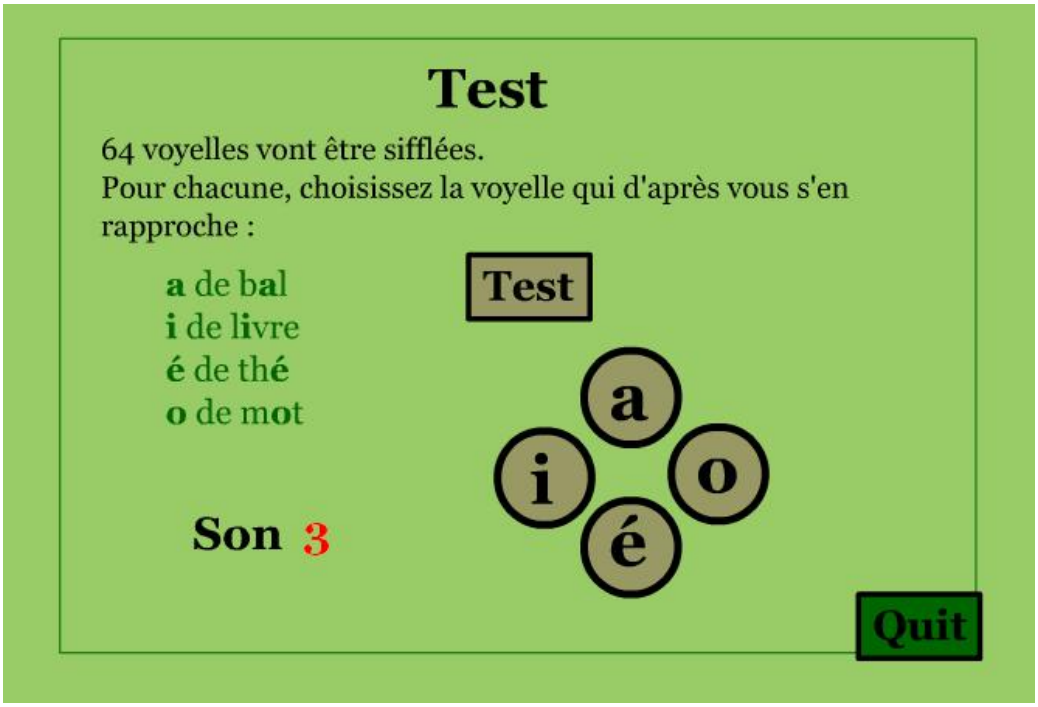

Figure 93 : Capture d'écran du déroulement de la phase de test de la variante 1

4.3.1.4. Résultats

4.3.1.4.1. Généralités sur le traitement des données

Les données collectées par l'interface graphique permettaient de connaître la liste des sons joués et celle des réponses. Ces données ont été analysées pour chaque individu puis ont été mises en commun. Un pré traitement a été effectué sur 5 listes avec *Excell*, puis le traitement de l'ensemble des données a donné lieu au développement d'un programme spécifique sous *Matlab* permettant d'extraire les réponses sous la forme de matrices de résultats et de les représenter sous différentes formes graphiques en réintégrant parfois des informations comme par exemple la répartition en fréquences des voyelles de la Figure 90.

4.3.1.4.2. Résultats de l'expérience de perception des voyelles isolées (variante 1)

Tendance générale

Réponses justes

Le taux de réussite moyen est de 55%. Il correspond au nombre de réponses justes. Compte tenu du protocole expérimental et de la tâche à effectuer ces résultats sont largement au-dessus de la chance qui est à 25 %. Mais les taux moyens de bonnes réponses varient en fonction des voyelles.

Tableau 28 : Taux de réussite moyen de bonne réponses sur 20 sujets

	/o/	/a/	/e/	/i/
Réponses justes en %	50.63	44.06	46.88	78.44

Ensemble des réponses

Si l'on considère maintenant l'ensemble des réponses on observe, outre les bonnes réponses, que la majorité des confusions peuvent être qualifiée de *logiques* en ce sens qu'une voyelle est en général confondue avec ses voisines fréquentielles dans 83% des cas de confusion.

Tableau 29 : Matrice de confusion de l'ensemble des réponses des 20 sujets (résultats en %).

	Voyelles répondues			
Voyelles jouées	/o/	/a/	/e/	/i/
/o/	50.63	40.31	7.50	1.56
/a/	13.44	44.06	31.56	10.94
/e/	5.94	22.19	46.88	25.00
/i/	0.00	4.38	17.19	78.44

La représentation graphique de la matrice sous forme de surface ou sur un plan avec un code couleur progréssif adapté illustre visuellement ces résultats (Figure 94).

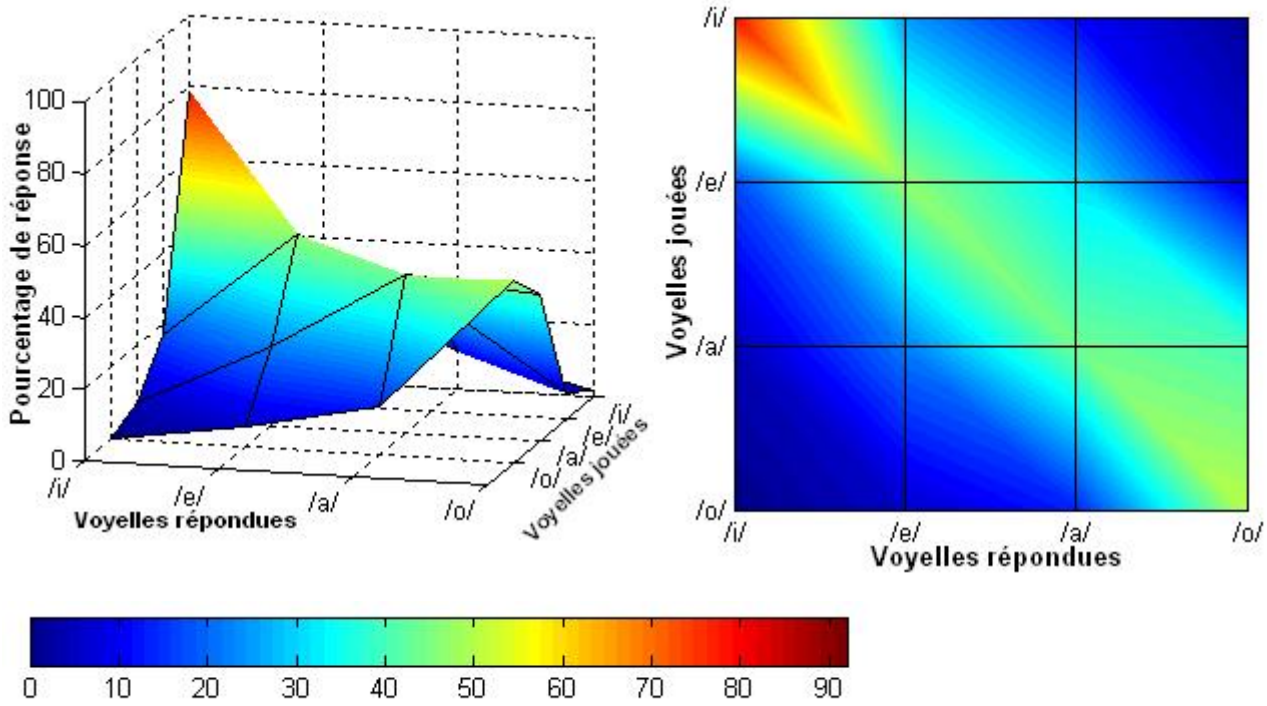

Figure 94 : Performances d'identification des voyelles sifflées seules

Afin de préciser l'influence des fréquences de chaque voyelle sifflée jouée sur les réponses des sujets nous avons également réprésenté les résultats en fonction de la répartition fréquentielle des voyelles de l'expérience (Figure 95). Sur cette nouvelle figure, nous avons représenté également les courbes de tendances des réponses des sujets. Elles permettent de retrouver une projection indirecte (car moyennée par la regression linéaire au second degré) de la forme en selle de cheval de la surface.

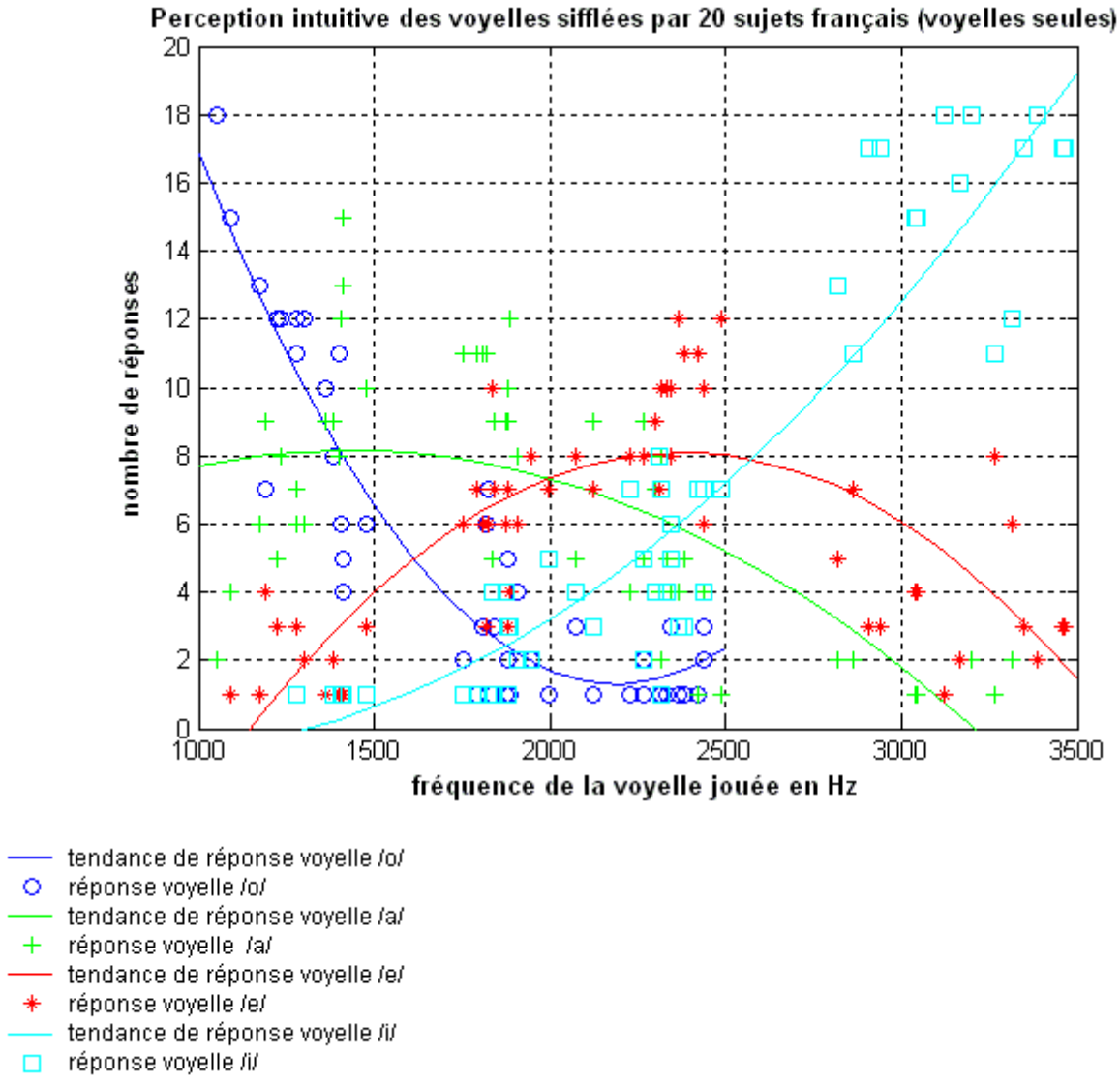

Figure 95 : Répartition des réponses en fonction des voyelles sifflées

Différences individuelles et confusions fréquentes

Considérations générales

Les taux de réussites et de confusion distinguent les sujets (les détails des résultats individuels sont disponibles en *Annexe E*). En effet, deux d'entre eux se détâchent positivement de la moyenne avec un taux de 73,5 %. Un groupe de 6 personnes est au dessus de 40 bonnes réponses sur 64 (62,5%). 4 autres personnes en sont très proches (taux supérieur à 37 sur 64 (58%)). Ce qui veut dire que la moitié des personnes réussit bien le test. Les 10 autres personnes ont toutes un taux de réussite supérieur à 37%. Les 4 moins performantes d'entre elles se situent entre 37 et 40% les 6 autres réussissent mieux le test puisqu'ils obtiennent des performances entre 45 et 54%.

Détails en fonction des voyelles

En général, les moins performants ont tout de même une matrice de confusion logique. Leur taux relativement faible est bien souvent dû à une confusion systématique entre deux voyelles sifflées voisines au niveau des fréquences. La variabilité des performances des sujets est donc grande suivant le type de voyelle :

- Pour le /i/ la grande majorité des sujets ont de très bons taux de réussite puisque 16 d'entre eux ont un score supérieur à 12 sur 16 et 2 d'entre eux ont 100% de réussite. Le moins performant a un score de 9 sur 16 (56%).
- Pour le /o/, 6 personnes un taux d'identification supérieur à 10 sur 16 (62,5%). Toutes les autres personnes prennent souvent le /o/ pour un /a/.
- Le /a/ est la lettre la moins bien identifiée par les sujets car elle est souvent prise pour un /e/ et assez fréquement prise pour un /o/.
- Le /e/ est confondu à part égales avec ses voisins sifflés /a/ et /i/

Les basses performances pour le /a/ et le /e/ s'expliquent par le fait qu'il ont chacun deux voisins perceptifs en termes de hauteur ce qui multiple les possibilités de confusion par rapport au voyelles plus isolées /i/ et /o/. Malgré cela les sujets les plus performants arrivent très bien à les catégoriser comme des voyelles différentes, uniquement à partir de leur fréquence sifflée.

Les confusions les plus fréquentes sont donc les suivantes : le /o/ est souvent pris pour un /a/, le /a/ et le /e / sont souvent pris réciproquement l'un pour l'autre. Cette dernière confusion s'explique en partie par la proximité des deux intervalles de fréquence de sifflement de ces deux types de voyelles.

Différences entre musiciens et non musiciens

Parmi les personnes ayant passé le test 6 étaient musiciennes. Les résultats de ce groupe se distinguent significativement de ceux du groupe des non musiciens ($F(1,18)=6,71$, $p<0,018$). On peut donc dire que les musiciens réussissent mieux la tâche que les non musiciens. Cette différence est clairement visible sur les Figure 96 et Figure 97 (voir aussi les courbes de tendance en fonction de la répartition en fréquence des voyelles jouées en Annexe E.2).

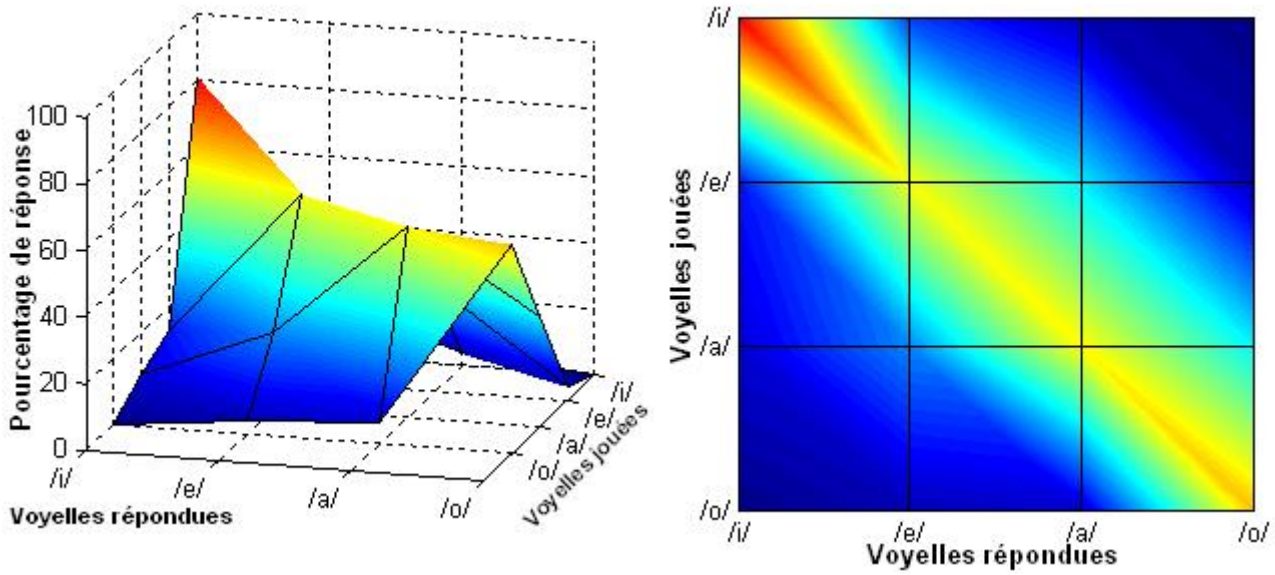

Figure 96 : Performances d'identification des voyelles sifflées seules pour les musiciens

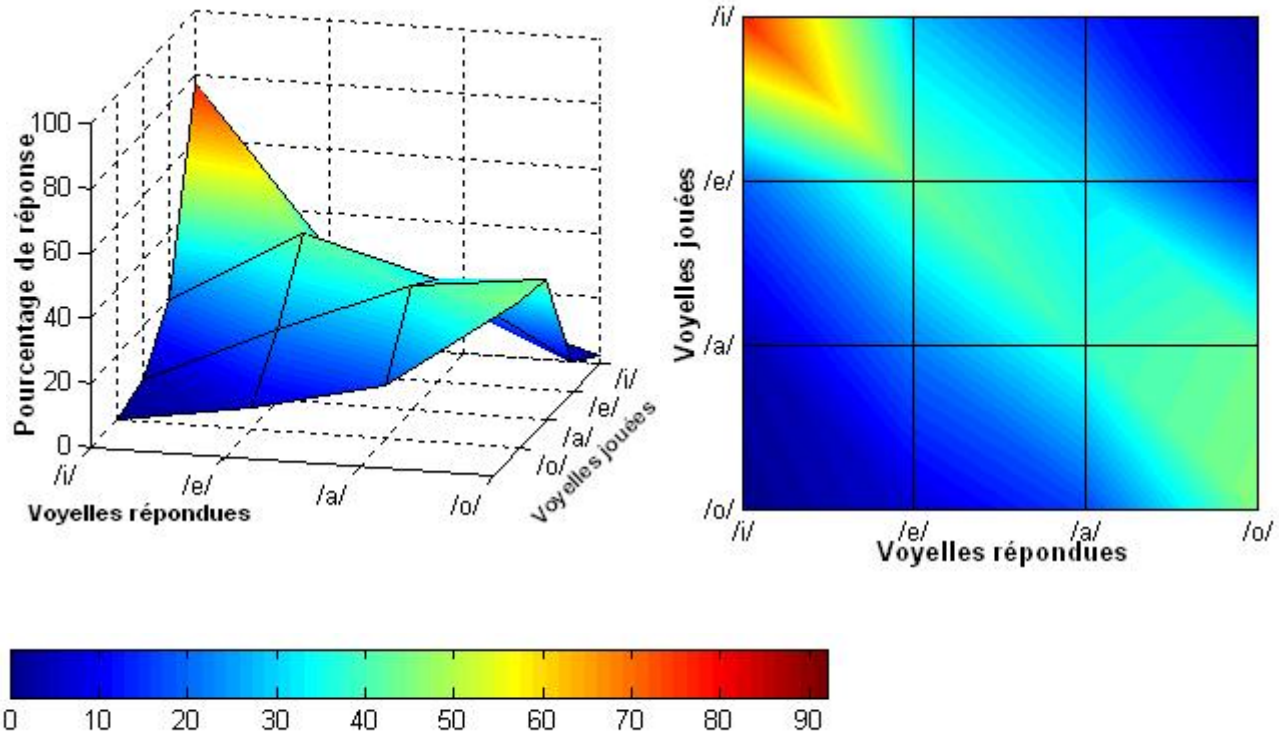

Figure 97 : Performances d'identification des voyelles sifflées seules pour les non musiciens

Conclusion

Les tendances de bonnes réponses et les tendances de confusions montrent que les sujets réussissent en général bien la tâche. Tous ces indices soutiennent le fait que les sujets français catégorisent les voyelles sifflées « a », « é », « i », « o » comme les siffleurs de la Gomera.

Un biais expérimental a pu être observé, qui tient au fait de présenter des voyelles isolées de tout contexte sonore hormis celui de la voyelle précédement écoutée. En effet, certains sujets répercutent des confusions sur la réponse suivante. Par exemple si a l'écoute d'un /e/ sifflé ils avaient répondu /a/ et que la voyelle jouée juste après était un /a/ il ont eu tendance à répondre /o/. Par conséquent, on observe un effet de cascade entre confusions logiques qui s'arrête quand un saut de fréquence conséquent intervient. Ce phénomène fausse les résultats des taux de réussite tout en confirmant que des sujets non-siffleurs étagent bien perceptivement les voyelles sur une répartition qui dépend de la fréquence.

Dans ces conditions, il n'est pas surprenant de constater que les musiciens réussissent mieux la tâche que les non musiciens car ils sont plus habitués à associer les hauteurs perçues de manière isolée à une référence sonore culturellement marquée.

4.3.1.4.3. Résultats pour l'expérience de perception des phrases se terminant par une voyelle (variante 2)

Cette expérience a été mise au point afin de vérifier l'effet du contexte sur la perception des voyelles sifflées par des français. En particulier nous avons émis l'hypothèse qu'en nous rapprochant des conditions d'écoute des siffleurs -qui ne percoivent pas des voyelles seules mais des voyelles sifflées intégrées à un flux sonore sifflé- nous pourrions observer une suppression ou au moins une atténuation de l'effet de répercussion en cascade des confusions logiques. Les voyelles présentées dans cette variante de l'expérience étaient donc précédées de leur contexte sifflé.

Tendance générale

On observe les mêmes tendances générales que lors de la variante 1 mais avec des performances de reussite à la tâche d'identification de 60,2%. Les voyelles sifflées /o/ et /i/ sont encore mieux identifiées que pour la variante 1 (respectivement 73,13% et 87,81%) alors que les voyelles /a/ et /e/ sont légèrement moins bien identifiées. Ces résultats apparaissent statistiquement et visuellement sur la matrice de confusion du Tableau 30 et sur les représentations graphiques en surface et planes de la Figure 98.

Tableau 30 : Matrice de confusion de l'identification par 20 sujets français de voyelles sifflées avec contexte

	Voyelles répondues				
Voyelles jouées		/o/	/a/	/e/	/i/
	/o/	73.13	23.13	2.81	0.94
	/a/	10.94	39.06	39.38	10.63
	/e/	5.00	19.38	40.94	34.69
	/i/	0.31	1.56	10.31	87.81

.

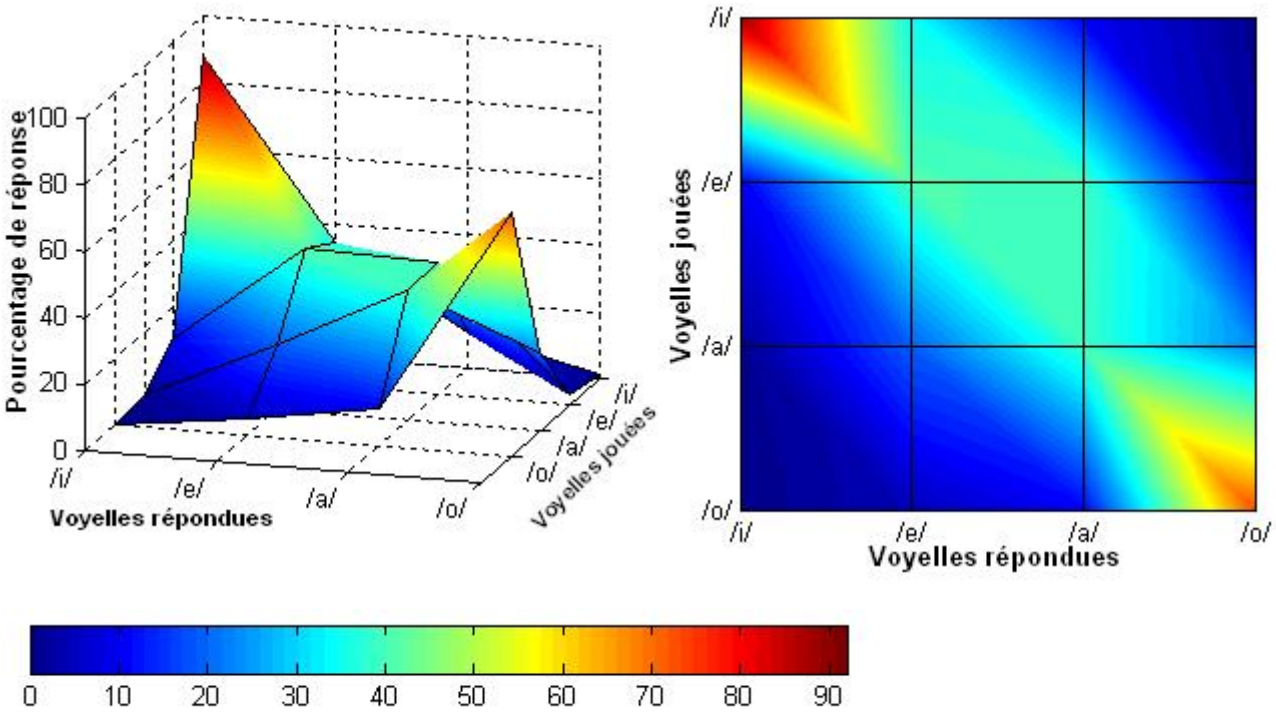

Figure 98 : Performance d'identification des voyelles sifflées avec contexte

Une représentation tenant compte des fréquences des voyelles jouées a été également produite :

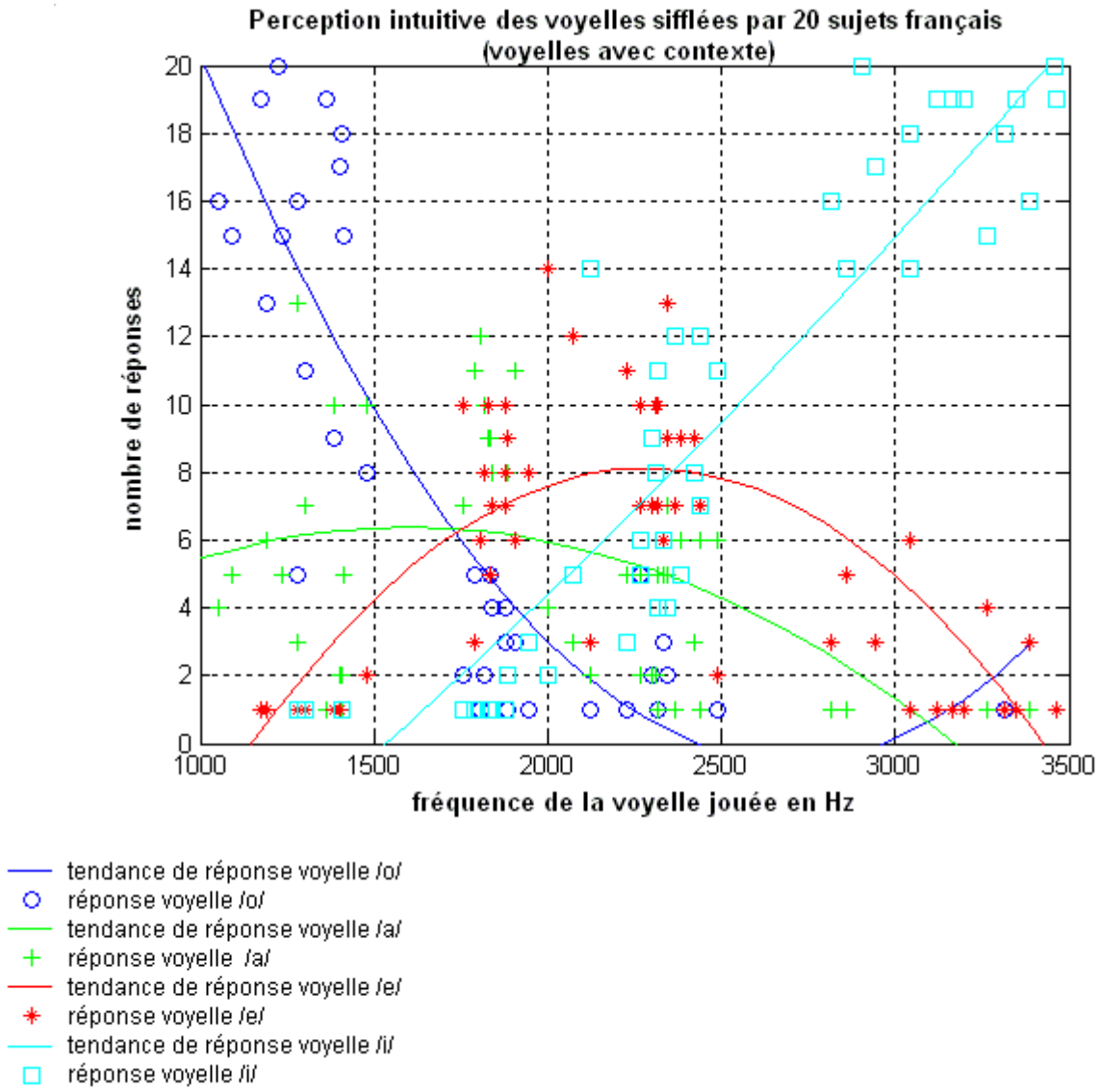

Figure 99 : Répartition des réponses en fonction des fréquences des voyelles sifflées

Différences individuelles et confusions

8 personnes ont un score de réussite supérieur à 40 bonnes réponses sur 64 (62,5%). 3 d'entre elles ont un score plus élevé que 47 sur 64 (73%). Le meilleur sujet, un saxophoniste ayant appris les notes suivant une technique particulière de repérage des sons par leur localisation dans son corps obtient un score de 75%. A l'opposé, le sujet le moins performant en identification a un score de 46%, ce qui est plus élévé que dans la variante 1.

En ce qui concerne les confusions, l'amélioration des scores de réussite sur le /o/ montrent qu'il est bien moins systématiquement pris pour un /a/, par contre le /a/ est toujours très souvent confondu avec /e/ et réciproquement. Enfin le /e/ est très souvent pris pour un /i/ avec des différences suivant les sujets. C'est souvent au niveau de l'identification du /a/ et du /e/ que se sont faites les différences entre les sujets les plus performants et les moins performants.

Une confusion du /i/ avec un /o/ qui fait remonter notre courbe de tendance des réponses sur les /o/ pourrait être due à l'effet d'ambigüités d'octaves sur les sifflements. L'autre explication plus simple est que le sujet n'a pas été attentif lors de la présentation de cette voyelle.

Différence entre musiciens et non musiciens

A nouveau dans ce cas il y avait 6 musiciens parmi les sujets (et ce malgré le fait que les sujets ayant passé la variante 2 ne soient pas les mêmes que ceux de la variante 1). Un calcul d'analyse de variance similaire à celui réalisé lors de la variante 1 de l'expérience a montré que cette fois ci les résultats des musiciens n'étaient pas significativement différents de ceux des non musiciens ($F(1,18)=6,71$, NS). L'effet de contexte semble avoir rapproché les non musiciens des musiciens car les résultats de ces derniers ne sont pas moins élevés que pour la variante 1 (l'ensemble des représentations graphiques de ces résultats est en annexe E.2).

Effet d'apprentissage

En raison de l'élimination des confusions favorisées dans la variante 1 par la présentation de voyelles isolées succéssives, il est pertinent de tester l'effet d'apprentissage dans cette variante.

	Voyelles répondues			
Voyelles jouées	/o/	/a/	/e/	/i/
/o/	59.00	22.00	13.00	6.00
/a/	20.00	32.00	36.00	12.00
/e/	6.00	16.00	49.00	29.00
/i/	2.00	5.00	9.00	84.00

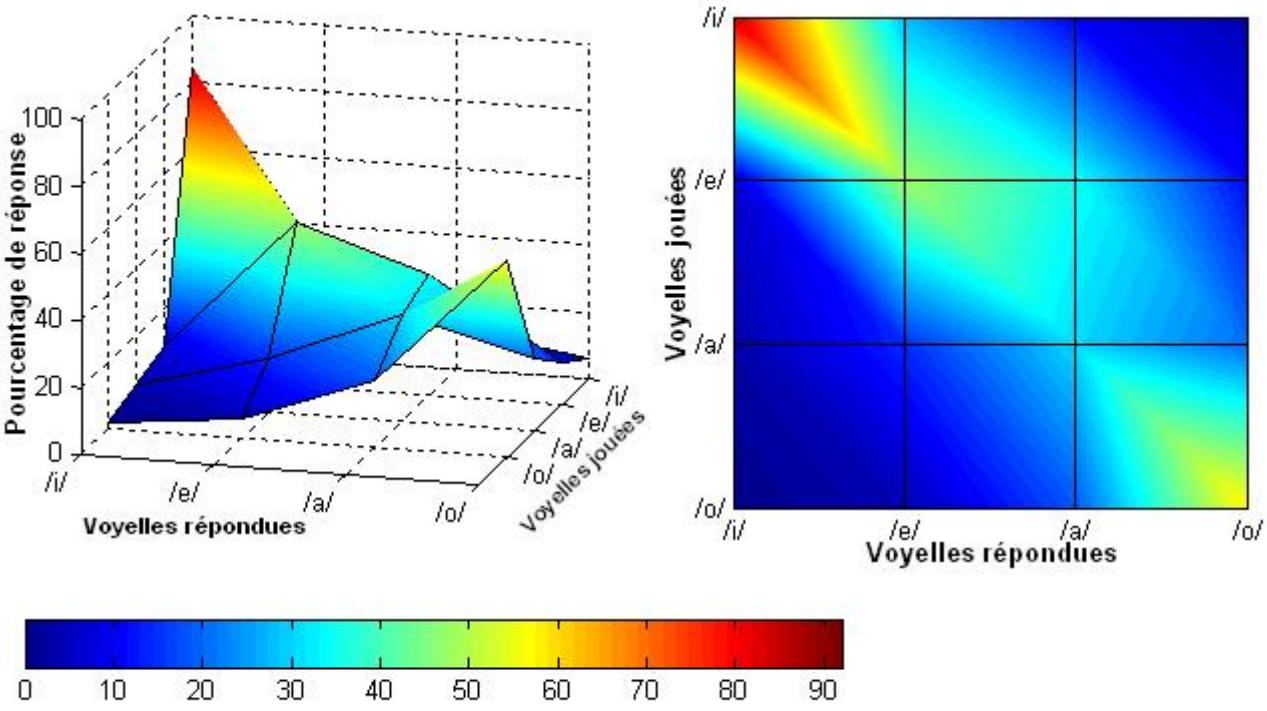

Figure 100 : Performance d'identification des voyelles sifflées avec contexte en phase d'apprentissage

On remarque que dès l'apprentissage les tendances observées dans le test sont affirmées. Etant donné le peu de voyelles de la phase d'apprentissage, ceci suggère que les sujets s'appuient sur des catégorisations déjà présentes dans leurs habitudes linguistiques.

4.3.1.5. Discussion générale

4.3.1.5.1. Première interprétation

Les résultats des deux variantes de cette expérience d'identification montrent que des sujets ignorants tout de la répartition des voyelles sifflées du Silbo espagnol et des langues sifflées en général mais ayant dans leur langue parlée le même type de voyelles que celles de l'espagnol arrivent à catégoriser de manière intuitive des voyelles sifflées du Silbo qu'ils écoutent pour la première fois. La répartition de leurs réponses est similaire à la représentation cognitive des siffleurs. Le fait que cette tendance soit déjà présente à l'apprentissage et que ce dernier ait été relativement court semble indiquer que la représentation perceptuelle des voyelles de manière étagée en fonction de la fréquence pré-existe chez les sujets testés. De plus, ces résultats semblent confirmer que les siffleurs s'appuient sur une réalité perceptive présente dans la version parlée pour transposer les voyelles en hauteurs sifflées.

4.3.1.5.2. Positionnement de nos résultats dans le cadre de la phonétique et de la perception des voyelles

Le mécanisme de la perception des voyelles a été testé par de nombreux expérimentateurs grâce à des tâches d'identification, de discrimination ou de *matching (qui consiste à apparier)* de stimuli. Celles qui obtiennent les résultats qui se rapprochent le plus de nos observations sont les expériences menées à partir d'un *« effective upper formant »* (Bladon et Fant, 1978 p.1) appelé F2'. F2' est la dérivée du F2 à des degrés variables de manière à tenir compte de la contribution des formants plus élevés en fréquence. Ce formant est donc considéré comme l'intégration perceptuelle de l'ensemble des formants élevés (au dessus du formant 1(F1)). Sur cette base, différentes équipes de recherche ont demandé à des sujets d'apparier des voyelles issues du suédois puis d'autres langues à une approximation synthétique à deux formants constituée du F1 et du F2' où le F2' était ajustable par les sujets eux-mêmes (Carlson et al., 1970). La justification d'une telle approche tient dans la remarque de Bladon et Fant : *"The notion that a vowel quality can be satisfactorily approximated by an acoustically-derived representation in two dimensions has held a long standing attraction for phoneticians, in the hope that such acoustic data might provide correlates for the articulatory dimensions of tongue height and tongue retraction which form the basis of conventional phonetic vowel quality diagram. More recently, perceptual studies have supported the view that a two formant model of a vowel is also a valid representation at some level (not necessarily peripheral) of auditory processing."* (Bladon et Fant, 1978, p.1). En effet, d'une part, le travail de Plomp et al (1975) avait montré que l'identification des voyelles du hollandais issues d'une représentation spectrale extraite par une analyse en composantes principales (PCA pour Principal Component Analysis) correspondait de manière très proche aux scores d'identification des mêmes voyelles dont seulement le F1 et le F2 étaient disponibles (plan logF1 versus logF2 utilisé en phonétique). D'autre part – et c'est l'aspect qui nous intéresse le plus- l'expérience

préalable de Carlson et al (1970) revenait à examiner les conséquences perceptuelles de la manipulation de l'espacement entre deux proéminences spectrales dans des voyelles synthétiques. Les résultats montraient que pour des formants F1 et F2 relativement proches (de 3 à 3,5 Bark[102]) les sujets placaient F2' au niveau de F2 réalisant ainsi une intégration sur le pic de densité spectrale le plus intense (voyelles /a/, /o/, /u/ Figure 101). Carlson et al (1970) avaient également remarqué que pour des voyelles où les formants F2 et F3 étaient proches, les sujets placaient le F2' entre ces deux pics (typiquement voyelle /e/ Figure 101). Enfin le /i/ du suédois était un cas à part dans leur expérience puisque les sujets placaient le F2'entre F3 et F4.

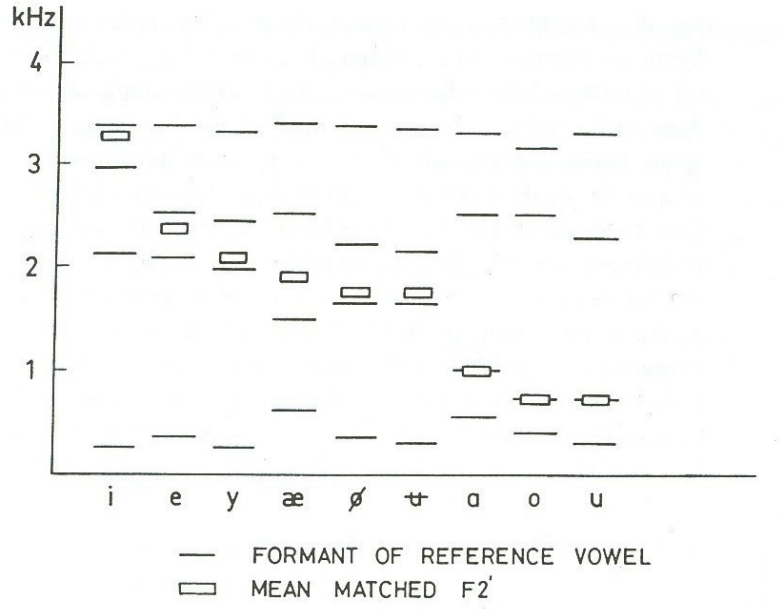

Figure 101 : Résultat s de l'expérience de Carlson et al (1970) (in Stevens, 1998, p. 240)

Plus tard, Chistovitch et son équipe (Chistovitch et Lublinskaya 1979 ; Chistovitch et al 1979 ; Chistovitch, 1985) ont montré que ce phénomène *d'intégration perceptuelle* apparait à partir d'une proximité de 3,5 bark entre deux formants, quels qu'ils soient.

Bladon et Fant (1978) firent la même expérience que Carlson et al (1970) mais à partir d'un corpus de voyelles cardinales[103]. Leurs résultats sont similaires, ils sont présentés Figure 102[104].

[102] 300 à 400 hz environ.

[103] Voyelles de l'A.P.I. (Association Phonétique Internationale)

[104] Dans cette étude les auteurs ont perfectionné une formule déjà établie en 1970, modifiée en 1975 (Carlson, Fant et Granström, 1975) permettant d'évaluer la valeur du F2' par un calcul, ce qui explique la présence d'un "F2' calculated" sur la Figure 102. Nous ne nous intéressons pas ici à cet aspect de leur approche.

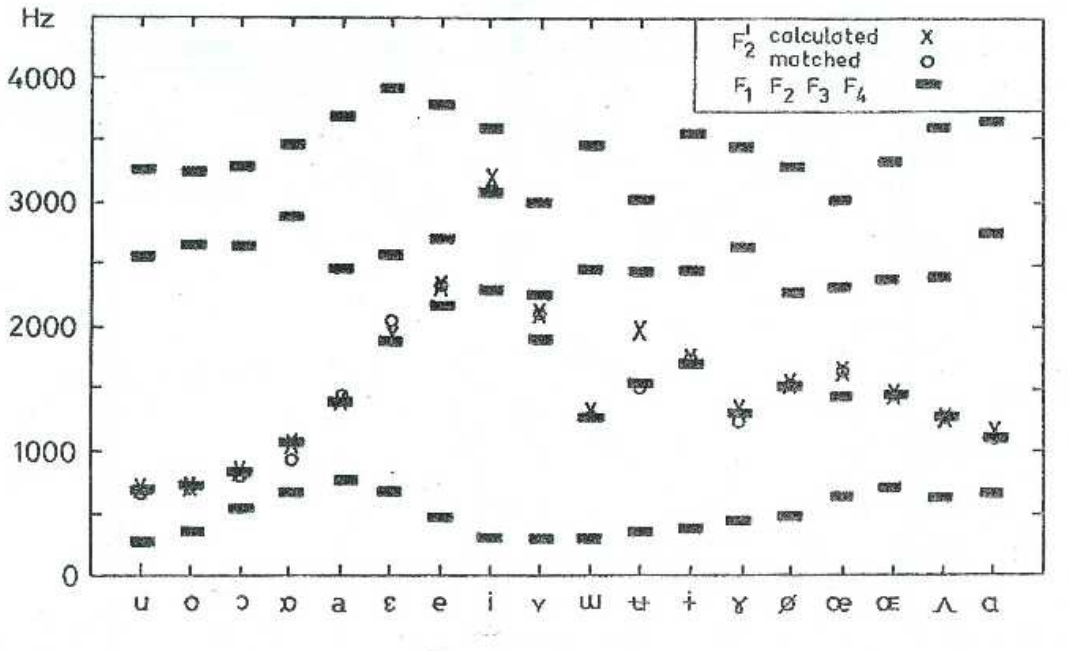

Figure 102 : Présentation séquentielle des voyelles entermes de formants F1, F2, F3, F4 mesurés et de F2' calculés et perceptuellement appariés (*matched*) (Bladon et Fant, 1978 p.9)

Timbre, F2′ et voyelles sifflées

La répartition du *effective formant* F2' nous semble particulièrement adaptée à une comparaison avec la stratégie de transposition des voyelles en langues non tonales sifflées. En effet on retrouve la distinction claire entre /i/ et les autres voyelles signalées dans la partie typologie pour le tuc, le grec et l'espagnol sifflé, mais également le groupement des voyelles postérieures et des voyelles centrales dans deux catégories différentes. De plus, elle fournit une explication plus plausible que celle d'une transposition du F2 seul.

En effet, en soulignant le rôle de l'intégration perceptuelle de deux formants de la parole en un seul F2' ces recherches mettent en évidence l'importance des zones de forte compacité dans la perception du timbre des voyelles et, d'après Stevens, le rôle de l'espacement critique mis en évidence par les expériences que nous avons citées, a des conséquences générales importantes pour la classification des voyelles : *« some aspects of the auditory system undergoes a qualitative change when the spacing between two spectral prominences becomes less than a critical value of 3,5 bark »* (Stevens 1998, p.241). Il a illustré son propos en situant l'emplacement des F2' sur une analyse spectrale de quelques voyelles (Figure 103). On peut y voir que la largeur de la bande de fréquence des deux formants les plus proches joue un rôle important. A l'intérieur de ce regroupement de fréquences, la profondeur et la largeur des vallées spectrales entre les formants sont relativement limitées. Par contre, les vallées spectrales sont plus profondes et larges autour. Ces conditions facilitent l'émergence de propriétés perceptives particulières liées à ces qualités du timbre des voyelles.

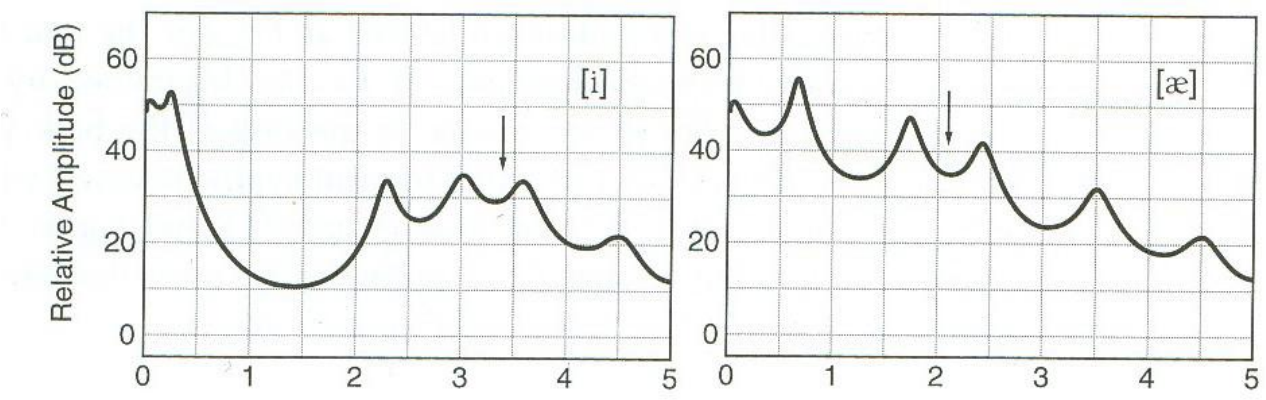

Figure 103 : Localisation du F2'(flèche) sur l'analyse spectrale de 2 types de voyelles (Stevens, 1998. p. 289)

Le /i/ est marqué par une concentration de fréquences élevées, son timbre peut être qualifié de brillant ou « aigu », le /e/ est intermédiaire (comme le [*æ*] de la Figure 103) et le /a/ ou /o/ ont des timbres parfois qualifiés de « grave ». Ces termes empruntés à l'acoustique musicale correspondent bien à la réalité des sifflements. Le lien fait sans grande difficulté par les sujets de notre expérience entre les voyelles sifflées et les quatre voyelles types du français nous fait penser qu'ils ont utilisé ce paramètre du prototype de voyelle qu'ils avaient en mémoire.

Remarques complémentaires :

Il semble que le même type de résultats ne puisse pas être obtenu avec des sujets de langue maternelle tonale, en effet, lorsque nous avons fait passer l'expérience à un étudiant arrivant du Mali et de langue maternelle Bambara, nous avons constaté qu'il n'avait pas du tout la même logique de réponse, malgré le fait qu'il ait fait tout son cursus scolaire en langue française. Il se peut qu'il ait donné la priorité à la Hauteur Fondamentale des voyelles types du français alors que les sujets Français donnaient la priorité à la Hauteur Brute marquée par la compacité du timbre. Ce type de différence de stratégie n'est pas visible dans la version parlée car les deux qualités de Hauteur perçue sont alors présentes dans la voix. Il permettrait de montrer qu'en fonction de sa culture le sujet écoute prioritairement des zones fréquentielles différentes de la voyelle. Son écoute serait donc culturellement orientée. Pour vérifier cette hypothèse il faut concevoir des expériences complémentaires.

4.3.1.6. Conclusion

Les résultats de l'expérience de psycholinguistique que nous avons présentés dans cette section ont été l'occasion de mener une réflexion sur la perception du timbre des voyelles car nous avons trouvé que des sujets français arrivent à avoir, sans exposition préalable, des résultats d'identification de 4 voyelles sifflées du Silbo espagnol montrant une catégorisation des voyelles sifflées similaire à celle des siffleurs eux-mêmes. Les langues sifflées définissent un bon modèle d'analyse pour poursuivre les recherches sur les regroupements perceptifs de formants. A partir de résultats issus de recherches de référence en phonétique nous avons développé une argumentation soutenant que la catégorisation des voyelles sifflées de nos sujets

s'est appuyée sur une référence implicite aux bandes de fréquences harmoniques les plus compactes et intenses des voyelles de référence du français parlé.

Comme les siffleurs transposent la voix parlée espagnole pour siffler ces voyelles et qu'une pratique linguistique est un équilibre entre perception et production, il se peut qu'ils utilisent aussi cet indice de compacité du timbre lors de l'écoute. Actuellement, nous ne pouvons pas conclure catégoriquement en ce sens. Pour pouvoir l'affirmer de manière plus sûre, il faudrait faire des tests psycholinguistiques complémentaires avec des siffleurs de plusieurs pays. Mais si c'était le cas, les langues sifflées comme le grec ou le turc qui se limitent souvent à un octave entre la fréquence du /o/ et celle du /i/ permettraient de préciser certains aspects de la réduction phonétique vocalique. En effet, le F2' varie sur plus de 2 octaves, il y aurait donc une concentration fréquentielle de l'information d'origine.

4.3.2. Analyse de la perception de non-mots et de logatomes par des siffleurs

4.3.2.1. Experience de perception des logatomes turcs

Moles (1970) a étudié la perception de 600 logatomes[105] turcs sifflés par 5 siffleurs. Avant d'analyser les résultats de cette expérience il est nécessaire de rappeler que les habitants de cette zone isolée, comme tous les siffleurs en général, sont très peu familiarisés avec ce type d'exercice de reconnaissance d'éléments sans sémantique. Tous les individus testés étaient illettrés (com. pers. Busnel 2005) et donc pour eux le découpage en lettres indépendantes est une abstraction qui n'a pas de sens. C'est pourquoi, même si la segmentation en syllabes ne leur pose pas de problème technique, le fait même de devoir siffler des éléments n'ayant pas de signification leur paraît incongru et tout à fait inutile.

Performances de reconnaissance des logatomes

Le tableau de résultats disponible dans l'étude publiée par Moles (1970) et que nous reproduisons ci-dessous concerne la perception de logatomes monosyllabes de la forme Consonne-Voyelle. Or ce type de présentation n'est pas celui permettant d'obtenir les meilleurs taux de reconnaissance en raison du fait que la consonne perd certains aspects temporels à l'initiale. L'auteur remarque à ce propos que d'autres types de combinaisons ont été essayées comme les ordres VC et V1CV1 . Il signale que l'ordre VC permet d'obtenir les mêmes scores que l'ordre CV, mais que l'ordre V1CV1 permet d'augmenter les performances de reconnaissance de 21% à 32%

[105] non- mots suivant des règles mises au point pour les besoins de la téléphonie

Tableau 31 : Résultats des scores de performance de reconnaissance de logatomes (Moles 1970, p. 85)

Siffleurs (% moyennes)	**Adil (22%)**	**Bebek**	**Ali (25%)**	**Osman (20%)**	**Sadik (22%)**
Adil (16,6%)	_	16,25	22	12,5	
Bebek	22,5	_			
Ali (20%)	19		_	17,50	24
Osman (22%)	23,75		21,25	_	21
Sadik (31%)			32,5	30	_

D'autre part, les résultats qu'il présente (Tableau 31) permettent de faire émerger les aptitudes des émetteurs (moyennes par lignes) et des récepteurs (moyennes par colonnes) et de conclure qu'il y a des différences notables entre individus. Moles a également observé, lors d'une autre expérience sur des logatomes que chaque récepteur possèdait des combinaisons favorites de syllabes et qu'il reconstruisait ses mots ou ses phrases en s'appuyant sur des éléments qu'il avait véritablement eu le temps de percevoir. A ces derniers, il ajoutait ses éléments préférentiels. Il cite par exemple *« les transformations systématiques de « gel » en « çer » par Saik et la prééminence de « çok » chez Osman »*(Moles 1970, p.98).

Perception des voyelles

Matrice de confusion

Les tests de netteté des logatomes réalisés par Moles étaient particulièrement adaptés à l'étude de la perception des voyelles. Une analyse a été fournie à ce propos. Les résultats ont été présentés sous la forme d'une matrice de confusion que nous reproduisons (Tableau 32). L'auteur a réalisé l'analyse d'un plus grand nombre de matrices non publiées : « *Quelques remarques générales émergent de celles-ci : D'abord la relative stabilité du i chez certains siffleurs, ensuite les transformations sélectives que le e donne en ö, le ay en oy, et le u en ö* »(Moles, 1970, p.100).

Tableau 32 : Matrice de confusion (couple Sadik-Osman) performances de reconnaissance des voyelles et des diphtongues sifflées turques.

« les matrices de confusion sont faites en reportant sur les bords verticaux et horizontaux de la matrice, par exemple, (respectivement) la série des phonèmes émis et la série des phonèmes reçus »(Moles 1970 p.100)

	i	E	A	Ü	I	U	Ö	AY	EY	OY	UY	O
i	9											
E		4					7					1
A			2	2		1	1			1		6
Ü				3	1		6					4
I					3	1		1				
U	3			1	2	4	1					
Ö		1			1	1	4	1				4
AY		1					1	5	1	6		
EY	1	1							3	1		1
OY			1					1		2	2	1
UY	4	5			1		1		1		1	
O							1					4

Comparaison avec notre analyse de la répartition des voyelles

Toutes les confusions citées ci dessus sont cohérentes avec les résultats statistiques que nous avons obtenus dans la la partie Typologie à propos des voyelles turques..

A partir de l'unique matrice à notre disposition qui concerne ce couple émetteur-récepteur, nous pouvons également remarquer un certain nombre d'autres tendances :

-Une forte confusion de « a » et « o ».

-Le uy est fortement confondu avec soit i soit e.

-Une seule confusion n'est pas très cohérente avec nos données il s'agit de la confusion de ü en ö (valeur de 6 items) alors que « ö » n'est pas confondu avec « ü » mais plutôt avec « o », ce qui est cohérent avec nos résultats. Nous avons cherché à comprendre pourquoi en observant les autres confusions :

> Les autres voyelles prises pour un « ö » sont « e » et « ö » lui-même. Ces deux confusions sont également cohérentes avec nos résultats. De son côté « ü » est légèrement confondu avec « a » et « u » et lui-même. Ces dernières confusions bien que moins marquées que celles de « ö » semblent indiquer que la différence entre nos données statistiques et les résultats de Moles viennent de la fréquence de sifflement du « ü ». Comme il existe des différences d'habitude d'élocution entre les siffleurs, il se peut que la fréquence de sifflement du « ü » de Sadik soit plus basse que celles des 3 siffleurs à partir desquels nous avons fait nos calculs.

Conclusion

Les siffleurs turcs obtiennent des taux d'identification de 20 à 50% de logatomes issus de leur langue (performances variant en fonction du type de logatome). Des matrices de confusion de la perception des voyelles ont été extraites de ces résultats. Elles confirment que des sifflements proches en fréquences entraînent des confusions perceptives entre les voyelles correspondantes.

4.3.3. Intelligibilité des mots sifflés

Pour l'analyse de l'intelligibilité des mots sifflés nous nous appuierons de nouveau sur les études de Busnel (1970) et Moles (1970) réalisées en Turquie. Busnel a mené une étude détaillée sur l'intelligibilité des mots et Moles a analysé l'impact de la fréquence des mots en constituant un début de dictionnaire sifflé.

4.3.3.1. Résultats d'une expérience de reconnaissance de mots sifflés

Dans le Tableau 33 nous avons regroupé les données issues de deux tableaux de résultats d'une même expérience d'intelligibilité des mots réalisée par Busnel (1970). Le taux élevé de reconnaissance des mots émis avec la voix parlée donne une bonne référence de la difficulté de la tâche qui puisait les mots dans une liste favorisant parfois les confusions (voir dernière ligne du tableau pour des précisions sur les ambigüités intentionelles du corpus)

Tableau 33 : Résultats de reconnaissance des mots parlés et sifflés issus d'une liste favorisant l'ambigüité (tableau que nous avons constitué à partir des données de différents tableaux de Busnel (1970), p. 46 et 49).

	Voix Parlée		Voix sifflée	
Nom du receveur	Nombre de mots écoutés	Réponse positive	Nombre de mots écoutés	Réponse positive
		Réponse négative		Réponse négative
Ali Cirit (m)	50	44	45	31
		6		14
Osman Cirit (m)	47	45	45	28
		2		17
Sadik Bebek (m)	46	44	46	35
		2		11
Adil Cindik (m)	45	47	45	30
		0		15
Mehmet Kosek (m)	47	47	45	32
		0		13
Récapitulation globale	235	227 (96,6%)	226	156 (69%)
		10 (3,4%)		70 (31%)
Exemple de mots de la liste:	Bana, sana, ormana, osmana, kapici, yapici, bakalim, yakalim, parali, yarali, yüzonbir, yüzonbin			

Les taux de reconnaissance des mots sifflés isolés restent relativement élevés (69%) mais on observe immédiatement une différence avec les 96,6 % de réussite de la voix parlée qui confirme qu'à courte distance et sans contexte lexical cette dernière reste la plus performante.

4.3.3.2. Eclairage par analyse du contexte lexical

4.3.3.2.1. Cadre de l'étude

Miller et al (1951) ont trouvé que les performances d'intelligibilité ne dépendaient pas uniquement des données inhérentes au matériel sonore utilisé mais également à d'autres facteurs comme la connaissance du sujet traité dans la phrase, ou la connaissance a priori du contenu des stimulus dans lequel l'expérimentateur puise. Leurs résultats montrent que les taux de reconnaissance de mots dépendent fortement de la taille de l'échantillon de test. Ainsi, ils ont obtenu des taux 60% plus élévés pour un matériel sonore constitué de 8 mots stimuli par rapport à un autre constitué de 256 mots stimuli. Ils en ont déduit que les auditeurs avaient besoin de moins d'information acoustique dans le cas d'un petit nombre d'alternatives.

Afin d'approfondir l'analyse de l'intéraction entre les performances d'intelligibilité et ce type d'informations contextuelles, plusieurs types d'approches générales sont possibles. Elles visent à quantifier l'effet de l'étendue de l'échantillon testé:

-La première et la plus souvent utilisée repose sur la constitution d'un dictionnaire fréquentiel des mots, des syllabes ou des phonèmes de la langue. Ces paramètres sont importants car ils indiquent d'une part l'homogénéité du vocabulaire et des indices phonémiques disponibles, d'autre part, ils permettent de *« prendre la température du langage. C'est à dire une notion statistique liée à l'effort intellectuel requis pour utiliser des mots de plus en plus spécifiques donc de plus en plus rares.* »(Moles 1970. p. 73)

-La deuxième approche qui peut compléter la première, tient compte de l'enchaînement relatif des mots et des phonèmes dans une phrase. Elle s'appuie sur la théorie de l'information (Shannon 1948). Elle utilise l'estimation de l'entropie linguistique[106] (Shannon 1951) d'un ensemble d'éléments pour quantifier l'effet du contexte.

-Enfin, d'autres approches statistiques se sont développées sur la base de ces deux premières. Ces modèles permettent de quantifier les effets du contexte à partir des relations obtenues entre les performances d'intelligibilité des sujets et les éléments constitutifs (phonèmes ou mots) du signal de parole hors contexte ou en contexte (Boothroyd 1978, Bosman 1989). Ils s'appuient sur des algorythmes statistiques développés pour la reconnaissance des mots (Boothroyd 1968).

[106] L'entropie linguistique, qui est une mesure de la densité d'information dans un flot de parole, décroît quand le nombre d'éléments d'une phrase augmente. De telles estimations sont obtenues en analysant les réponses des sujets confrontés à la reconstitution de phrases dont ils n'ont qu'une présentation partielle (Taylor 1953).

4.3.3.2.2. Application aux langues sifflées

Seule la première des trois approches a été appliquée aux langues sifflées. Nous présentons le résultat de cette analyse du dictionnaire fréquentiel dans les paragraphes suivants.

De rares analyses du dictionnaire des langues sifflées

Notre souci de recréer des ambiances d'usage traditionnel de la langue sifflée fait que notre corpus est un bon révèlateur des centres d'intérêts des siffleurs mais le peu de temps passé dans chaque lieu (deux mois au maximum) explique pourquoi il n'est pas suffisant une étude statistique des fréquences d'usage des mots. C'est pourquoi, nous nous appuierons ci-dessous sur les résultats d'autres études.

Constitution de dictionnaire

La tâche de constitution d'un dictionnaire est facilitée par le fait que les siffleurs utilisent un vocabulaire sifflé correspondant aux mots du quotidien. En outre, la grammaire de la parole sifflée n'est pas différente de celle de la langue locale, il n'y a pas de syntaxe spécifique. Le dictionnaire d'un langage sifflé n'a été évalué qu'en deux lieux: dans la région de Kuskoy en Turquie (Moles 1970) et dans l'île de la Gomera aux Canaries. Les estimations varient suivant les sources mais l'étendue lexicale du sifflement porte au moins sur 2000 mots. Aux Canaries une dernière estimation réalisée par les siffleurs dans le cadre de la revitalisation de la langue recense plus de 4000 mots très courants. Tout autre mot de la langue est toutefois susceptible d'être transposé en sifflements et intégré à ce vocabulaire au bout d'un certain temps d'usage. C'est par exemple le cas du mot turc signifiant « camion » qui n'était presque pas utilisé dans les années 70 lors de l'expédition organisée par Busnel et qui était d'un usage courant lors de notre passage.

Analyse du dictionnaire

Pour les raisons exposées plus haut une seule étude approfondie a été menée jusqu'à aujourd'hui pour évaluer l'impact des effets de contexte dans une langue sifflée. Ces résultats sont à notre avis révélateurs de l'ensemble des situations même si l'auteur de l'étude, Moles, indique qu'elle n'est qu'une approche du problème en raison du peu de temps passé sur place à collecter des données et du corpus de mots sifflés limité au 120 mots sifflés les plus courants.

« La liste ainsi établie qui émergeait d 'un corpus d'environ 1000 mots, comportait environ 120 types qui ont été sifflés sur ordre accompagnés de leur traduction en turc parlé[...] Nous avons enfin étudié la répartition des fréquences, tirées de ce dictionnaire.[...] Du point de vue linguistique, l'un des avantages du dictionnaire fréquentiel est de permettre d'établir, même de façon très approximative, une courbe de Zipf, dont la pente exprime une certaine « température du langage », grandeur strictement statistique et, donc, de la comparer avec d'autres courbes analogues, en particulier, au sujet de la langue parlée correspondante »(ibid., p84).

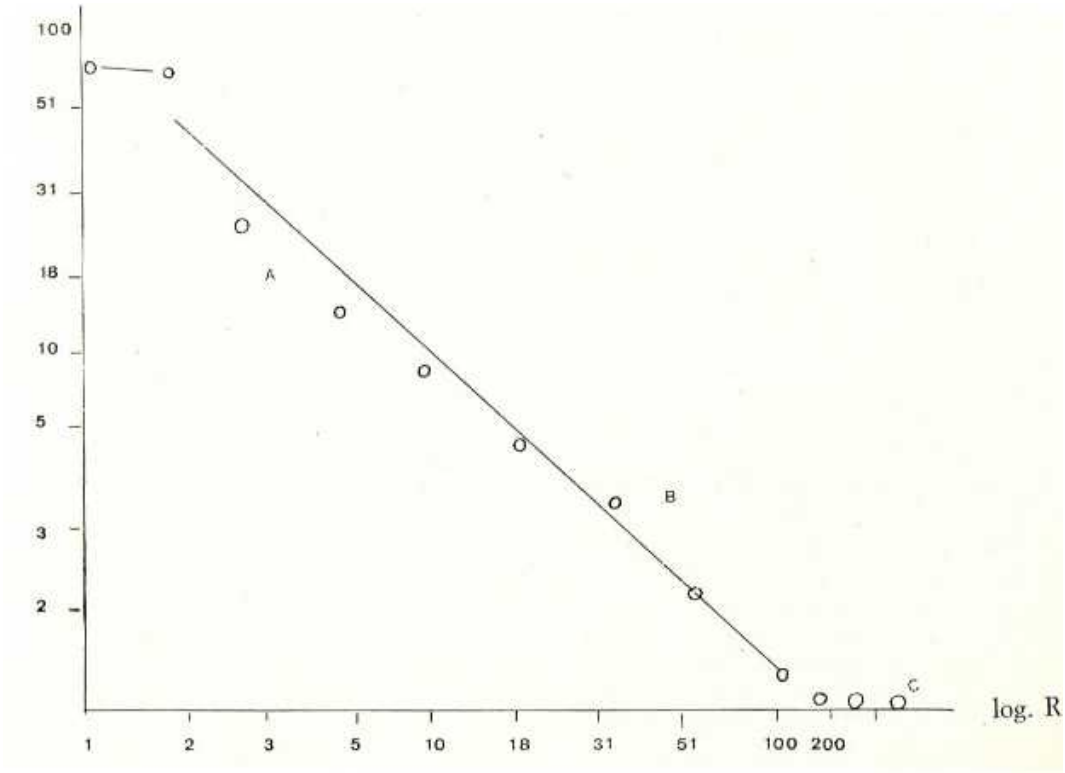

Figure 104 : Courbe de Zipf d'un corpus de 120 mots sifflés turcs avec en abscisse le Log du rang du mot dans le dictionnaire et en ordonnée le log de l'ordre d'apparition (Moles 1970, p. 113)

Moles analyse cette courbe ainsi : *« L'examen de cette courbe révèle en dehors d'erreurs locales au départ qui, avec un corpus aussi restreint ne méritent aucune interprétation, trois zones dans le vocabulaire: la zone A, de pente 1/1,2 environ, n'est pas très éloignée de la pente relative au langage parlé, elle l'épouse en quelque sorte. Dès le vingtième mot déjà, la courbe change de pente révèlant une nouvelle zone B du dictionnaire, celle qui correspondrait à un effort de « pensée sifflée » ; enfin la zone C (queue de la courbe) […] qui correspond à des mots utilisés une à deux fois ».* (Moles 1970, p.113) D'après l'auteur il se peut aussi que ces derniers mots soient liés à des imprécisions de la traduction.

La constitution d'un dictionnaire fréquentiel passe par l'analyse de l'étendue du vocabulaire testé et de ses domaines d'application en prenant en compte le milieu d'évolution des sujets dont on estime les performances d'intelligibilité. Les mots les plus communéments sifflés sont liés aux activités quotidiennes des siffleurs. Ce sont eux qui correspondent à la zone B de la courbe de Zipf de la Figure 104. Cet aspect est important dans la forme sifflée car le vocabulaire d'usage le plus courant est celui utilisé quand une personne est isolée, ce qui arrive surtout en dehors du village. Dans bien d'autres contextes modernes, cette tendance de la spécification du vocabulaire en fonction du métier est tangible dans la forme parlée: c'est par exemple le cas dans les métiers de la recherche.

4.3.3.2.3. Conclusion

Le taux d'intelligibilié des mots isolés atteint environ 70% d'après des études menées en turquie dans les années 60-70. Ces pourcentages sont relativement stables d'un auditeur-siffleur à l'autre. Ils représentent une augmentation de plus de 20 à 30% par rapport aux logatomes (du type Voyelle-Consonne-Voyelle). Cette différence tient en priorité à la reconnaissance de la sémantique qui repose sur l'ensemble des connaissances des sujets testés. Le vocabulaire courant en sifflement a été évalué entre 2000 et 4000 mots à la Gomera et en Turquie. Il dispose de plusieurs couches dont certaines sont plus spécifiques à l'usage sifflé.

Dans les mêmes conditions d'écoute, les mots en voix parlée sont identifiés à 96%. Mais comme la langue sifflée est adaptée, avant tout, à des conditions de communication dans le bruit de fond de la nature, il est interessant de comparer ces mesures d'intelligibilité à celles de la voix dans le bruit. En effet, la langue sifflée serait justifiée, pour l'intelligibilité des mots parlés, qu'à partir du moment où le bruit ambiant entraîne une identification inférieure à 70 %. D'après les données de Bronkhorst et al (1993) une telle valeur est atteinte pour la voix dans des conditions d'écoute avec un rapport S/B de 3 dB alors qu'un taux d'intelligibilité des mots de 90% est obtenu pour une écoute avec un rapport S/B de 10 dB. [107]

[107] D'après nos mesures réalisées sur le terrain et pour des phrases (§ 2.3.6.2), la voix parlée inférieure à 80 dB porte moins loin que 40-50 m dans un bruit de fond d'environ 40 dB. C'est donc avant cette limite que le sifflement devient plus efficace que la voix parlée.

4.3.4. Intelligibilité des phrases sifflées (courte et longue distance)

4.3.4.1. Performances usuelles

En considérant l'entité de la phrase nous atteignons la réalité de la structuration des dialogues pratiqués par les siffleurs. Tous les éléments favorisant l'intelligibilité s'intègrent à ce niveau de reconnaissance de la parole.

Sensation sonore	Perception phonétique	Perception des mots	Perception des phrases
Audibilité	Netteté	Intelligibilité	Compréhension

En grèce par exemple nous avons mesuré un taux d'intelligibilité de 95% à 150 m sur 10 phrases échangées spontanément dans une conversation vive et sur le sujet peu commun de la visite d'étrangers. Les phrases enregistrées à cette occasion sont caractéristiques d'une conversation courante. Elles sont assez courtes mais contiennent jusqu'à neuf mots. L'analyse de leur structure n'a montré aucune différence avec la syntaxe habituelle du langage parlé. Les 5% d'intelligibilité manquant de l'interlocutrice (Mrs. Kula) peuvent être attribués à la prise de contact à distance. Mais la même conversation a été perçue avec un taux d'intelligibilité de 100% par la fille d'un des siffleurs effectuant, depuis une distance de 10 m, en direct, la traduction du dialogue en grec parlé.

Comme nous l'avons vu, les communications ne sont pas stéreotypées car le vocabulaire courant est aussi étendu que la voix parlée utilisée dans le même type de contexte (plus de 2000 mots courants). Les arrangements syntaxiques des phrases qui suivent les règles de la langue locale réduisent la probabilité de confusion, marquent la mélodie, tout en donnant une consistence interne aux propos.

Donc, grâce au phénomène de la phrase, l'intelligibilité la plus complète est atteinte, comparable à celle de la voix parlée -quand la pratique de la langue sifflée est encore quasi-quotidienne- et ce, d'après les pratiques observées, jusqu'à des distances pouvant aller jusqu'à un à deux kilomètres suivant le milieu écologique. Ce dernier point est une spécificité sur laquelle nous allons nous attarder car ses conséquences pour la phonétique n'ont jamais été mesurées.

4.3.4.2. Non-linéarité de l'intelligibilité à distance

Une expérience pilote, réalisée en pays Mazatèque en 2003 et que nous décrivons en détails en Annexe F.1 nous a permis d'observer que le phénomène d'intelligibilité de la parole sifflée a une évolution non linéaire avec la distance (comme pour la voix). Cela signifie qu'à deux distances d'écoute différentes, la même émission permet toujours au siffleur de comprendre le message avec les mêmes performances à condition de ne pas dépasser une zone limite tampon où la perte d'intelligibilité est rapide. Les éléments du signal sonore parvenant à l'auditeur sont *suffisants* pour atteindre un taux élevé d'intelligibilité des phrases, malgré la dégradation qu'ils ont subit depuis leur émission à cause des effets de la propagation et du bruit de fond.

Busnel et Classe (1976) avaient également, de manière succinte, observé le phénomène non linéaire de l'intelligibilité des langues sifflées puisqu'ils écrivaient à l'époque *"We recorded a whistled signal at the source and simultaneously, from a point 2000 m away but in visual contact, in a flat country and on a windless day. Analysis by the sonagraph revealed an interesting fact: a significant portion of the frequency band had disappeared. Nevertheless the signal remains intelligible, for what remains of the informational skeleton amply suffices for a mental reconstruction of the message."* (Busnel et Classe,1976).

Dès lors les langues sifflées présentent un cas extrême de la capacité du cerveau humain à reconstruire un message linguistique à partir d'indices acoustiques parcellaires. Du point de vue de l'analyse de l'intelligibilité du langage, le cadre expérimental offert par ce phénomène naturel est extrêmement prometteur. En effet, à des distances même moyennes (permettant un taux d'intelligibilité élevé), on peut s'attendre à ce que le signal parvenant à l'auditeur soit particulièrement réduit puisque qu'il est le résultat de la propagation d'un style de parole qui opère lui-même une concentration et une sélection physique de l'information linguistique d'une phrase dans la bande de fréquence du sifflement.

4.3.4.3. Expérimentation en conditions de distance

4.3.4.3.1. Introduction

Afin d'étudier la répartition des indices acoustiques qu'un individu perçoit à grande distance et ses conséquences phonétiques, nous avons mis au point une expérience en plein air dont le protocole nous a permis d'enregistrer exactement les mêmes phrases sifflées à plusieurs distances pour plusieurs langues dont nous avions étudié la phonologie et la phonétique de manière approfondie. Il est important de connaître leur forme dégradée qui parvient à des auditeurs placés à différentes distances car nous pourrions alors comprendre non seulement quels sont les indices réduits qui suffisent à l'auditeur pour décoder le signal de parole, mais aussi nous observerions l'évolution de la dégradation du signal émis en fonction de la distance.

Dans la partie suivante nous donc développons la présentation de cette expérience qui porte sur plusieurs phrases de notre corpus de turc, de grec et de silbo. Puis nous faisons une première analyse des enregistrements obtenus.

4.3.4.3.2. Préparation

Choix du lieu

Nous avons choisi de réaliser notre expérience en milieu montagneux avec une vallée dégagée et un environnement semi boisé, comme c'est le cas dans des milieux permettant des communications jusqu'à 2 km. Notre but était en effet de faire des mesures sur des distances intermédiaires (100, 150 et 300 mètres) et une distance longue (550 m) dans un milieu qui n'était pas trop réverbérant. Après différentes investigations dans les Alpes et une expérience pilote dans le Vercors, nous avons opté pour la vallée présentée sur la Photo 27 qui forme un guide d'onde de qualité intermédiaire.

Photo 27 : Vallée cadre de l'expérimentation à distance

Equipe expérimentatrice

Une équipe de travail de deux personnes a été constituée[108]. L'un des expérimentateurs était chargé de la diffusion des sons et l'autre des enregistrements. Les deux ayant évalué auparavant les distances avec un mètre de maçon de 50m.

Corpus et stimuli

Choix des langues

Trois langues ont été sélectionnées. Il s'agit des langues non tonales dont nous avons analysé la version sifflée le plus en détails : le turc, le grec et l'espagnol. Le turc a été diffusé en version sifflée et parlée (limite voix criée). Une partie du corpus de turc a l'avantage de présenter la même phrase énoncée trois fois de suite par 2 personnes, à la fois en version parlée et sifflée, ce qui donne plusieurs points de vue sur la même réalité linguistique et permet donc une meilleure interprétation des résultats.

[108] Je tiens ici à remercier Rémi Saudax pour sa collaboration.

Choix des phrases

Parmi tout notre corpus, les phrases énoncées par les meilleurs siffleurs dans les conditions d'enregistrement les plus optimales et réalisées à quelques mètres de la bouche ont été sélectionnés. En tout nous avons diffusé et enregistré à chaque distance : 10 phrases de turc, 5 phrases de grec, 4 phrases d'espagnol.[109]

Préparation des stimuli

La version parlée du turc a été étalonnée en amplitude sur la version sifflée afin de pouvoir réaliser une comparaison liée aux bandes de fréquences. Cela entraîne que, compte tenu des niveaux d'émission sonore, une telle aisance dans l'énonciation ne peut être obtenue naturellement (97 dB en moyenne avec des maxima à 104 dB), mais l'analyse de la voix criée qui a été réalisée par la même occasion nous a permis de le remarquer.

Matériel de diffusion et d'enregistrement

Le matériel utilisé est présenté en Annexe F.

4.3.4.3.3. Déroulement de l'expérience

Le protocole expérimental concu afin de gérer le déroulement de l'expérience consistait tout d'abord en la mesure des distances. Puis au réglage des niveaux d'émission pour chaque langue. Puis la diffusion et l'enregistrement ont été réalisés en faisant attention a ce que le bruit du vent soit minimal et qu'il n'y ait aucun bruit industriel (tracteur, avion) . La Figure 105 résume les configurations d'enregistrement (150, 300 m, 550 m). On peut voir qu'une variante a été introduite à 100m : à cette distance l'enregistrement a eu lieu dans deux positions du haut-parleur (face et 45°) afin de mesurer d'éventuels effets sur l'amplitude de l'onde directe.

[109] Nous avons également diffusé 4 phrases de Français énoncées par un Gomero. Mais ces données ne sont pas traitées ici

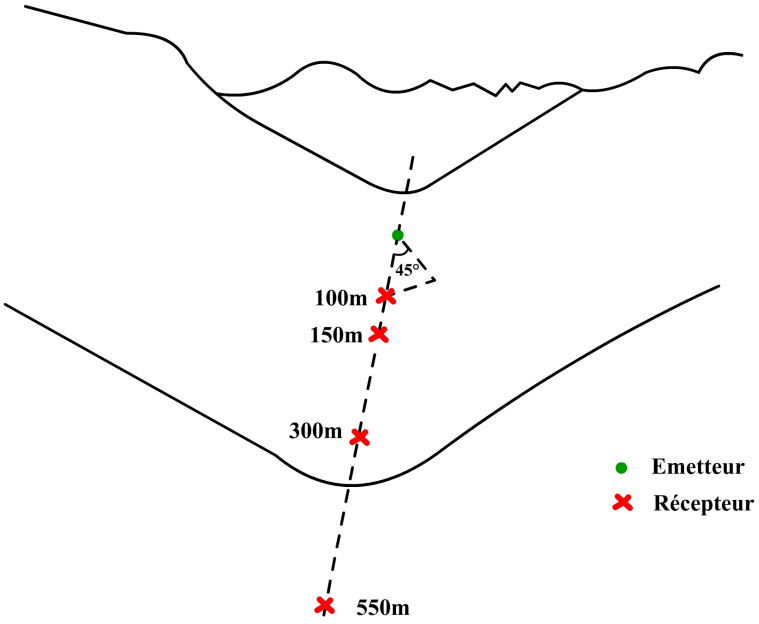

Figure 105 : Configurations de l'expérience

Contrôle du niveau d'émission

Le son était contrôlé uniquement à partir de l'amplificateur. Un niveau de réglage a été opté pour chaque langue en fonction des conditions d'enregistrement et de la technique utilisée par les siffleurs de manière à être le plus proche possible des conditions réèlles d'émission.

Tableau 34 : Niveaux d'émission des phrases pour chaque langue

	grec	espagnol	turc
Niveau moyen (dB)	89	86	97
Maximum de niveau (dB)	101	97	104

Contrôle de l'intelligibilité

La limite d'intelligibilitité n'a été atteinte que pour la langue espagnole qui est celle qui avait été émise aux niveaux de puissance les plus bas (par le siffleur et donc par nous). Pour les autres langues nous sommes restés bien en deçà de cette limite car nous n'avons pas dépassé la distance à laquelle le sifflement original était destiné. L'intelligibilité a été contrôlée dans le casque afin de correspondre à l'enregistrement réalisé. Nous avons estimé que nous étions qualifiés pour juger de l'intelligibilité des phrases car nous avons appris la technique de sifflement à la Gomera et que nous parlons l'espagnol couramment.

4.3.4.3.4. Résultats et première analyse

Dégradation progressive 150, 300, 550m

Exemple du turc

a) Résultat des enregistrements

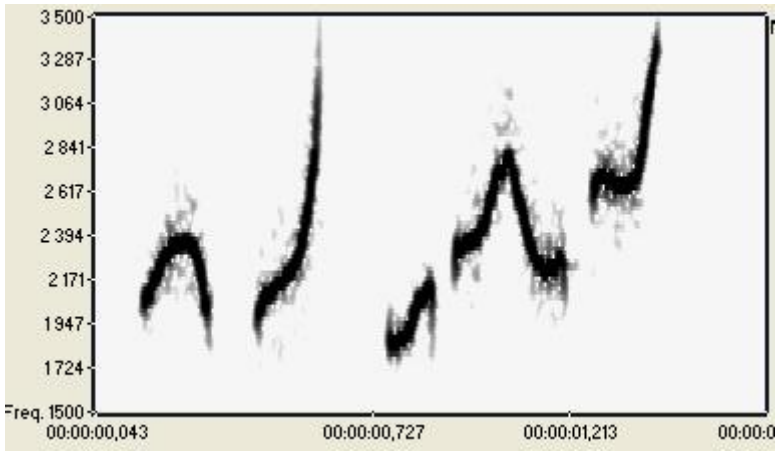

Figure 106 : enregistrement à la source de la phrase turque « mehmet okulagit »

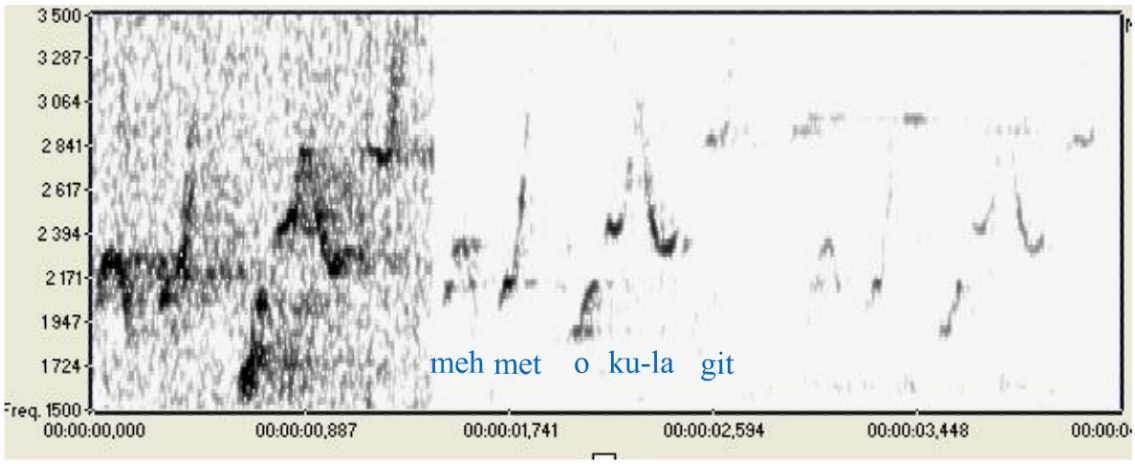

Figure 107 : Même phrase enregistrée à 150, 300 et 550 m, même règlage du spectrogramme aux 3 distances

b) Commentaires

On observe une dégradation progressive de l'émergence du signal par rapport au bruit. L'intelligibilité est toujours complète à 550 m même si elle requiert de l'attention. En ce qui concerne le bruit de fond, il était différent en nature aux 3 distances mais toujours entre 40 et 50 dB. A 150m, le bruit était le plus élevé de toutes les distances d'enregistrement en grande partie à cause de la forte réverbération de cette partie du champ qui renvoyait le bruit des feuilles des sommets des arbres et celui de la présence des grillons.

c) Dégradation de la voix turque

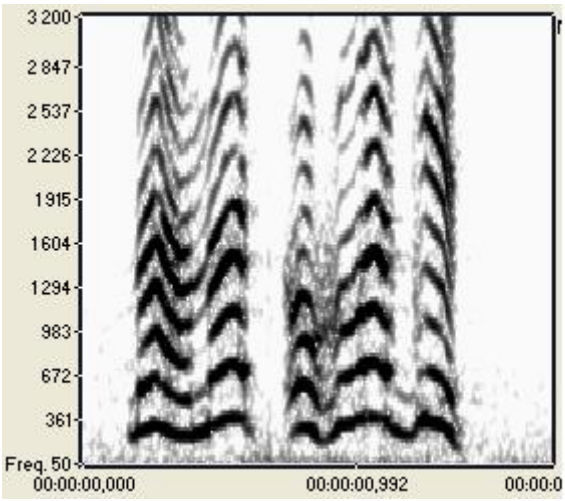

Figure 108 : « Mehmet okulagit » version parlée (limite criée) à la source.

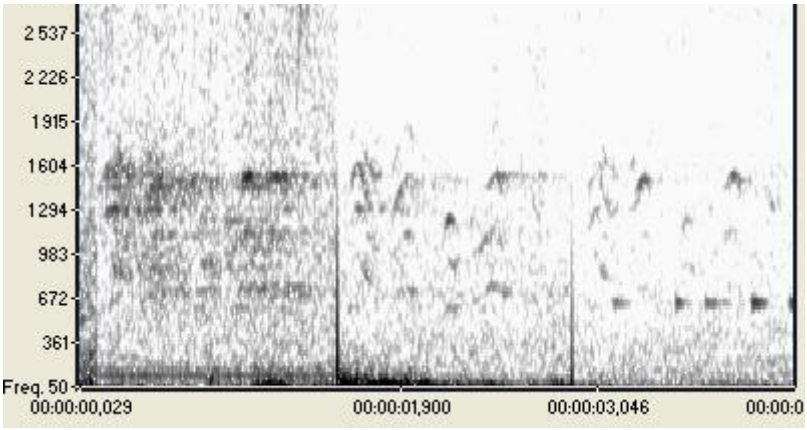

Figure 109 : « Mehmet okulagit » version parlée enregistrée à 150, 300, 550 m (à 550 m les fréquences visibles vers 672 Hz sont issues du tintement d'une cloche de vache)

A 550 m la voix diffusée était à la limite de l'intelligibilité. Nous avons déjà précisé qu'aucun être humain ne peut garder l'aisance d élocution au niveau auquel nous avons diffusé cette voix. Il ne s'agit donc pas de la reproduction de conditions réèlles de communication. Par contre nous pouvons remarquer que les différents milieux d'enregistrement ont eu un impact sur le signal de la voix en le déformant ou en le dégradant dans de grandes proportions. Malgrè cela l'intelligibilité est toujours complète à 150 m (forte réverbération) et 300 m. Les éléments portés par la plus forte amplitude subsistent. Nous n'avons pas encore produit d'analyse phonétique détaillée comparée mais l'on remarque qu'à grande distance ce sont les fréquences situées entre 1000 et 1900 Hz qui sont le mieux transmises à travers le filtre du milieu ambiant.

Cas de l'espagnol

a) *Résultats d'enregistrement*

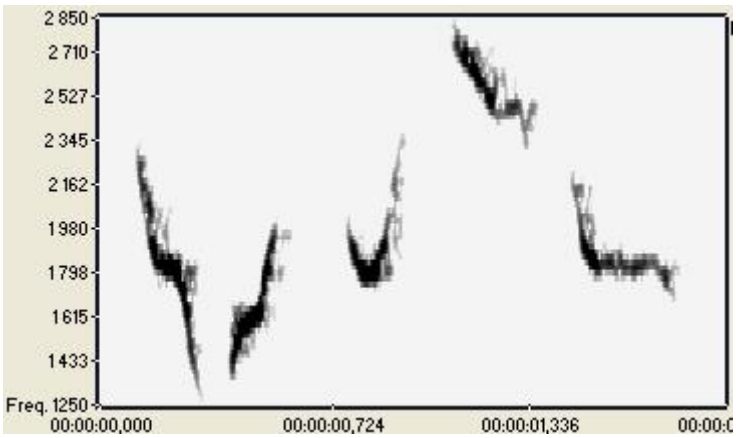

Figure 110 : Enregistrement à la source de l'extrait de phrase espagnole : « la mon-ta-ñe-ta »

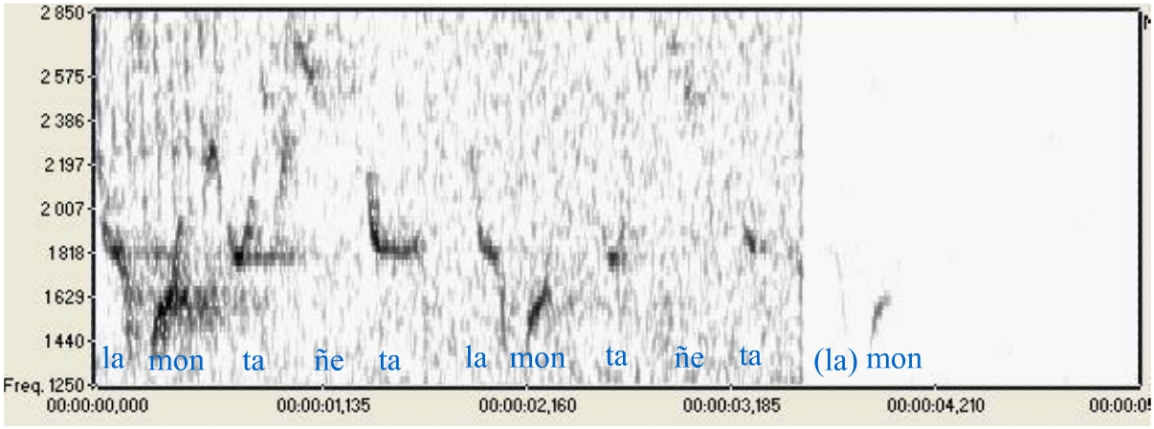

Figure 111 : Même extrait de phrase « la montañeta » enregistré à 150m, 300 m et 550 m. Règlages identiques du spectrogramme aux 3 distances.

b) *Commentaires*

On observe la même dégradation progressive que pour le turc, mais comme le signal est émis à un niveau moins élevé, il émerge difficilement à 550 m.

Le signal est encore clairement intelligible à 300 m mais il n'est plus intelligible à 550 m. Il est pourtant net pour certaines syllabes. Mais la dégradation est trop grande pour reconstituer le signal. Nous avons par exemple confondu « silbo » avec «amigo ».

Remarques sur le bruit

Dans tous les cas nous avons observé une réduction progressive de la réverbération du signal dans le milieu de l'expérience. Le signal à 300 m est plus agréable à l'écoute qu'à 150 m en raison de la forte réverbération à cette dernière distance. De plus, parfois, des modifications locales apparaissent à 300 m mais ne sont pas présentes à 550 m. Ceci est dû à des conditions de vent ou de réverbération différentes.

Mais le point le plus intéressant de cette étude est la possibilité qu'elle nous donne d'encadrer la limite de netteté des syllabes. En effet des cas tels que celui présenté dans la configuration 550m de la Figure 111 (syllabes « la » et « mon ») nous ont permis de conclure que le signal, pour être détecté et reconnu devait émerger d'environ 20 dB de la bande des sifflements. Si dans un mot, plusieurs syllabes ont des parties émergentes à 20 dB et quelsques unes à 6 dB, il est éventuellement possible d'en reconstruire le sens mentalement, tout dépend de leur répartition. A ce niveau de difficulté, la connaissance du sujet joue beaucoup. L'exemple de la Figure 111 est typiquement représentatif de ce type de situation puisque le signal émerge de 6 dB de la bande des sifflements pour « la », et de 20 dB pour « mon » (Figure 112).

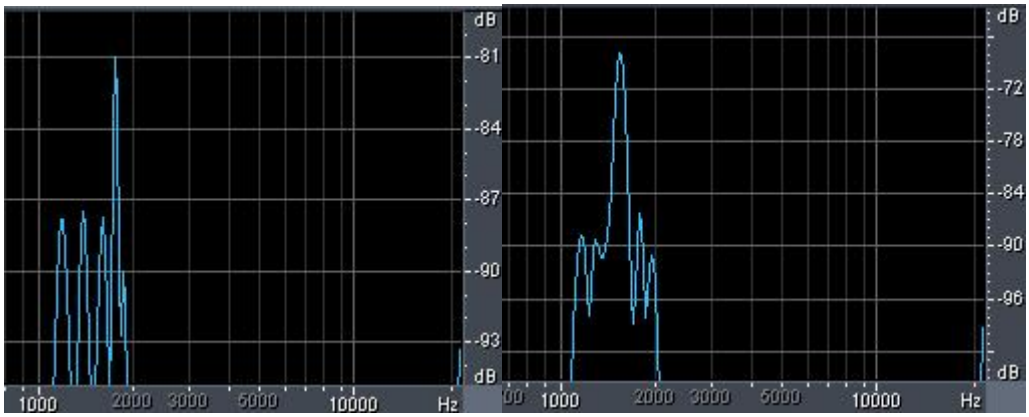

Figure 112 : Emergence de syllabes sifflées du bruit de fond dans la bande 1200-2000 Hz, « la »(S/B=6dB) et « mon » (S/B=20dB) (voir Figure 111)

Une émergence de 6 dB permet de détecter le signal mais pas de s'en servir de base pour une projection mentale permettant l'intelligibilité. Il ne peut que confirmer une impression donnée par un signal émergant à plus de 20 dB.

Remarques:

- les valeurs que nous donnons en décibels ici sont pertinentes pour une écoute monaurale car l'analyse spectrale ne nous permet pas de faire une évaluation de l'émergence avec écoute binaurale. En général celle ci améliore la perception de + 6 dB. Comme nos enregistrements ont été réalisés en stéréo nous en avons bénéficié dans l'écoute au casque (casque englobant les oreilles).
- D'autres considérations de l'écoute en contexte semblent jouer pour la préhension du cerveau sur les sons environnants, comme l'évaluation de la distance de la source (voir la remarque finale de l'expérience en pays Mazatèque en Annexe F.1)

Effet de l'angle à 45°

Résultat type

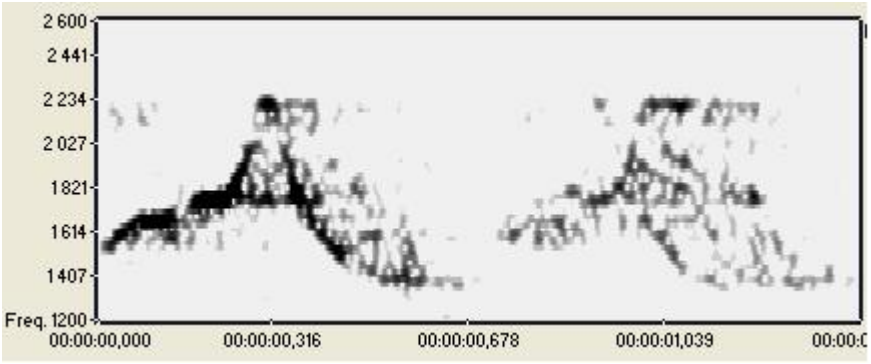

Figure 113 : Mot grec enregistré à 100m dans deux positions différentes de la source (face et 45° par rapport au récepteur).

Commentaires

La zone d 'expérimentation à 100 m était déjà une zone de forte réverbération, avec le changmeent d'orientation de 45° cet effet est beaucoup amplifié (Figure 113). Malgré la forte déformation du signal par réverbération et la moindre intensité de l'onde directe dans le deuxième enregistrement, les phrases sont toujours clairement audibles.

Effet du chant des oiseaux

Lors de l'écoute nous avons également été perturbés par les sifflements d'oiseaux situés dans le même domaine de fréquence que le signal de parole sifflée. Surtout lorsqu'ils venaient de la même direction. Malgré cela, l'intelligibilité restait possible. Mais la dynamique très différente du signal des oiseaux de cette vallée par rapport à celle des sifflements humains explique que l'intelligibilité n'est que partiellement affectée. Les merles de la Gomera, dont le sifflement est très proche de celui des humains, engendrent plus de confusions chez les siffleurs en créant les conditions typiques de l'*effet cocktail party*.

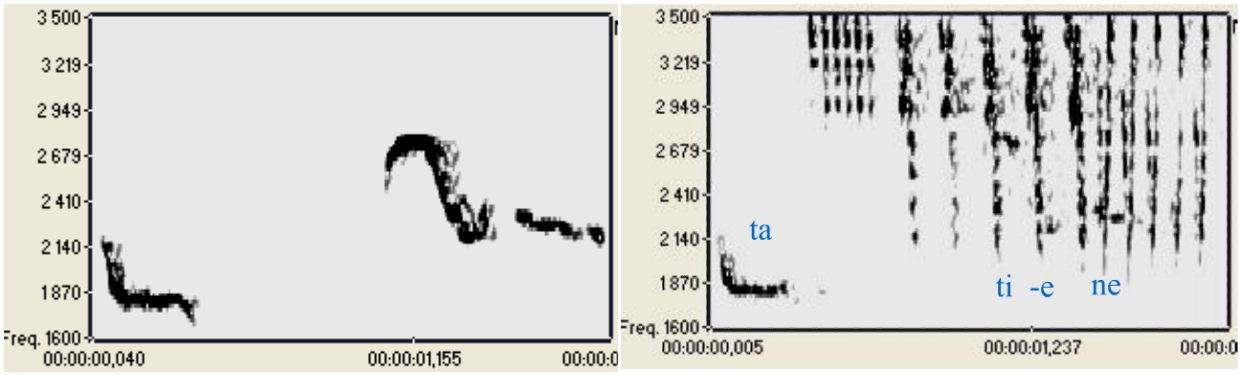

Figure 114 : extrait espagnol sifflé d'origine et à 150 m, « -ta tiene ». Le mot « tiene » est visible à travers la grille du chant d'oiseau

4.3.4.3.5. Première approche phonétique

L'analyse détaillée des structures phonétiques qui subsistent à moyenne et grande distance permet de constater que les signaux entraînant une bonne intelligibilité des phrases ont une répartition qui dépend des points de plus forte énergie du signal d'origine.

-Nous avons remarqué que les voyelles apparaissent le plus clairement : elles portent la plus grande partie de l'énergie sonore de la phrase. Elles sont toujours modulées, ce qui semble indiquer la direction de variation de la (des) consonne(s) qui l'entoure(nt).

-Les éléments de consonnes qui subsistent ne sont pas uniquement ceux proches du noyau de la syllabe, ils concernent aussi parfois des points intermédiaires de la modulation de fréquence, ce qui permet d'en saisir clairement la direction et les aspects temporels liés à l'articulation. Le /t/ est par exemple souvent marqué par un silence intermédiaire qui indique une baisse puis une remontée en amplitude.

4.3.4.3.6. Discussion

Il est important de considérer les problèmes d'émergence du bruit de fond que nous avons souligné grâce à l'espagnol. Dans tous les cas que nous avons observé, la parole sifflée est encore intelligible lorsque les points d'intensité les plus élevés de la syllabe émergent de 20 dB du bruit de fond de la bande des sifflements. Dans le cas particulier de la Figure 111, nous avons trouvé que la netteté des syllabes disparaît entre 6 dB et 20 dB d'émergence S/B de la même bande de fréquence des sifflements. Or, au § 2.3.7.1 nous avons montré que la dynamique intra-sifflements de la parole sifflée est en général de 20 dB (en mesure sur analyse spectrale). Cela veut dire que les éléments les moins intenses de la syllabe « mon » (Figure 111, 550 ont une émergence (S/Bbande sifflé) de 0dB. Compte tenu de nos mesures il y aurait bien une dégradation des éléments perçus par l'auditeur. Pour la même Figure à 300 m on observe de nombreux cas de syllabes pour lesquelles des parties importantes du signal d'origine sont en dessous de la limite des 6 dB de rapport (S/B (Bande des sifflements)). Il y a donc souvent reconstruction cognitive de la parole pour les éléments perçus.

Facteurs généraux d'intelligibilité de la phrase

Parole dans le bruit

De nombreuses recherches sur la parole dans le bruit donnent une première justification des taux élévés d'intelligibilité obtenus par les siffleurs grâce aux phrases même après dégradation. En effet, la reconnaissance de la parole dans le bruit ou en milieux contraignants ne nécessite pas la perception de tous les phonèmes, de toutes les syllabes et de tous les mots car elle s'appuie en grande partie sur l'ensemble des connaissances apprises par le sujet au cours de sa vie (Warren et Warren 1970, Warren 1970[110]). Les éléments manquants peuvent être facilement comblés grâce aux contraintes phonologiques et lexicales, et aux relations syntaxiques et sémantiques qui gouvernent leur enchassement linguistique. Bosman (1989) a

[110] voir en Annexe F.2 pour une explication détaillée de l'expérience des Warren.

montré que les effets de ces indices sur l'intelligibilité de la parole se révèlent être extrêmement élevés lorsqu'elle est présentée dans des conditions d'écoute éloignées des conditions idéales. Il a par exemple mesuré les taux de reconnaisance de la parole de plusieurs types de corpus dans des conditions de bruits contraignants. Il a obtenu des niveaux d'intelligibilité jusqu'à 55 % plus élevés pour des mots insérés dans des phrases par rapport à des mots isolés (ce qui indique un effet combiné de la syntaxe et de la sémantique), et des taux jusqu'à 80% plus elevé par rapport à des non-mots (effet additionel des contraintes lexicales). Il a également constaté que la reconnaissance des syllabes était favorisée par le fait qu'elles ne soient pas prononcées de manière isolée mais intégrée dans un continuum de parole. Ce dernier aspect souligne que la parole contient des indices de coarticulation, d'allophonie et de durée qui fournissent des informations sur la position et l'identité des syllabes et, par extension, des phonèmes.

Rôle clef des modulations

Nous avons remarqué que même à grande distance, les modulations sifflées, qui sont si présentes et rapidement apparentes en sifflements proches, semblent jouer un rôle primordial à la fois au niveau du noyau de la syllabe et au niveau de la constriction ou de l'explosion des consonnes.

a) Modulations et noyau de la syllabe

Un aspect intéressant pour l'étude des langues sifflées tient au fait qu'il a été montré qu'une modulation du type vibrato appliquée à l'ensemble des harmoniques d'un son de la voix permet de créer une plus grande proéminence (McAdams 1984) et cohérence (Darwin et Sandell 1995, Darwin et al 1994) qu'un son non modulé. Marin et McAdams (1991) ont précisés les conditions importantes dans lesquelles une modulation favorise la proéminence d'une voyelle face à d'autres voyelles non modulées simultanées:*" We conclude that coherent, sub-audio frequency modulation on a harmonic sound source contributes to its segregation from other concurrent sounds if the modulation width is large enough*".(Marin et McAdams, p. 350). Des études perceptives récentes de Zeng et al (2005) mettent également en évidence le rôle joué par les modulations cohérentes de fréquence pour la ségrégation et l'association des éléments de parole dans des environnements de parole complexes comme un cocktail party ou un milieu bruyant. Cet aspect du problème nous donne un élément d'explication supplémentaire sur l'efficacité de la parole sifflée dans le bruit. En effet, les voyelles sifflées sont toujours affectées par une modulation, au moins à l'initiale ou à la finale du noyau de la syllabe et nous avons montré que leur bande de fréquence est assez large pour activer plusieurs filtres auditifs simultanément.

De plus, Darwin (1981) a analysé l'effet de la cohérence de modulation de l'ensemble des harmoniques d'une consonne d'attaque d'une syllabe. Il a souligné l'importance des modulations à la fois de fréquence et d'amplitude qui donnent une plus grande valeur perceptive à l'ensemble CV qu'à un phonème seul. Dans les langues sifflées non tonales du groupe 1 de la Typologie que nous avons établie, la modulation reste visible à l'attaque de la consonne, même à distance et reflète ainsi l'importance de ces cohérences en appuyant l'effet perceptif transposé.

b) Double rôle des combinaisons de modulation AM (modulation d'amplitude) et FM (modulation de fréquence)

Plusieurs facteurs ont jusqu'ici souligné l'importance des modulations dans l'organisation générale de la phrase que ce soit au niveau de la continuité temporelle observée grâce aux modulations d'amplitude (AM) en prosodie ou au niveau de la continuité fréquentielle grâce aux modulations de fréquence (FM).

Les études de Zeng et al (2005) ont à nouveaux été intéressantes pour faire avancer notre compréhension à ce niveau. En effet, ces auteurs soulignent que l'avantage de la FM dépend du type de AM qu'ils ont imposé artificiellement et simultanément. Pour eux c'est le signe que *la phase* joue un rôle non négligable dans l'intelligibilité de la parole car c'est le type de combinaison de AM et de FM qui va être important pour l'auditeur, en particulier dans le bruit : « *phase information may not be needed in simple listening task but is critically needed in challenging tasks, such as speech recognition with a competing voice* » (Zeng et al 2005, p.2298).

Or, la phase a recu relativement peu d'attention de la part des études sur la parole. Pour comprendre et analyser les modulations des langues sifflées il convient donc de comprendre le rôle de la phase dans la perception du langage. On sait aujourd'hui que la modification de la phase ne change pas la qualité phonétique des éléments considérés à court terme (Carlson et al, 1979). Mais une des autres rares études sur le rôle de la phase dans la parole a été réalisée par Oppenheim et al (1979), elle a montré qu'un signal de parole avec un spectre d'amplitude maintenu à l'unité alors que le spectre de phase reste intact produit un type de parole appelé *"phase-only"* ayant une qualité non naturelle car bruitée mais permettant de garder un haut degré d'intelligibilité. Les spectrogrammes de ce type de parole montrent que la structure formantique est maintenue. Une telle manipulation du signal peut être comparée à l'utilisation de la phase pour la reconstruction d'images. Ces auteurs reconnaissent que la phase représentée dans des spectres étudiés sur des temps courts joue un rôle insignifiant car ce type d'observation spectrale modélise la parole comme un signal quasi-stationaire pour lequel les composantes fréquentielles sont assimilées à des éléments stationnaires. D'un autre côté, ils remarquent que lors de l'observation du spectre de phase sur le long terme, la parole n'est plus modélisée comme un signal quasi stationnaire, elle est un signal pour lequel les composants fréquentiels changent dans le temps de manière dynamique. Ils en concluent que pour la parole, comme c'est le cas pour les images, l'information de la phase préserve la « situation relative des éléments » *("location of events"*) en signalant les bords des syllabes et tous les autres évènements à variation rapide *("lines edges and other narrow events"*) (ibid. p534). En d'autres termes, les spectres tenant compte de la phase sur une longue durée encodent l'emplacement des changements les plus importants dans le signal qui, dans la parole, se traduisent par des changements dans l'amplitude des composants fréquentiels. Par conséquent ont comprend que les types de combinaisons AM-FM jouent un rôle important dans les langues sifflées *au niveau de la phrase*. De surcroît, ce phénomène perceptif nous permet de comprendre pourquoi le sifflement devient si efficace sur les phrases alors qu'il est encore moyennement efficace sur les mots.

Ces résultats mettent donc en évidence que les changements fréquentiels s'accompagnant de changements d'amplitude rapides (associés à des ouvertures, des fermetures, des constrictions des couplages ou des

découplages des résonateurs du tracus vocal) jouent un rôle important d'indices acoustiques de localisation relative pour la perception. Ceci est cohérent avec les conclusions de l'analyse de la scène auditive.

L'analyse de la phase de la parole révèle également certaines limites des modes d'observation spectrale utilisant l'analyse de Fourier. Ils confirment la nécessité d'utiliser des observations sur le long terme (large bande) pour les éléments temporels. Dès lors, il nous est possible d'identifier les phonèmes pour lesquels l'analyse de Fourier à bande étroite entraîne un rendu fortement incomplet des changements de valeurs portés par la phase. Ce sont par exemple les consonnes occlusives, les affriquées (provoquées par l'ouverture et la fermeture des cavités de la bouche), les consonnes nasales (qui sont produites par un couplage et un découplage des cavités nasales et par l'ouverture et la fermeture des cavités orales), et les latérales /l/ ou /r/ (produites par un couplage et un découplage de deux cavités orales paralllèles). Les langues sifflées confirment que ces phonèmes entraînent des variations rapides de AM ou de FM (voir partie Typologie, Chapitre 3)

Ouverture sur le rythme

On remarque sur nos enregistrements réalisés à distance que les phrases enregistrées à longue distance sont parfois très limitées à une modulation du noyau de la syllabe. C'est par exemple le cas de l'expression « la montañeta » à 300 m (Figure 107), avec seulement les voyelles pour éléments de base et de légères modulations comme un tambour parleur peut le faire (Essien 2000)[111]. Cet aspect a le mérite de nous rappeler l'importance du rythme dans la phrase sifflée. L'intensité jouant le rôle de marqueur rythmique soulignant un événement fréquentiel. Nous n'avons pas fait de mesures temporelles précises sur le signal sifflé jusqu'ici car nous attendions les résultats de nos enregistrements à distance. Certains aspects de la dégradation du signal à distance permettent effectivement de préciser le choix des éléments temporels à mesurer. La forme dégradée des modulations de consonnes observées sur sonagrammes suggère que tous les éléments visibles à courte distance ne jouent pas le même rôle dans le rythme et que ce ne sont pas seulement les pics d'amplitude qui sont à prendre en compte mais plutôt des zones temporelles dans chaque syllabe qui participent à identifier les consonnes qui sont marquées par la fréquence des voyelles sur lesquelles elles s'appuient.

[111] Essien a étudié le « tambour parlant » Yoruba ou « tambour bata » qui est un tambour d'aisselle biface portable (de la forme d'un sablier) dont les peaux de percussion peuvent être tendues par un jeu de courroies pressées sous l'aisselle en même temps que l'on frappe la peau, ceci permet de faire de légères modulations des sons.

4.3.4.3.7. Conclusion et perspectives

Nous avons engagé ici une recherche qui se révèle très riche d'enseignements sur la compréhension de l'intelligibilité des langues sifflées. Le respect des conditions naturelles de dégradation du signal, bien que contraignant d'un point de vue expérimental est le point clef et déclencheur de nombreuses réflexions relativement nouvelles sur l'intelligibilité de la parole.

Nous avons donc commencé à explorer quelques pistes de recherche à la faveur d'une première analyse des enregistrements réalisés :

- Nous avons ainsi souligné la nécessité de considérer le rôle des combinaisons AM-FM en lien avec leur signification phonétique.
- Nous avons pu identifier des paramètres importants à considérer lors de la mesure du rythme d'une parole sifflée.
- Nous avons souligné l'importance de regarder le signal avec une méthodologie adaptée aux types de variations des différents indices acoustiques portés par les éléments segmentaux de la parole. Cette observation doit être réalisée à la fois à court terme et à long terme. C'est pour cela que nous avons commencé à mettre au point, avec l'aide de chercheurs spécialisés en traitement du signal, des outils adaptés au sifflement. Nous avons cherché des outils mathématiques permettant une résolution spectro temporelle plus stable que l'analyse de Fourier. Deux voies ont été ouvertes dans ce sens en fin de thèse. Nous les présentons succinctement en fin d'Annexe A. Il s'agit d'une part d'une approche par l'intermédiaire des ondelettes de Morlet, et d'autre part, d'une approche par l'intermédiaire d'une technique de paramétrisation du signal qui a l'avantage de s'adapter aux caractéristiques du sifflement tout en étant rapide et souple (cf.A.4.2.4). Nous n'avons pas cependant terminé la phase d'optimisation de ces outils. De plus, dans la perspective d'une comparaison avec le rythme de la voix parlée, il est nécessaire de faire un choix parmi les différentes méthodes de mesures temporelles proposées par les phonéticiens de la voix. Nous en sommes au stade de bilan des techniques afin de choisir celle qui est la plus cohérente avec la réalité des langues sifflées et des données perceptives de la psychoacoustique.

4.4. Conclusion pour l'intelligibilité

Pour étudier l'intelligibilité des langues sifflées nous avons décidé de procéder par étapes en nous appuyant sur des connaissances de différents domaines. De cette manière nous avons pu souligner que toutes les connaissances actuelles en physiologie acoustique de l'oreille et en psychoacoustique convergent pour soutenir que l'audition humaine privilégie la bande de fréquences des sifflements mais aussi sa dynamique (en termes de masquage et de résolution temporelle notamment). Des considérations perceptives sont aussi particulièrement adaptées pour mieux comprendre l'équilibre perception-production utilisé par les langues sifflées pour reproduire les voyelles ou les tons. En effet, leurs stratégies de transposition de la voix choisissent entre deux qualités différentes de la hauteur: la Hauteur Fondamentale (HF : tons et intonation) et la Hauteur Brute (HB : timbre). De plus, nous avons pu faire un lien entre le fait que la HF entraîne parfois des ambigüités d'octave et l'habitude des siffleurs de maintenir leur signal de parole dans un intervalle d'un octave (pour une même phrase et ce dans toutes les cultures rencontrées sauf une[112]).

Comme la parole n'est pas statique et dispose de nombreux paramètres relatifs, nous avons cherché à éclairer notre compréhension de la parole continue sifflée avec les résultats de l'analyse de la prosodie et avec ceux de l'*analyse de la scène auditive.* L'évocation de ces deux domaines de recherche a été l'occasion, d'une part de décrire comment l'audition humaine sépare et regroupe les sons de son entourage en des flux perceptifs, d'autre part de remarquer que les langues sifflées font une proposition alternative et phonologiquement pertinente de méthodologie d'étude de la prosodie du langage. Nous avons montré que les relations entre les attributs de la perception permettent de comprendre de nombreux usages phonétiques et phonologiques et se révèlent des points clefs d'une bonne intelligibilité dans le bruit. Ceci est valable à la fois pour la forme parlée et pour la forme sifflée d'une langue.

Sur ces bases, nous avons ensuite pu présenter et concevoir des expériences testant l'intelligibilité des langues sifflées non tonales du groupe 1 de notre typologie. Lors de l'expérience de psychoacoustique sur les voyelles sifflées, il est apparu que les Français identifient les voyelles sifflées /i/, /e/, /a/, /o/ avec la même logique que les siffleurs espagnols. Il est intéressant de noter que les voyelles de l'expérience étant sur plus d'un octave, il semble que les Français, comme les Espagnols soient peu perturbés par l'ambiguïté d'octave pour l'identification des voyelles. Ces résultats nous ont suggéré un rapprochement avec ceux des expériences sur l'*effective formant* de Carlson et al (1970). La pratique sifflée semble être un modèle adequat pour étudier la perception du timbre des voyelles. Elle nous a permis de confirmer qu'un des principaux éléments de perception de la voyelle tient à la compacité de la répartition de ses harmoniques fréquentielles et même de ses formants les mieux soulignés en amplitude.

Nous avons ensuite montré que les taux de netteté des logatomes, et d'intelligibilité des mots et des phrases étaient respectivement de l'ordre de 20-50%, 70%, 90-100% à courte distance. Nous avons remarqué que l'intelligibilité reste souvent élevée à grande distance. Nous avons donc concu une expérience pour observer

[112] Il semble que les langues sifflées les plus marquées par ce phénomène soient celles qui transposent les tons de la voix (mazatèque, hmong, surui) ou bien celles qui ont un grand nombre de voyelles (turc)

de manière contrôlée la dégradation du signal de parole sifflée. Notre analyse des résultats qui n'est que partielle actuellement nous a permis de comprendre l'importance des modulations dans le bruit et de trouver des limites cohérentes de netteté des syllabes. Nous avons enfin retenu que la phase et le mode d'observation des éléments segmentaux et suprasegmentaux de la parole devaient être pris en compte pour analyser l'intelligibilité de la phrase et pour mesurer les paramètres temporels de la parole.