

Chapitre 1

Perception de la parole

1.1. Introduction

En phonétique et en phonologie, l’accent est traditionnellement placé sur les mécanismes mis en œuvre dans la production de la parole, et la perception est mise au second plan. La description d’un système phonémique se fonde le plus souvent sur des critères de nature articulatoire, en faisant prévaloir le point de vue du locuteur sur celui de l’auditeur. Le fait que le mot “langue” puisse désigner à la fois l’organe articulatoire et le système linguistique reflète peut-être les excès de ce “fétichisme de la langue”, selon l’expression de Jakobson cité par [DUR 00]. Il est rarement fait référence aux mécanismes employés par l’auditeur dans les manuels de phonétique et de phonologie, et si la phonologie articulatoire a connu le succès que l’on sait (Fougeron, ce volume), une phonologie perceptive reste encore à inventer [FRA 96].

Cela constitue un paradoxe dans la mesure où la forme sonore du langage est d’abord appréhendée à travers l’oreille. En phonétique descriptive et en phonologie, notamment dans les enquêtes de terrain, l’oreille est le premier instrument d’investigation utilisé. Il est ainsi important de bien connaître le fonctionnement du système perceptif puisque les propriétés de ce système conditionnent la façon dont nous appréhendons les phénomènes phonétiques et phonologiques¹. Mais l’enjeu des recherches sur la perception de la parole dépasse largement ces aspects instrumentaux. Il touche

Chapitre rédigé par Noël NGUYEN. Je remercie Jacques Durand et Sophie Wauquier-Gravelines pour leurs commentaires sur une version antérieure de ce texte..

1. Pour citer un exemple connu, Ladefoged [LAD 93] souligne que, selon toute vraisemblance, les termes “haut-bas” et “avant-arrière” traditionnellement utilisés dans la description des voyelles, renvoient en fait à des dimensions acoustico-perceptives.

à la question de savoir quelle place doit être attribuée à l'auditeur dans les théories phonétiques et phonologiques, et de manière plus générale encore, aux relations qui s'établissent entre phonétique, phonologie et cognition (voir Laks, ce volume).

Cet enjeu a été saisi il y a longtemps déjà par Jakobson, dont la théorie de la communication parlée assignait au locuteur et à l'auditeur des places symétriques. Dans le système de traits distinctifs de Jakobson, Fant & Halle [JAK 52], les traits étaient dotés pour la plupart d'entre eux d'une dimension perceptive. Dans *The Sound Pattern of English* [CHO 68], Chomsky & Halle présentaient également les représentations phonétiques employées comme ayant une réalité perceptive². Plus récemment, la perception de la parole a été mise en avant dans différentes théories phonologiques, telles que la théorie de l'optimalité (selon laquelle certaines contraintes phonologiques trouvent leur origine dans la perception ; voir Lyche, ce volume), ou la théorie AIU dans la version proposée par [HAR 00]. Bien évidemment, la perception de la parole occupe une place centrale dans les travaux rassemblés sous le terme de phonologie de laboratoire (voir D'Imperio, ce volume).

Ce chapitre a pour objectif de présenter un aperçu général des recherches sur la perception de la parole, dans leur relation avec la phonétique et la phonologie³. Nous commençons par exposer les travaux visant à explorer les processus employés dans l'identification des phonèmes. Nous abordons ensuite les questions relatives à la forme et à la fonction des représentations phonétiques et phonologiques dans le traitement de la parole.

1.2. L'identification phonémique

Dans cette partie, nous explorons les mécanismes susceptibles d'être mis en œuvre par l'auditeur dans l'identification des phonèmes. Nous retraçons le chemin suivi par les recherches réalisées depuis un demi-siècle dans ce domaine, en montrant qu'après avoir longtemps porté sur les frontières entre phonèmes, l'accent s'est aujourd'hui déplacé sur la structure interne des catégories phonémiques⁴.

2. Même si ces représentations ont été définies en termes articulatoires, et que la façon dont le signal de parole est converti en une représentation phonétique de la séquence entendue n'a pas été explicitée.

3. Les mécanismes de traitement mis en œuvre par l'auditeur sont appréhendés ici au plan fonctionnel et nous ne présentons pas les bases neurales de la perception. Nous n'abordons pas non plus le traitement de la prosodie (voir [VAI 05] pour une récente synthèse sur le sujet).

4. Notons que les études rassemblées ici se fondent sur le postulat selon lequel le phonème possède une réalité psychologique, et font appel à des procédures expérimentales assujetties à cette hypothèse de base. Les travaux visant à établir de manière empirique la nature des unités d'analyse utilisées dans le traitement de la parole sont abordés dans la section 1.3.

1.2.1. La perception catégorielle

Des études fondamentales sur l'identification phonémique ont été menées dès les années 1950 aux laboratoires Haskins par Liberman et ses collaborateurs [LIB 96]. Ces études avaient pour objectif général d'identifier les indices acoustiques servant de support aux oppositions entre phonèmes. Les expériences sur le rôle des transitions de formant dans l'identification du lieu d'articulation, ou sur celui du VOT (*voice onset time*) dans l'identification du trait de voisement par exemple, sont aujourd'hui célèbres.

Ces premières expériences ont abouti à mettre en évidence le phénomène désigné sous le terme de perception catégorielle. On a donné différentes définitions de la perception catégorielle [REP 84]. Celle que nous utiliserons ici renvoie au fait qu'il est, dans certaines circonstances, plus facile à l'auditeur de percevoir les différences entre deux sons lorsque ces derniers se rangent dans deux catégories phonémiques différentes, plutôt que dans une même catégorie. La figure 1.1 représente de manière schématique les réponses produites par l'auditeur dans une expérience de ce type.

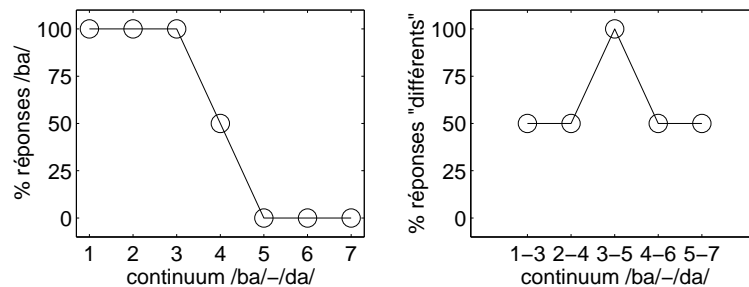


Figure 1.1. Réponses-type dans un test d'identification (à gauche) et dans un test de discrimination (à droite).

Dans cet exemple théorique, les sept stimuli utilisés sont espacés à intervalles égaux sur un continuum acoustique entre la syllabe /ba/ et la syllabe /da/⁵. Lorsque ces stimuli sont présentés à plusieurs reprises chacun et dans un ordre aléatoire à l'auditeur, dans une tâche d'identification avec choix forcé entre deux réponses possibles (/ba/ et /da/), les données généralement obtenues sont illustrées à gauche sur la figure 1.1, sous la forme d'une courbe d'identification (pourcentage de réponses /ba/). La figure montre que l'auditeur associe les stimuli 1-3, à gauche du continuum, à la catégorie /ba/, mais que sa réponse "bascule" vers la droite du continuum au profit de l'autre catégorie proposée (/da/), le stimulus 4 situé au centre du continuum étant

5. Phonétiquement, chaque stimulus se présente sous la forme d'une voyelle synthétique dont les formants varient en fréquence dans la partie initiale.

jugé ambigu. Lorsque les stimuli sont présentés par paires, et que la tâche de l'auditeur est de dire pour chaque paire si les stimuli dont elle se compose sont identiques ou différents, le patron de réponse observé est représenté à droite sur la figure, sous la forme d'une courbe de discrimination (pourcentage de réponses "différents"). La courbe révèle que l'auditeur ne perçoit pas les différences entre stimuli à gauche et à droite sur le continuum (en répondant "différents" une fois sur deux, c.à.d. au hasard), alors que la différence qui s'établit entre les stimuli 3 et 5 est correctement détectée. Tout se passe donc comme si l'auditeur faisait abstraction des variations acoustiques à l'intérieur de chaque catégorie phonémique, pour se focaliser sur les variations inter-catégorielles susceptibles de différencier les phonèmes les uns des autres.

Ces recherches ont été étendues quelques années plus tard à la perception des voyelles [FRY 62]. Contrairement aux résultats établis pour les occlusives, il a été montré que l'auditeur percevait les différences entre voyelles que celles-ci soient associées ou non à une même catégorie phonémique. Les auditeurs se montraient ainsi capables de discriminer deux stimuli associés tous les deux à /ɛ/, aussi bien qu'ils percevaient la différence entre un /ɛ/ et un /æ/.

Ce résultat laissait supposer que les consonnes et les voyelles faisaient intervenir deux modes de traitement différents. Il a en cela été interprété comme un argument majeur en faveur de la théorie motrice de la perception de la parole, développée à Haskins [LIB 67, LIB 85], et dans laquelle un lien étroit est postulé entre perception et production de la parole. Selon cet argument, il n'y a pas de frontières bien établies entre voyelles dans l'espace articulatoire, et il est ainsi possible de passer d'une voyelle (/ɛ/) à une autre voyelle (/æ/) de manière continue et graduelle, en modifiant progressivement la configuration géométrique du conduit vocal. En revanche, les consonnes occlusives seraient séparées par un ensemble de frontières articulatoires naturelles. Le passage entre un /p/ et un /t/ par exemple, revêtirait nécessairement un caractère discontinu, dans la mesure où ces deux consonnes font appel à des articulateurs actifs différents. Dans la théorie motrice, l'auditeur interprète le signal de parole en référence aux mouvements articulatoires dont celui-ci est le produit, et la présence de ces frontières articulatoires dans l'espace des consonnes expliquerait le fait que les consonnes soient perçues sur un mode catégoriel, alors que les voyelles le seraient sur un mode continu.

À la fin des années 60, des investigations réalisées par Fujisaki et Kawashima [FUJ 71] sur de nouvelles classes de consonne (fricatives, glides, liquides) ont révélé que la perception de ces stimuli n'était pas aussi catégorielle que celle des consonnes occlusives. Fujisaki et Kawashima ont également montré que même les voyelles pouvaient être perçues sur un mode catégoriel dans des conditions défavorables, et notamment lorsqu'elles étaient de courte durée. Ces travaux entraient en contradiction avec la théorie motrice en donnant à penser que la perception catégorielle n'était pas un mode de traitement propre à une certaine classe de sons. Ils ont permis d'établir qu'un même son pouvait être perçu de façon plutôt catégorielle, ou plutôt continue,

en fonction de la situation expérimentale. Fujisaki et Kawashima ont interprété leurs résultats dans le cadre d'un modèle qui attribuait un rôle central à la mémoire auditive à court terme⁶. Ce modèle apportait une première explication au fait que l'auditeur conserve dans certaines conditions la capacité de percevoir des différences entre stimuli à l'intérieur d'une même catégorie phonémique, chose que les partisans de la théorie motrice avaient d'abord tenue pour inintéressante et dépourvue de véritable signification psychologique.

Ces résultats ont soulevé la question de savoir dans quelle mesure la perception catégorielle était conditionnée par la nature de la réponse demandée au sujet. Dans la grande majorité des expériences réalisées, le sujet répondait en effectuant un choix parmi un petit nombre de réponses prédéfinies (/p/, /b/; "pareil", "différent"). La perception dite catégorielle était peut-être partiellement induite par cet éventail limité de choix possibles. Dans les années qui ont suivi, de nouvelles expériences ont apporté une confirmation à cette idée, en conduisant les chercheurs à établir une distinction entre perception catégorielle et réponse catégorielle. Cette distinction a notamment été mise en avant par Massaro [MAS 83] dans les années 80. Plutôt que d'imposer une réponse discrète, Massaro a demandé à ses auditeurs d'employer une échelle numérique (en attribuant à chaque stimulus une certaine valeur sur cette échelle, entre X et Y). Les réponses observées présentaient un caractère continu. Selon Massaro, l'auditeur peut dans certaines circonstances répondre sur un mode catégoriel (lorsque l'éventail des réponses proposées le force à le faire par exemple). En revanche, les informations extraites par l'auditeur du signal de parole conservent dans tous les cas un caractère continu et graduel.

Pour Massaro, il serait en fait désavantageux pour l'auditeur de traiter les sons de la parole sur un mode catégoriel, en passant par le filtre d'un ensemble d'étiquettes phonétiques. En faisant abstraction des informations acoustiques détaillées sur le signal de parole, l'auditeur perdrait la possibilité d'exploiter ces informations ultérieurement, pour corriger de possibles erreurs d'identification relatives au phonème ou au mot précédent par exemple. La stratégie la plus efficace consisterait au contraire à conserver autant d'information que possible sur le signal le plus longtemps possible. La perception continue de la parole répondrait à ces conditions-là. Ces hypothèses sont à l'origine du modèle FLMP (*Fuzzy Logical Model of Perception*) développé par Massaro.

Il n'est pas exclu de penser que l'auditeur puisse procéder à un double codage de l'information, continu (sous la forme d'une sorte de spectrogramme auditif par exemple) et discret (sous la forme d'un ensemble d'étiquettes phonétiques). Dans une telle hypothèse, la perception catégorielle constituerait davantage qu'un artefact expérimental, et serait à mettre en relation avec ce processus de codage symbolique et

6. Modèle porté à la connaissance des chercheurs occidentaux par Pisoni [PIS 75].

discret. L'importance relative de ces deux formes de codage pourrait dans une certaine mesure dépendre de la tâche assignée au sujet.

1.2.2. La structure interne des catégories phonémiques

Les chercheurs de Haskins avaient commencé par considérer que l'auditeur était insensible aux différences intra-catégorielles. Comme nous l'avons vu, cette hypothèse a ensuite été rejetée. Des études détaillées ont alors été entreprises sur la manière dont l'auditeur perçoit les sons rattachés à une même catégorie phonémique. Selon Joanne Miller [MIL 94], en particulier, les catégories phonémiques sont dotées d'une véritable structure interne, riche et graduée. Ces investigations marquent un nouveau tournant dans l'histoire des travaux sur la perception de la parole, dans la mesure où l'attention des chercheurs se détourne progressivement des frontières phonémiques, pour se focaliser sur l'organisation interne de chaque catégorie.

Les expériences réalisées dans ce domaine ont fait appel à différentes tâches, dont le jugement de qualité (*judgement of category goodness*). Cette tâche vise à explorer la structure interne d'une catégorie de manière explicite en demandant à l'auditeur d'attribuer au stimulus une valeur sur une échelle de qualité. Par exemple, on fait entendre à l'auditeur une série de syllabes /pi/ présentant des variations de VOT, et l'auditeur a pour tâche de dire dans quelle mesure la consonne initiale représente bien la catégorie /p/, en utilisant une échelle de 1 (mauvais exemple) à 10 (excellent exemple). Les résultats montrent que tous les stimuli ne sont pas perçus comme représentant au même degré la catégorie phonémique qui leur est associée. Une distinction se dégage ainsi entre les "bons" et les "mauvais" exemplaires de la catégorie. Le meilleur d'entre eux est souvent désigné sous le terme de prototype, et la qualité perçue des stimuli se dégrade au fur et à mesure que l'on s'éloigne du prototype dans l'espace acoustique. La figure 1.2 représente sous une forme schématique le patron de réponse obtenu. Le prototype coïncide avec le sommet de la courbe de qualité.

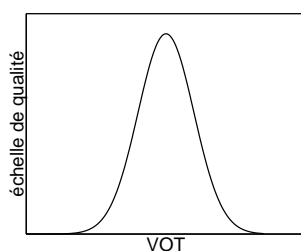


Figure 1.2. Représentation schématique de la qualité perçue de l'occlusive dans /pi/, en fonction de la durée du VOT.

Ces expériences ont été en partie inspirées par un courant de recherche en psychologie cognitive dont l'objectif est d'étudier les processus généraux employés dans la

catégorisation des formes sensorielles. On peut notamment citer Rosch [ROS 76] et Nosofsky [NOS 86]. L'une des questions centrales soulevées dans ces travaux porte sur la façon dont les catégories sont représentées (sous la forme de prototypes ou de liste d'exemplaires).

Dans cette perspective, un son de parole est associé à une valeur continue pour l'auditeur, selon que ce son représente plus ou moins bien la catégorie phonémique correspondante. Cette valeur est parfois interprétée comme une probabilité d'appartenance à la catégorie (elle est alors comprise entre 0 et 1). À notre connaissance, la plupart des modèles actuels du traitement de la parole se placent dans ce cadre probabiliste. Cela est vrai pour les modèles connexionnistes (réseaux de neurones artificiels, voir [MCC 86]), le modèle FLMP de Massaro, etc. Les modèles d'aujourd'hui se différencient fortement sur ce point d'un modèle générique plus ancien, basé sur une série de décisions binaires, et dans lequel un son de parole appartiendrait de façon univoque à une catégorie et à elle seule (*cf.* le modèle du type traitement de l'information proposé par Jakobson par exemple).

Notons malgré tout que la notion de structure interne reste définie de façon assez élémentaire, dans la mesure où elle renvoie essentiellement au fait que les membres d'une catégorie phonémique n'appartiennent pas tous au même degré à cette catégorie pour l'auditeur.

1.2.3. *Des frontières aux prototypes*

La notion de prototype est au centre de la théorie de la perception de la parole proposée par Kuhl au début des années 90, la théorie des aimants perceptifs (*perceptual magnets*, [KUH 91]). Kuhl défend elle aussi l'hypothèse selon laquelle tous les membres d'une catégorie ne sont pas perçus par l'auditeur comme étant équivalents. Selon Kuhl, les "bons" exemplaires, ou prototypes, ont un rôle central à jouer dans la structuration perceptive de l'espace phonétique, dans la mesure où c'est à partir de ces prototypes que s'opère la catégorisation des sons de la parole. On s'éloigne davantage encore de la théorie d'abord défendue à Haskins, avec son attachement à la notion de frontière inter-catégorielle, et le peu d'attention qu'elle accordait aux différences perçues par l'auditeur à l'intérieur de chaque catégorie.

Kuhl attribue un statut particulier aux prototypes, qui selon elle exercent un effet d'"attraction" perceptive sur les sons qui les entourent. Prenons l'exemple d'une voyelle placée dans le voisinage du prototype associé à /i/ dans l'espace vocalique. Kuhl a montré que l'auditeur a tendance à assimiler perceptivement la voyelle entendue au prototype. En d'autres termes, l'auditeur perçoit plus difficilement les différences entre voyelles dans le voisinage du prototype, qu'à la périphérie de la catégorie. Ce phénomène peut être décrit de façon métaphorique en disant que le prototype attire

à lui les sons qui lui ressemblent sur le plan acoustique, ou encore que l'espace perceptif se contracte autour du prototype. La figure 1.3 illustre de manière schématique ce phénomène.

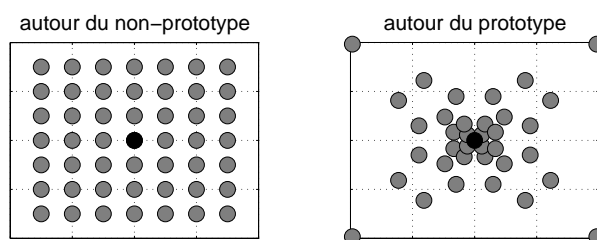


Figure 1.3. Représentation schématique des distances perçues autour d'une voyelle non-prototype et autour d'une voyelle prototype. Le non-prototype et le prototype sont tous les deux représentés en noir.

Hawkins [HAW 99] souligne que la différence de point de vue avec les études sur la perception catégorielle est intéressante. Dans l'effet d'aimant perceptif, ce que nous voyons n'est pas tant la mise en relief des différences de part et d'autre d'une frontière catégorielle, que la réduction des différences au centre de la catégorie. La perception d'une frontière entre catégories est alors attribuée au simple fait que les sons situés à mi-chemin entre deux prototypes échappent à l'attraction de ces prototypes. Les frontières inter-catégorielles se présentent ainsi comme étant assujetties aux sons prototypes.

1.2.4. Effets de contexte dans l'identification phonémique

De nombreux travaux ont été réalisés dans le but d'explorer l'influence du contexte sur l'identification phonémique. Le terme de contexte est employé ici au sens large, et recouvre le contexte phonétique et le contexte lexical, entre autres choses. On sait bien sûr que les phonèmes présentent d'importantes variations dans leur forme articulaire et acoustique en fonction du débit, du contexte phonétique dans lequel ils apparaissent, ou du locuteur, pour ne citer que quelques facteurs. La fricative /s/ par exemple, ne sera pas prononcée de la même manière dans "sou" et "si". Dans le mot "sou", la fricative est articulée avec les lèvres arrondies sous l'influence de la voyelle suivante, et cet arrondissement labial a d'importantes conséquences sur la forme spectrale de la fricative, avec un déplacement du pic spectral affilié à la cavité antérieure vers les basses fréquences. Selon de nombreux auteurs, ces phénomènes de coarticulation ne font pas obstacle à l'identification des phonèmes par l'auditeur. Il est au contraire supposé que l'auditeur exploite ces effets de contexte pour identifier les phonèmes plus facilement [PER 86]. Cela lui est possible dans la mesure où les effets de contexte revêtent un caractère systématique et régulier. On considère que l'auditeur utilise ces

régularités à son profit en se rapportant d'une manière ou d'une autre au contexte pour identifier chaque phonème. Cette hypothèse est implémentée sous une forme numérique dans le modèle TRACE [MCC 86], par exemple⁷.

Les études réalisées dans ce domaine montrent que l'auditeur se comporte bien comme s'il prenait le contexte phonétique en compte dans l'identification des phonèmes. Selon certains auteurs, l'auditeur ajuste les frontières qu'il établit entre phonèmes en fonction du contexte [MIL 94, REP 87]. On désigne parfois ce phénomène sous le terme de compensation perceptive.

On s'est également beaucoup interrogé sur le rôle possible du lexique dans la perception de la parole. Ces travaux visent à déterminer dans quelle mesure l'auditeur tire parti de ses connaissances lexicales pour identifier les phonèmes dont un mot est composé. On peut par exemple supposer que ces connaissances lexicales nous permettront de reconnaître un phonème mal prononcé ou masqué par un bruit. Les célèbres expériences de [WAR 70] sur la restauration phonémique ont fait apparaître que le lexique permettait à l'auditeur de reconstituer la forme sonore d'un phonème remplacé par du bruit dans une phrase. Selon Warren, le lexique donne lieu à un véritable phénomène d'illusion perceptive, dans la mesure où l'auditeur "entend" le phonème bien que le segment acoustique correspondant ait été supprimé. Selon l'une des hypothèses avancées pour expliquer ce phénomène, un transfert d'information de haut en bas (du lexique vers les niveaux inférieurs de traitement) a lieu dans la perception de la parole. On parle d'un effet descendant du lexique sur la perception de la parole, ou effet top-down [ELM 88]. L'existence de ces effets est loin d'être reconnue par tous les auteurs (voir par exemple [NOR 00, MCQ 01] pour un état des discussions actuelles).

1.3. Forme et fonction des représentations phonétiques et phonologiques dans le traitement de la parole

Les travaux que nous venons de passer en revue visent à étudier comment l'auditeur identifie des phonèmes lorsque la situation expérimentale l'amène à le faire. Ces travaux ne se placent pas en position de déterminer si les unités phonémiques jouent un véritable rôle dans la compréhension du langage oral par l'auditeur. Nous abordons à présent les études cherchant à explorer plus directement la façon dont l'auditeur se représente le signal de parole à chaque étape de traitement.

7. On peut se reporter à [NGU 01] pour une synthèse récente sur le rôle de la coarticulation dans la reconnaissance des mots.

1.3.1. *Unités perceptives de base*

Il est souvent supposé que le signal de parole est décomposé sous la forme d'une séquence d'unités perceptives de base (ex. : phonèmes) à partir desquelles il est possible à l'auditeur d'accéder à des unités linguistiques à la fois plus larges (ex. : mot) et plus petites (ex. : traits distinctifs). On considère souvent que ces unités se situent à un niveau infralexical intermédiaire entre le signal de parole et le lexique. De très nombreux travaux ont été réalisés dans le but de déterminer la nature de ces unités perceptives⁸.

Ces travaux ont souvent fait appel à la tâche de détection de fragment. Dans cette tâche, le sujet a pour consigne de déterminer le plus vite possible si une cible préétablie (phonème ou séquence de phonèmes) est contenue ou non dans le stimulus acoustique présenté. En 1970, Savin et Bever [SAV 70] ont montré qu'une syllabe entière était détectée plus rapidement que la consonne initiale de cette même syllabe, dans une séquence de syllabes sans signification en anglais. Selon Savin et Bever, ces résultats démontraient que les syllabes priment sur les phonèmes dans le traitement de la parole. D'importantes améliorations à la méthode employée ont été apportées quelques années plus tard par Mehler, Dommergues, Frauenfelder et Segui [MEH 81]. Dans cette célèbre expérience réalisée sur le français, la cible à détecter était de type CV (ex. : BA), ou de type CVC (ex. : BAL), et la séquence entendue comportait deux syllabes dont la première était elle-même soit de type CV (ex. : "balance"), soit de type CVC (ex. : "balcon"). Les résultats ont révélé que la cible visuelle était détectée plus vite lorsqu'elle coïncidait avec la syllabe initiale du mot porteur, indépendamment de la longueur de cette cible (2 ou 3 phonèmes). Mehler et coll. ont eux aussi interprété leurs résultats en faveur de l'idée selon laquelle la syllabe est une unité perceptive de base.

Ultérieurement, différentes études ont suggéré que cet "effet syllabique" était peut-être spécifique à certaines langues dont le français⁹. Cutler, Mehler, Norris et Segui [CUT 86] ont ainsi montré que les anglophones faisaient probablement appel à une stratégie de traitement différente. Ces variations inter-langues dans les patrons de réponse ont été mises en relation avec les différences entre le système phonologique de l'anglais et celui du français. Dans une série d'expériences récentes, Content, Frauenfelder et coll. [CON 02] ont entrepris de répliquer l'effet obtenu par Mehler et coll. [MEH 81] pour le français. Malgré le caractère bien établi que l'on a pu attribuer à cet effet dans la littérature depuis plus de vingt ans, toutes leurs tentatives se sont soldées par des échecs. Leurs résultats donnent cependant à penser que certains constituants syllabiques, et notamment ce que les auteurs désignent sous le terme d'attaque, ont un

8. Notons qu'une distinction – qui dépasse le cadre de cette présentation générale – est parfois établie entre unités de segmentation et unités de catégorisation, voir [CUT 88].

9. Voir [CHR 91] et [SEG 02] pour des synthèses de ces travaux.

rôle important à jouer dans la perception de la parole, en servant de point d’ancrage pour l’accès au lexique [CON 01].

D’autres chercheurs soutiennent une hypothèse concurrente en vertu de laquelle l’unité primaire de traitement dans la parole est le trait. Cela est notamment le cas dans le modèle Cohort de Marslen-Wilson, dans sa version révisée [MAR 94], dans le modèle LAFF (*Lexical Access from Features*) de Stevens [STE 02], et dans le modèle FUL (*Featurally Underspecified Lexicon*) de Lahiri [LAH 02]. Ces trois modèles postulent que l’auditeur établit une correspondance directe entre une représentation en traits du signal de parole et le lexique, sans passer par un niveau phonémique ou syllabique. Différents résultats expérimentaux ont été invoqués en faveur de cette hypothèse. Streeter et Nigro [STR 79] par exemple, ont montré qu’un auditeur se montrait sensible à une distorsion locale artificiellement introduite dans le stimulus pour des mots mais pas pour des non-mots, dans une tâche dite de décision lexicale. Un tel patron de réponse suggère que l’accès au lexique ne s’effectue pas par l’intermédiaire d’une séquence de phonèmes ou de syllabes¹⁰. Les résultats obtenus par Marslen-Wilson et Warren [MAR 94], à partir d’un design expérimental beaucoup plus élaboré, vont dans le sens d’un accès lexical direct à partir du trait. Notons cependant que ces résultats ont été récemment remis en question par [MCQ 99].

En dépit de ce long débat, il ne semble pas se dégager de consensus en ce qui concerne la nature de l’unité perceptive de base. Goldinger & Azuma [GOL 03] soulignent que la liste des candidats proposés s’est en fait allongée au fil des années. Certains auteurs ont récemment suggéré que le problème a en fait été peut-être mal posé, et que les auditeurs sont simultanément sensibles à des unités de taille différente dans le signal de parole. Dans la théorie de la résonance adaptative proposée par Grossberg [GRO 03] par exemple, des unités de petite et de grande taille sont simultanément activées, avec un biais naturel en faveur des unités les plus larges. Goldinger & Azuma [GOL 03] ont obtenu des résultats qui vont dans ce sens-là, en montrant que l’on peut faire émerger des unités de taille différente à la conscience de l’auditeur en fonction du protocole expérimental utilisé. Il est ainsi possible qu’une compétition ait lieu parmi différentes unités dont le domaine temporel est conditionné par des facteurs phonologiques, lexicaux et grammaticaux, mais aussi par la dynamique de l’interaction conversationnelle et par les contraintes de la situation expérimentale dans laquelle l’auditeur est placé. Nous renvoyons le lecteur à [HAW 03a, LOC 03] pour une récente discussion à ce sujet.

10. Dans un tel cas en effet, la distorsion devrait être détectée à ce niveau infra-lexical et exercer un effet sur la réponse du sujet que le stimulus soit un mot ou un non-mot.

1.3.2. Représentations lexicales

Dans les articles de synthèse sur la perception de la parole, il est fréquent que l'auteur commence par souligner la grande variabilité du signal de parole, et les problèmes qui en découlent pour l'auditeur (voir par exemple [FOW 86]). La façon dont l'auditeur fait face à cette variabilité a donné lieu à différentes hypothèses contradictoires. Nous établirons ici une distinction schématique entre les modèles abstractionnistes et les modèles à exemplaires. Dans l'approche abstractionniste, l'auditeur associe à chaque mot une représentation phonologique abstraite et indépendante des caractéristiques individuelles du locuteur. Dans les modèles à exemplaires, l'auditeur se représente mots et constructions grammaticales de manière concrète et détaillée, sous la forme de listes d'exemplaires et/ou de prototypes.

Les tenants de l'approche abstractionniste insistent souvent sur le fait que les mécanismes de traitement de la parole semblent très résistants au bruit et à la variabilité intra- et inter-individuelle. Par ailleurs, certaines études semblent montrer que les auditeurs sont sensibles aux variations temporelles dans la forme spectrale globale du signal, par opposition aux changements à court terme portant sur des propriétés acoustiques détaillées. Ces résultats ont été établis par Remez et ses collaborateurs [REM 94], grâce à la parole dite sinusoïdale (*sinewave speech*)¹¹. Le fait que la parole sinusoïdale soit intelligible a conduit Remez à affirmer que ces patrons de variations prédominent sur la structure acoustique fine du signal pour l'auditeur.

L'approche abstractionniste revêt peut-être sa forme la plus radicale avec le modèle FUL (*Featurally Underspecified Lexicon*) de Lahiri et coll. [LAH 02]. Dans ce modèle, chaque morphème est associé à une représentation phonologique unique, abstraite et sous-spécifiée, indépendamment des variations dont ce morphème peut être le siège en surface. Dans un célèbre travail sur la perception des voyelles nasalisées en bengali et en anglais [LAH 91], Lahiri et Marslen-Wilson ont affirmé que l'auditeur serait en fait *insensible* aux variations présentées par un mot dans sa forme de surface, en particulier lorsque ces variations sont conditionnées par des phénomènes d'assimilation, parce que le signal de parole serait interprété directement à partir des représentations phonologiques stockées dans le lexique mental.

FUL suppose en particulier que certains traits, tels que [coronal] en anglais et en allemand, ne sont pas spécifiés dans le lexique. Le caractère sous-spécifié des représentations phonologiques lexicales conduit Lahiri à supposer que la relation entre signal de parole et lexique peut revêtir trois formes possibles, a) correspondance (*match*), b)

11. La parole sinusoïdale est formée par un petit nombre de sinusoïdes dont la fréquence et l'amplitude sont manipulées de façon à refléter l'évolution temporelle des pics spectraux dans un signal naturel pris pour modèle.

non-correspondance (*mismatch*), et c) absence de non-correspondance (*no-mismatch*), comme dans les exemples suivants.

<i>Trait détecté dans le signal</i>	<i>Lexique</i>	<i>Type de relation</i>
labial	labial	match
coronal	labial	mismatch
labial	place non spécifiée	no-mismatch
coronal	place non spécifiée	no-mismatch

Selon Lahiri, ce dispositif permet d'expliquer qu'un mot soit correctement reconnu par l'auditeur malgré les déformations dont il est parfois le siège, notamment en vertu des phénomènes d'assimilation. En anglais, on sait que les coronales en fin de mot sont sujettes à l'influence assimilatrice de la consonne suivante, ex. : "Byrd concert" [bɜ :gkɔnsət], "did gardens" [dɪgga :dnz], "bride must" [braɪbmɪst] [WRI 89]. On s'est beaucoup interrogé sur la nature des mécanismes permettant à l'auditeur d'identifier correctement un mot assimilé, malgré le fait que la consonne finale soit ainsi transformée. Dans FUL, il est supposé que l'auditeur est insensible aux modifications présentées par le lieu d'articulation de la consonne finale, dans la mesure où les coronales ne sont pas spécifiées pour ce trait. Si une consonne labiale ou dorsale est détectée dans le signal, cela n'exclut pas la possibilité que le mot entendu se termine par une coronale. L'assimilation n'est pas traitée comme une "déviation" vis-à-vis d'une forme idéale, dans la mesure où cette forme est inexistante dans le lexique. Lahiri et Reetz [LAH 02] donnent les exemples suivants :

<i>Contexte</i>	<i>Forme prononcée</i>	<i>Forme lexicale</i>	<i>Type de relation</i>
"Where could Mr Bean be ?"	[bi :m]	/biN/	no-mismatch
" Green bag"	[gri :m]	/griN/	no-mismatch
" Green grass"	[gri :ŋ]	/griN/	no-mismatch

Dans ces exemples, *Bean* est correctement identifié à partir de [bi :m] dans [bi :m bi], *Green* est identifié à partir de [gri :m] dans [gri :m bæŋ], et à partir de [gri :ŋ] dans [gri :ŋ grɑs], parce qu'il y a une relation de type no-mismatch entre la forme de surface et la forme phonologique sous-jacente. L'assimilation ne fait pas obstacle à l'identification du mot cible pour la simple raison qu'elle n'est pas détectée.

Il est important de souligner que la relation de no-mismatch est asymétrique. Une coronale détectée dans le signal par l'auditeur n'activera pas les mots se terminant par des labiales ou des dorsales, dans la mesure où ces consonnes sont spécifiées pour le lieu d'articulation, et sont ainsi incompatibles avec la coronale. La validité du modèle

FUL repose en partie sur ces asymétries perceptives, dont Lahiri et coll. s'attachent aujourd'hui à démontrer l'existence.

Les modèles abstractionnistes font souvent intervenir une procédure de normalisation [JOH 05], mise en application de manière automatique à une étape précoce du traitement, et permettant à l'auditeur de faire abstraction de la variabilité inter-individuelle (en rapportant par exemple les caractéristiques acoustiques d'une voyelle à la longueur estimée du conduit vocal du locuteur). Ces modèles s'inscrivent dans une perspective dualiste postulant l'existence de deux niveaux de représentation : une représentation de surface, physique et variable, et une représentation phonologique sous-jacente, abstraite et symbolique. Dans l'approche adoptée par Coleman [COL 03] et Local [LOC 03] (dérivée de la *Firthian Prosodic Analysis*), le lien entre représentations phonologiques et formes sonores est établi par l'intermédiaire d'une relation d'interprétation (*phonetic exponency*) qui revêt un caractère arbitraire. [COL 03] souligne ainsi la nature abstraite d'un trait phonologique tel que [voice], arbitrairement associé avec un vaste ensemble de propriétés phonétiques distribuées sur un intervalle étendu dans le signal et qui ne présentent pas toujours de relation intrinsèque les unes avec les autres. Les modèles abstractionnistes visent d'abord à rendre compte du traitement de la parole chez l'adulte. Appliqués à l'acquisition de la parole par l'enfant, ils donnent à supposer que les représentations phonologiques sont prédéterminées génétiquement et que l'input acoustique présenté à l'enfant a simplement pour fonction de déclencher leur mise en place (Wauquier-Gravelines, ce volume). Les travaux menés dans une perspective abstractionniste sur la perception des langues secondes se sont focalisés sur les limites de la plasticité présentée par le système perceptif chez l'adulte, sous l'influence du système phonologique maternel, notamment au travers de l'étude de certaines illusions perceptives [DUP 99, PAL 00, PAL 01].

Dans l'approche abstractionniste, l'auditeur se représente les mots sous une forme minimaliste réduite à un petit nombre de traits distinctifs. La perception de la parole s'apparente à un processus de réduction de l'information, et elle conduit à convertir le signal de parole en une séquence de représentations symboliques dont le degré d'abstraction augmente à chaque étape du traitement. Cette approche est en partie basée sur un postulat selon lequel la mémoire humaine est de taille limitée, et qu'il est donc essentiel à l'auditeur d'économiser ces ressources en mémoire dans le traitement de la parole. Ce postulat est peut-être lié à la métaphore informatique employée pour appréhender le fonctionnement du système cognitif dans les années 1950-60, à une époque où les ordinateurs étaient dotés d'une mémoire très limitée, avec les contraintes qui en découlaient sur le traitement des données.

Nous savons aujourd'hui que l'étendue de la mémoire humaine est bien supérieure à ce que l'on croyait alors. Les modèles à exemplaires attribuent à l'auditeur la capacité de stocker en mémoire les différentes formes de surface associées à un mot. Ces

modèles se présentent sous de multiples versions élaborées par des chercheurs d’horizons différents : phonéticiens (ex. Hawkins [HAW 03a], Johnson [JOH 97]), phonologues (ex. Bybee [BYB 01], Coleman [COL 02], Pierrehumbert [PIE 02]), psycholinguistes (ex. Goldinger [GOL 96, GOL 98]), spécialistes de l’acquisition (ex. Jusczyk [JUS 93])¹². Les modèles à exemplaires offrent une solution “représentationnelle”, selon l’expression proposée par Johnson, au problème de la variabilité du signal de parole, en supposant que cette variabilité est encodée par l’auditeur au niveau lexical. Dans le modèle X-MOD développé par Johnson [JOH 97], toutes les formes de surface associées à un mot et entendues par l’auditeur sont stockées en mémoire, en permettant qu’un accès direct soit établi dans le traitement avec toutes les variations déjà rencontrées. Lorsque l’auditeur doit reconnaître un mot, les propriétés acoustiques de ce mot sont comparées à chacun des exemplaires, et l’exemplaire est activé proportionnellement à son degré de similarité avec le mot d’entrée. La somme des activations pour tous les exemplaires associés à la même unité lexicale permet à l’auditeur de savoir si le mot entendu doit être ou non considéré comme appartenant à cette catégorie. Par opposition avec l’approche abstractionniste, une relation naturelle d’analogie vient s’établir entre le signal d’entrée et les exemplaires stockés dans le lexique.

Nous avons brièvement exposé la façon dont les mots se terminant par une coronale assimilée sont identifiés selon l’approche abstractionniste. La réponse apportée à ce problème par les modèles à exemplaires se montre très différente. Les partisans de ces modèles font d’abord valoir un point fondamental : même dans les coronales qui se présentent comme étant complètement assimilées, le signal de parole contient peut-être une information résiduelle associée à la coronalité. Différentes études articulatoires (voir par ex. Nolan [NOL 92]) donnent à penser que les alvéolaires assimilées font malgré tout intervenir ce que l’on désigne sous le terme de geste alvéolaire résiduel, et ce geste peut à son tour laisser une trace dans le signal [LOC 03]. On suppose que l’auditeur prête attention à cette information dans la reconnaissance des mots (au lieu de lui être insensible, comme cela est postulé dans l’approche abstractionniste). On suppose enfin que différentes représentations phonétiques détaillées et assujetties au contexte sont contenues dans le lexique mental de l’auditeur pour le même mot, et qu’elles incluent notamment une version avec coronale finale assimilée et une version avec coronale finale non-assimilée de ce mot.

On peut considérer que le modèle LAFS (*Lexical Access from Spectra*), proposé par Klatt [KLA 79], constitue l’un des premiers modèles à exemplaires. Fortement inspiré par le système Harpy de reconnaissance automatique de la parole développé à cette époque, ce modèle n’attribuait aucune place aux représentations symboliques

12. Cette liste ne vise pas à être exhaustive.

dans le traitement de la parole. Chaque mot était représenté sous la forme d'une séquence de spectres acoustiques, et le lexique était assimilé à un immense treillis comportant toutes les séquences de spectres possibles associées à toutes les combinaisons possibles de mots en anglais. Les variations dans la forme de surface d'un mot en fonction du contexte (ex. : déletion du /t/ final dans "list some" [lɪs :ʌm]) étaient ainsi précompilées dans ce treillis. La reconnaissance d'un mot s'accomplissait en établissant une correspondance directe entre le signal de parole (lui-même converti en une séquence de spectres) et ce treillis. Il est intéressant de constater que ces hypothèses resurgissent aujourd'hui dans certains travaux de phonologie (ex. [COL 02]).

Goldinger [GOL 96, GOL 97, GOL 98] a entrepris d'appliquer à la perception de la parole un modèle générique de la formation des concepts appelé MINERVA [HIN 86]. MINERVA est un modèle dit épisodique¹³ et il repose sur l'hypothèse selon laquelle chaque expérience crée une trace indépendante en mémoire, dans laquelle sont intégrés tous les détails perceptifs, le contexte, etc. La reconnaissance d'un mot se produit dans MINERVA de la manière suivante. Pour chaque mot connu de l'auditeur, un vaste ensemble de traces partiellement redondantes sont présentes dans la mémoire. Lorsqu'un mot est entendu par l'auditeur, une sonde à l'image de ce mot (*analog probe*) est mise en correspondance avec toutes les traces en parallèle. Les traces sont activées par cette sonde proportionnellement à leur degré de similarité mutuelle.

Les modèles connexionnistes de la perception de la parole s'inscrivent également dans cette approche [PRO 99]. Ils présupposent que tous les contrastes phonétiques observables entre mots peuvent être mis à profit dans la reconnaissance des mots (pour autant du moins que ces contrastes soient introduits dans la représentation d'entrée), dans la mesure où l'information circule à l'intérieur du réseau sous la forme d'un ensemble de paramètres continus. Dans TRACE [MCC 86] par exemple, de petits détails dans la structure acoustique du signal se traduisent par des variations quantitatives dans le niveau d'activation des détecteurs de phonème, et ces variations ont à leur tour un effet sur le niveau d'activation des détecteurs de mots. Les réseaux de neurones récurrents [ELM 90] ont été introduits plus récemment dans le but de simuler le décours temporel du traitement de l'information dans la perception de la parole et la reconnaissance des mots [GAS 95, GAS 03, NOR 94, PLA 99, SHI 93]. Gaskell [GAS 95, GAS 03] par exemple, a montré comment l'identification des mots avec coronale finale assimilée peut être modélisée au moyen d'un réseau de neurones récurrents exposé à des patrons d'assimilation à différents degrés, tels qu'ils se rencontrent dans la parole naturelle.

Contrairement à ce qui est parfois supposé, les modèles à exemplaires présentent d'importantes différences avec une approche du type WYSIWYG (*what you see is*

13. La notion d'épisode renvoie à la trace laissée dans la mémoire à long terme par un stimulus.

what you get) dans laquelle le lexique mental se réduirait à une vaste liste a-structurée de traces auditives. Ces modèles se caractérisent par un certain nombre de propriétés émergentes dont la pertinence sur le plan phonétique et phonologique commence à se faire jour. L'activation des exemplaires associés à chaque mot donne ainsi lieu à la formation d'une trace générique assimilable à un prototype, et revêtant un caractère plus abstrait. La particularité essentielle des modèles à exemplaires tient au fait que ces phénomènes de généralisation se produisent en temps réel et de manière dynamique¹⁴, pendant l'activation des mots dans le lexique, au lieu de s'établir de manière permanente lors la mise en place de ce lexique [JOH 05]. Les modèles à exemplaires se montrent également capables de segmenter le signal de parole sous la forme d'une séquence d'unités infra-lexicales (traits, phonèmes, syllabes), qui émergent dynamiquement dans la mise en relation entre le signal et le lexique [COL 03, GOL 03, GRO 03, JOH 97]. Il est supposé que ces processus de segmentation tiennent en partie à la structure hautement élaborée du lexique, à l'intérieur duquel des liens s'établiraient entre les différentes unités en fonction de leur degré de similarité sur le plan phonétique et sur le plan sémantique (cf. les réseaux associatifs de [BYB 01]).

Plusieurs arguments empiriques ont été mis en avant en faveur des modèles à exemplaires. En premier lieu, de nombreuses expériences récentes ont fait apparaître que l'auditeur est sensible à la structure détaillée du signal acoustique dans la compréhension de la parole, au moins dans certaines circonstances. Il a ainsi été montré que l'identification des mots était soumise à l'influence d'un certain nombre de paramètres phonétiques fins : variations "subphonémiques" dans la durée du VOT pour les occlusives initiales [AND 94, MCM 02] ; effets de coarticulation régressive entre voyelles [HAW 94, MAR 81], effets de résonance à long terme associés aux liquides [TUN 99, WES 99], indices acoustiques distribués, relatifs au voisement des occlusives en position de coda [HAW 03b] et au lieu d'articulation des coronales en fin de mot [WRI 89], pour ne citer que quelques exemples sur l'anglais. Ces données suggèrent que l'information extraite par l'auditeur à partir du signal de parole dans l'accès au lexique présente un caractère riche et gradué. En deuxième lieu, il a été montré à de multiples reprises que l'auditeur est sensible aux caractéristiques individuelles de la voix du locuteur, en parvenant par exemple à reconnaître plus facilement un mot lorsque ce mot a été prononcé antérieurement par le même locuteur plutôt que par un locuteur différent (voir par ex. [GOL 96]). Enfin, différents travaux font apparaître aujourd'hui que l'enfant prête attention à la forme phonétique détaillée du signal dans l'acquisition de la parole (voir [WER 03] pour une récente synthèse). Dans ce domaine, les modèles à exemplaires se différencient fortement des modèles abstractionnistes en s'attachant à expliquer la mise en place du système phonologique chez

14. En d'autres termes, les prototypes sont recalculés en permanence en fonction des tokens nouvellement intégrés à une catégorie.

l'enfant à partir d'un ensemble de propriétés perceptives et cognitives générales, qui ne sont pas spécifiquement linguistiques.

Malgré le succès qu'ils rencontrent aujourd'hui, les modèles à exemplaires se heurtent à différentes contradictions et laissent plusieurs questions majeures en suspens. Ainsi, les mécanismes permettant aux propriétés acoustiques détaillées d'exercer une influence dans l'accès au lexique restent à élucider. Paradoxalement, les modèles à exemplaires ont souvent été implémentés sous une forme très abstraite, en schématisant le mot d'entrée par une matrice de valeurs binaires par exemple [GOL 98]. En outre, l'hypothèse selon laquelle chaque mot est associé à une liste d'exemplaires qui s'allonge au fil du temps devrait paradoxalement rendre le système insensible à de petits détails dans la forme sonore du mot d'entrée (un peu comme la position du centroïde au sein d'un vaste nuage de points est peu sensible à l'introduction d'un élément supplémentaire). Un autre problème fondamental tient au fait que les modèles à exemplaires font intervenir un ensemble de catégories – le plus souvent les mots – qui sont prédéfinies, et dont les mécanismes qui président à leur possible mise en place semblent échapper au pouvoir explicatif du modèle. Le modèle à exemplaires développé (en production) par Pierrehumbert [PIE 01, PIE 02] par exemple, définit une liste d'exemplaires $E(L)$ comme l'ensemble $\{e_1^L, \dots, e_n^L\}$ associé à un label L dont l'existence est posée *a priori*. Les labels munissent ce modèle et ceux qui lui sont apparentés d'un niveau de représentation symbolique dont la relation avec les exemplaires (formes spatio-temporelles continues) n'est pas explicitée. Une théorie permettant d'expliquer l'émergence de ce niveau de représentation à partir du signal de parole reste à élaborer.

1.4. Bibliographie

- [AND 94] ANDRUSKI J., BLUMSTEIN S., BURTON M., « The effect of subphonetic differences on lexical access », *Cognition*, vol. 52, p. 163–187, 1994.
- [BYB 01] BYBEE J., *Phonology and Language Use*, Cambridge University Press, Cambridge, UK, 2001.
- [CHO 68] CHOMSKY N., HALLE M., *The Sound Pattern of English*, MIT Press, Cambridge, Mass., 1968.
- [CHR 91] CHRISTOPHE A., PALLIER C., BERTONCINI J., MEHLER J., « À la recherche d'une unité : segmentation et traitement de la parole », *L'Année Psychologique*, vol. 91, p. 59–86, 1991.
- [COL 02] COLEMAN J., « Phonetic representations in the mental lexicon », DURAND J., LAKS B., Eds., *Phonetics, Phonology, and Cognition*, Oxford University Press, Oxford, UK, 2002.
- [COL 03] COLEMAN J., « Discovering the acoustic correlates of phonological contrasts », *Journal of Phonetics*, vol. 31, 2003, 351–372.
- [CON 01] CONTENT A., KEARNS R., FRAUENFELDER U., « Boundaries versus onsets in syllabic segmentation », *Journal of Memory and Language*, vol. 45, p. 177–199, 2001.

- [CON 02] CONTENT A., FRAUENFELDER U., « La syllabe comme unité de perception de la parole : un état de la question », *Actes des XXIVèmes Journées d'Etudes sur la Parole*, Nancy, 24–27 juin 2002.
- [CUT 86] CUTLER A., MEHLER J., NORRIS D., SEGUI J., « The syllable's differing role in the segmentation of French and English », *Journal of Memory and Language*, vol. 25, p. 385–400, 1986.
- [CUT 88] CUTLER A., NORRIS D., « The role of strong syllables in segmentation for syllable access », *Journal of Experimental Psychology : Human Perception and Performance*, vol. 14, p. 113–121, 1988.
- [DUP 99] DUPOUX E., KAKEHI K., HIROSE Y., PALLIER C., MEHLER J., « Epenthetic vowels in Japanese : a perceptual illusion ? », *Journal of Experimental Psychology : Human Perception and Performance*, vol. 25, p. 1–11, 1999.
- [DUR 00] DURAND J., « Les traits phonologiques et le débat articulation/audition », BUSUTTIL P., Ed., *Points d'interrogation : Phonétique et phonologie de l'anglais*, Presses Universitaires de Pau, Pau, 2000.
- [ELM 88] ELMAN J., MCCLELLAND J., « Cognitive penetration of the mechanisms of perception : compensation for coarticulation of lexically restored phonemes », *Journal of Memory and Language*, vol. 27, p. 143–165, 1988.
- [ELM 90] ELMAN J., « Finding structure in time », *Cognitive Science*, vol. 14, p. 179–211, 1990.
- [FOW 86] FOWLER C., SMITH M., « Speech perception as 'vector analysis' : an approach to the problems of invariance and segmentation », PERKELL J., KLATT D., Eds., *Invariance and Variability in Speech Processes*, p. 123–136, Lawrence Erlbaum, Hillsdale, NJ, 1986.
- [FRA 96] FRANCIS A., JONES E., « Phonetics and phonological theory », *Language & Communication*, vol. 16, p. 381–393, 1996.
- [FRY 62] FRY D., ABRAMSON A., EIMAS P., LIBERMAN A., « The identification and discrimination of synthetic vowels », *Language and Speech*, vol. 5, p. 171–189, 1962.
- [FUJ 71] FUJISAKI H., KAWASHIMA T., « A model of the mechanism for speech perception : quantitative analysis of categorical effects in discrimination », *Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo*, vol. 30, p. 59–68, 1971.
- [GAS 95] GASKELL M., HARE M., MARSLER-WILSON W., « A connectionist model of phonological representation in speech perception », *Cognitive Science*, vol. 19, p. 407–439, 1995.
- [GAS 03] GASKELL M., « Modelling regressive and progressive effects of assimilation in speech perception », *Journal of Phonetics*, vol. 31, p. 447–463, 2003.
- [GOL 96] GOLDINGER S., « Words and voices : episodic traces in spoken word identification and recognition memory », *Journal of Experimental Psychology : Learning Memory and Cognition*, vol. 22, p. 1166–1183, 1996.
- [GOL 97] GOLDINGER S., « Words and voices – Perception and production in an episodic lexicon », JOHNSON K., MULLENIX J., Eds., *Talker Variability in Speech Processing*,

p. 33-66, Academic Press, 1997.

- [GOL 98] GOLDINGER S., « Echoes of echoes ? An episodic theory of lexical access », *Psychological Review*, vol. 105, p. 251–279, 1998.
- [GOL 03] GOLDINGER S., AZUMA T., « Puzzle-solving science : the quixotic quest for units in speech perception », *Journal of Phonetics*, vol. 31, 2003, 305–320.
- [GRO 03] GROSSBERG S., « Resonant neural dynamics of speech perception », *Journal of Phonetics*, vol. 31, 2003, 423–445.
- [HAR 00] HARRIS J., LINDSEY G., « Vowel patterns in mind and sound », BURTON-ROBERTS N., CARR P., DOCHERTY G., Eds., *Phonological Knowledge : Conceptual and Empirical Issues*, p. 185–205, Oxford University Press, Oxford, UK, 2000.
- [HAW 94] HAWKINS S., SLATER A., « Spread of CV and V-to-V coarticulation in British English : Implications for the intelligibility of synthetic speech », *Proceedings of ICSLP 94*, vol. 1, Yokohama, p. 57–60, 1994.
- [HAW 99] HAWKINS S., « Auditory capabilities and phonological development : animal, baby, and foreign listeners », *The Acoustics of Speech Communication : Fundamentals, Speech Perception Theory, and Speech Technology*, p. 183–197, Allyn and Bacon, Boston, 1999.
- [HAW 03a] HAWKINS S., « Roles and representations of systematic fine phonetic detail in speech understanding », *Journal of Phonetics*, vol. 31, p. 373–405, 2003.
- [HAW 03b] HAWKINS S., NGUYEN N., « Effects on word recognition of syllable-onset cues to syllable-coda voicing », LOCAL J., OGDEN R., TEMPLE R., Eds., *Papers in Laboratory Phonology VI*, p. 38–57, Cambridge University Press, Cambridge, UK, 2003.
- [HIN 86] HINTZMAN D., « “Schema abstraction” in a multiple-trace memory model », *Psychological Review*, vol. 93, p. 411–428, 1986.
- [JAK 52] JAKOBSON R., FANT G., HALLE M., *Preliminaries to Speech Analysis : The Distinctive Features and their Acoustic Correlates*, Massachusetts Institute of Technology, Cambridge, MA, 1952.
- [JOH 97] JOHNSON K., « Speech perception without speaker normalization », JOHNSON K., MULLENIX J., Eds., *Talker Variability in Speech Processing*, p. 145–166, Academic Press, 1997.
- [JOH 05] JOHNSON K., « Speaker normalization in speech perception », PISONI D., REMEZ R., Eds., *Handbook of Speech Perception*, Blackwell, 2005, sous presse.
- [JUS 93] JUSCZYK P., « From general to language-specific capacities : the WRAPSA model of how speech perception develops », *Journal of Phonetics*, vol. 21, p. 3–28, 1993.
- [KLA 79] KLATT D., « Speech perception : a model of acoustic-phonetic analysis and lexical access », *Journal of Phonetics*, vol. 7, p. 279–312, 1979.
- [KUH 91] KUHL P., « Human adults and human infants show a ‘perceptual magnet effect’ for the prototypes of speech categories, monkeys do not », *Perception and Psychophysics*, vol. 50, p. 93–107, 1991.
- [LAD 93] LADEFOGED P., *A Course in Phonetics*, Harcourt Brace Jovanovich, Fort Worth, 3ème édition, 1993.

- [LAH 91] LAHIRI A., MARSLEN-WILSON W., « The mental representation of lexical form : a phonological approach to the recognition lexicon », *Cognition*, vol. 38, p. 245–294, 1991.
- [LAH 02] LAHIRI A., REETZ H., « Underspecified recognition », GUSSENHOVEN C., WARNER N., Eds., *Papers in Laboratory Phonology VII*, p. 637–675, Mouton de Gruyter, Berlin, Germany, 2002.
- [LIB 67] LIBERMAN A., COOPER F., SHANKWEILER D., STUDDERT-KENNEDY M., « Perception of the speech code », *Psychological Review*, vol. 74, p. 431–461, 1967.
- [LIB 85] LIBERMAN A., MATTINGLY I., « The motor theory of speech perception revised », *Cognition*, vol. 21, p. 1–36, 1985.
- [LIB 96] LIBERMAN A., *Speech : A Special Code*, MIT Press, Cambridge, Mass., 1996.
- [LOC 03] LOCAL J., « Variable domains and variable relevance : interpreting phonetic exponents », *Journal of Phonetics*, vol. 31, 2003, 321–339.
- [MAR 81] MARTIN J., BUNNELL H., « Perception of anticipatory coarticulation effects », *Journal of the Acoustical Society of America*, vol. 69, p. 559–567, 1981.
- [MAR 94] MARSLEN-WILSON W., WARREN P., « Levels of perceptual representation and process in lexical access - words, phonemes, and features », *Psychological Review*, vol. 101, p. 653–675, 1994.
- [MAS 83] MASSARO D., COHEN M., « Categorical or continuous speech perception : a new test », *Speech Communication*, vol. 2, p. 15–35, 1983.
- [MCC 86] MCCLELLAND J., ELMAN J., « The TRACE model of speech perception », *Cognitive Psychology*, vol. 18, p. 1–86, 1986.
- [MCM 02] MCMURRAY B., TANENHAUS M., ASLIN R., « Gradient effects of within-category phonetic variation on lexical access », *Cognition*, vol. 86, p. B33-B42, 2002.
- [MCQ 99] MCQUEEN J., NORRIS D., CUTLER A., « Lexical influence in phonetic decision making : evidence from subcategorical mismatches », *Journal of Experimental Psychology : Human Perception and Performance*, vol. 25, p. 1363–1389, 1999.
- [MCQ 01] MCQUEEN J., CUTLER A., Eds., *Language and Cognitive Processes, special issue on Spoken Word Access Processes*, vol. 16, 2001.
- [MEH 81] MEHLER J., DOMMERGUES J., FRAUENFELDER U., SEGUI J., « The syllable's role in speech segmentation », *Journal of Verbal Learning and Verbal Behavior*, vol. 20, p. 298–305, 1981.
- [MIL 94] MILLER J., « On the internal structure of phonetic categories : a progress report », *Cognition*, vol. 50, p. 271–285, 1994.
- [NGU 01] NGUYEN N., « Rôle de la coarticulation dans la reconnaissance des mots », *L'Année Psychologique*, vol. 101, p. 125–154, 2001.
- [NOL 92] NOLAN F., « The descriptive role of segments : evidence from assimilation », DOCHERTY G., LADD D., Eds., *Papers in Laboratory Phonology II : Gesture, Segment, Prosody*, p. 261–280, Cambridge University Press, Cambridge, 1992.

- [NOR 94] NORRIS D., « Shortlist — A connectionist model of continuous speech recognition », *Cognition*, vol. 52, p. 189–234, 1994.
- [NOR 00] NORRIS D., MCQUEEN J., CUTLER A., « Merging information in speech recognition : Feedback is never necessary », *Behavioral and Brain Sciences*, vol. 23, 2000.
- [NOS 86] NOSOFSKY R., « Attention, similarity, and the identification-categorization relationship », *Journal of Experimental Psychology : General*, vol. 115, p. 39–57, 1986.
- [PAL 00] PALLIER C., « Word recognition : do we need phonological representations ? », CUTLER A., MCQUEEN J., ZONDERVAN R., Eds., *Proceedings of the Workshop on Spoken Word Access Processes (SWAP)*, Nijmegen, p. 159–162, 29-31 May 2000 2000.
- [PAL 01] PALLIER C., COLOMÉ A., SEBASTIÁN-GALLÈS N., « The influence of native-language phonology on lexical access : exemplar-based vs. abstract lexical entries », *Psychological Science*, vol. 12, p. 445–449, 2001.
- [PER 86] PERKELL J., KLATT D., Eds., *Invariance and Variability in Speech Processes*, Lawrence Erlbaum, Hillsdale, N.J., 1986.
- [PIE 01] PIERREHUMBERT J., « Exemplar dynamics : Word frequency, lenition, and contrast », BYBEE J., HOPPER P., Eds., *Frequency effects and the emergence of linguistic structure*, p. 137–157, John Benjamins, Amsterdam, 2001.
- [PIE 02] PIERREHUMBERT J., « Word-specific phonetics », GUSSENHOVEN C., WARNER N., Eds., *Papers in Laboratory Phonology VII*, p. 101–140, Mouton de Gruyter, Berlin, Germany, 2002.
- [PIS 75] PISONI D., « The role of auditory short-term memory in vowel perception », *Memory & Cognition*, vol. 3, p. 7–18, 1975.
- [PLA 99] PLAUT D., KELLO C., « The emergence of phonology from the interplay of speech comprehension and production : a distributed connectionist approach », MACWHINNEY B., Ed., *The Emergence of Language*, p. 381–415, Lawrence Erlbaum, Mahwah, NJ, 1999.
- [PRO 99] PROTOPAPAS A., « Connectionist modeling of speech perception », *Psychological Bulletin*, vol. 125, p. 410–436, 1999.
- [REM 94] REMEZ R., RUBIN P., BERNS S., PARDO J., LANG J., « On the perceptual organization of speech », *Psychological Review*, vol. 101, p. 129–156, 1994.
- [REP 84] REPP B., « Categorical perception : issues, methods, findings », LASS N., Ed., *Speech and Language : Advances in Basic Research and Practice*, vol. 10, p. 243–335, Academic Press, Orlando, Flor., 1984.
- [REP 87] REPP B., LIBERMAN A., « Phonetic category boundaries are flexible », HARNAD S., Ed., *Categorical perception : the groundwork of cognition*, p. 89–112, Cambridge University Press, New York, 1987.
- [ROS 76] ROSCH E., MERVIS C., GRAY W., JOHNSON D., BOYES-BRAEM P., « Basic objects in natural categories », *Cognitive Psychology*, vol. 8, p. 382–349, 1976.
- [SAV 70] SAVIN H., BEVER T., « The nonperceptual reality of the phoneme », *Journal of Verbal Learning and Verbal Behavior*, vol. 9, p. 295–302, 1970.

- [SEG 02] SEGUI J., FERRAND L., « The role of the syllable in speech perception and production », DURAND J., LAKS B., Eds., *Phonetics, Phonology, and Cognition*, p. 151–167, Oxford University Press, Oxford, UK, 2002.
- [SHI 93] SHILLCOCK R., LEVY J., LINDSEY G., CAIRNS P., CHATER N., « Connectionist modelling of phonological space », ELLISON T., SCOBIE J., Eds., *Computational Phonology, Edinburgh Working Papers in Cognitive Science*, vol. 8, p. 179–195, 1993.
- [STE 02] STEVENS K., « Toward a model for lexical access based on acoustic landmarks and distinctive features », *Journal of the Acoustical Society of America*, vol. 111, p. 1872–1891, 2002.
- [STR 79] STREETER L., NIGRO G., « The role of medial consonant transitions in word perception », *Journal of the Acoustical Society of America*, vol. 65, p. 1533–1541, 1979.
- [TUN 99] TUNLEY A., Coarticulatory influences of liquids on vowels in English, PhD thesis, Cambridge University, Cambridge, UK, 1999.
- [VAI 05] VAISSIÈRE J., « Intonation », PISONI D., REMEZ R., Eds., *Handbook of Speech Perception*, Blackwell, 2005, sous presse.
- [WAR 70] WARREN R., « Perceptual restoration of missing speech sounds », *Science*, vol. 167, p. 392–393, 1970.
- [WER 03] WERKER J., « The acquisition of language specific phonetic categories in infancy », *Proceedings of the XVth International Congress of Phonetic Sciences*, Barcelone, Espagne, p. 21–25, 3–9 août 2003 2003.
- [WES 99] WEST P., « Perception of distributed coarticulatory properties in English /l/ and /ɫ/ », *Journal of Phonetics*, vol. 27, p. 405–426, 1999.
- [WRI 89] WRIGHT S., KERSWILL P., « Electropalatography in the analysis of connected speech processes », *Clinical Linguistics & Phonetics*, vol. 3, p. 49–57, 1989.