REPRÉSENTATIONS PHONOLOGIQUES EN RECONNAISSANCE DES MOTS PARLÉS

Christophe Pallier

Laboratoire de Sciences Cognitives et Psycholinguistique
EHESS-CNRS, 54 bd Raspail, 75006 Paris.
pallier@lscp.ehess.fr

Résumé

En perception visuelle, un débat important porte sur la question du degré d'abstraction des représentations qui sous-tendent la reconnaissance des objets. Les objets sont-ils mémorisés sous forme de représentations symboliques structurées ou bien plutôt sous la forme d'exemplaires qui conservent de nombreux détails de surface ? Ces deux types d'approches correspondent à des conceptions très divergentes des processus d'appariement de l'entrée perceptive avec les représentations mentales des objets. Dans un cas, on suppose que le cerveau effectue des calculs symboliques complexes, alors que dans l'autre, la reconnaissance résulte essentiellement d'une recherche du plus proche voisin dans un ensemble d'exemplaires représentés de façon détaillée.

Ce débat peut être transposé dans le domaine de la reconnaissance auditive des mots. Alors que certaines théories (Lahiri & Marslen-Wilson, 1991), inspirées par la phonologie, postulent que des représentations phonologiques très abstraites sont calculées par le cerveau, d'autres supposent qu'à chaque mot est associé un ensemble de représentations "de surface", conservant de nombreux détails acoustiques (Klatt, 1979; Pisoni, 1996). Nous présentons plusieurs travaux expérimentaux sur la perception de la parole qui suggèrent que, lors de la reconnaissance des mots parlés, le cerveau calcule effectivement des représentations phonologiques abstraites.
MOT-CLES: perception de la parole, reconnaissance des mots, psycholinguistique. phonologie, représentations.

Abstract

In the field of visual object recognition, there is a debate about the degree of abstractness of the mental representations involved. Are objects stored as structural, symbolic, representations or are they rather memorised as multiple quasi-pictorial views ? In the first case, recognition would involve complex computations, while in the second, it would consists in a search of a nearest neighbour in a large exemplar space. This debate can be translated in the field of auditory word recognition: while some theories assume that abstract phonological representations of the words are elaborated by the perceptual system, others suppose that words are stored as acoustic tokens. We present experimental data supporting the existence of phonological representations for word recognition.
KEYWORDS: speech perception, word recognition, psycholinguistics, phonology, representations.

Introduction

L'identification des mots par le cerveau peut-elle se faire par une voie acoustique (modèle "direct" dans la figure 1), où bien doit-elle nécessairement passer par l'élaboration d'une représentation phonologique (modèle "pré-lexical" dans la figure 1) ? En d'autres termes : sous quel format un mot est-il mémorisé par le cerveau ? Comme de multiples exemplaires acoustiques ou bien sous une forme phonologique ?

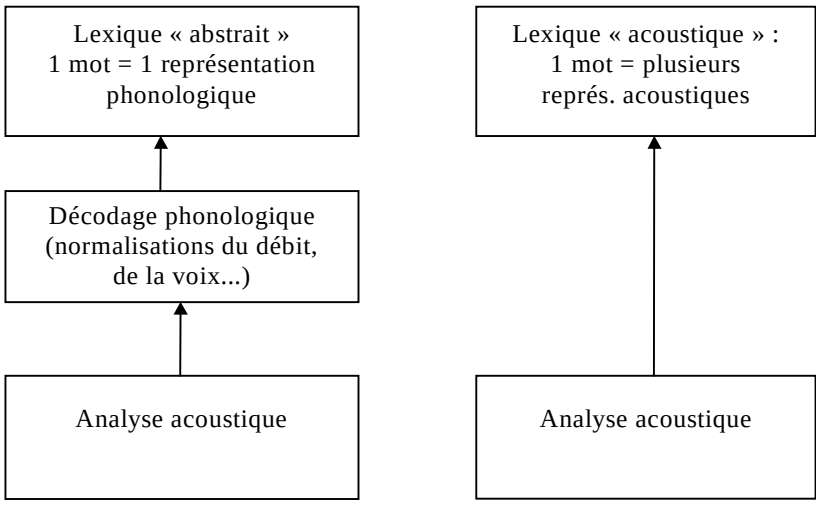

Figure 1: deux modèles possibles de la reconnaissance des mots parlés

Jusqu'à maintenant, pratiquement tous les chercheurs qui étudient la reconnaissance des mots se placent, implicitement, dans le cadre de théories de traitement de l'information qui *postulent* que le signal subit des transformations successives qui élaborent des représentations de plus en plus abstraites. Par conséquent, la plupart des modèles psycholinguistiques de reconnaissance de l'accès au lexique mental (i.e. de l'identification des mots) appartiennent de facto à la catégorie « pré-lexicale » : ils supposent que le signal est transformé dans une suite de phonèmes (ou bien d'autres unités : traits, syllabes, etc.) qui est ensuite comparée avec les représentations stockées dans le lexique (Pisoni & Luce, 1987). Cela est parfaitement illustré par la séparation du travail entre les spécialistes qui se concentrent sur le décodage acoustico-phonétique, et les spécialistes qui étudient comment l'on passe d'une représentation phonémique à une représentation lexicale (Frauenfelder, 1992).

Le modèle acoustique peut paraître, au premier abord, surprenant. Pourtant, considérez la situation dans le domaine de la reconnaissance visuelle des objets. Les premières théories supposaient que chaque objet est stocké en mémoire sous la forme d'une représentation abstraite et structurale (l'objet étant analysé en composants) (Biederman, 1985; Marr & Nishihara, 1978). Cependant ces théories ont été mises à mal par la découverte que la reconnaissance d'un objet tridimensionnel dépend de la façon dont cet objet a été vu précédemment, et n'est donc pas invariante en fonction du point de vue. Elles ont laissé place à l'idée que la reconnaissance d'un objet est en fait le résultat d'une interpolation entre des vues acquises de cet objet (Edelman & Buelthoff, 1992; Ullman, 1996).

Dans les domaines plus spécifiques de la reconnaissance des visages et des mots *écrits* (Jacoby & Brooks, 1984) des modèles « fondés sur des exemplaires » (« instance-based ») ont

également été proposés. Ces modèles supposent que pour chaque visage ou mot, de multiples « traces » (« instances », « exemplaires ») superficielles sont stockées en mémoire ; lorsqu'un stimulus visuel est présenté, il est comparé à toutes ces traces. Ces propositions se fondent empiriquement sur des effets de facilitation du traitement quand un stimulus est répété : ces effets (dit de « medium and long-term repetition priming ») sont très sensibles à des variations de propriétés de surface. [1]

Les mêmes principes peuvent-ils s'appliquer à la perception des mots parlés ? Il est attirant de supposer que des principes similaires sous-tendent la reconnaissance des mots parlés et la reconnaissance des objets en général. Il est donc important de réexaminer les arguments en faveur de l'hypothèse que des calculs symboliques sur des représentations linguistiques se déroulent quand les mots parlés sont reconnus.

L'existence même des représentations phonologiques n'est pas remise en cause ; il existe des arguments assez convaincants en faveur de la réalité psychologique d'objets comme le phonème : erreurs de production, processus phonologiques, acquisition de l'écriture alphabétique (Fromkin, 1973; Halle, 1990). Cependant même si l'on accepte l'existence de représentations phonologiques, cela n'implique pas qu'une telle représentation soit nécessaire pour l'identification des mots : la phonologie pourrait n'exister que pour le versant « production » de la parole. Ainsi, un examen critique des arguments en faveur de la réalité psychologique du phonème montre qu'aucun ne concerne *l'identification* des mots (Klatt, 1981; Pallier, 1994). Si l'on accepte la possibilité que le lexique de perception soit distinct du lexique de production, alors il est logiquement possible que le premier n'utilise pas de représentations phonologiques.

Expérience 1

Selon un modèle acoustique de l'accès au lexique, l'acquisition d'un mot est simplement l'enregistrement des différents exemplaires auxquels on a été exposé ; la représentation d'un mot est donc « véridique », c'est à dire spécifie toutes les propriétés du mot. Dans un modèle à phonologie pré-lexicale, par contre, l'expérience linguistique à pu modifier les calculs pré-lexicaux, filtrant certaines caractéristiques acoustiques avant l'accès au lexique. Les modèles « acoustiques » prédisent, eux, que toute l'information est dans le lexique.

Il existe, à Barcelone, une population de bilingues espagnol-catalan, qui semble très intéressante pour tester ces modèles. Tous ont un très haut degré de bilinguisme car ils ont été exposés intensivement aux deux langues dès le jardin d'enfant ou la première année d'école primaire. Cependant, certains sont nés dans des familles où l'on parle principalement le catalan et d'autres viennent de familles où l'espagnol domine. Le catalan et l'espagnol diffèrent au plan phonologique : seul le catalan possède une distinction entre un /ɛ/ ouvert et un /e/ fermé, alors que l'espagnol n'utilise qu'une seule voyelle /e/, distincte des deux autres.

Dans des tâches de catégorisation et discrimination avec des voyelles synthétisées sur un continuum /e/-/ɛ/, nous avons découvert que les bilingues « nés espagnol » se comportaient différemment des bilingues « nés catalan » (Pallier, Bosch, & Sebastián-Gallés, 1997). Tous percevaient les différences acoustiques entre les stimuli, mais seuls les « catalans » catégorisaient le continuum en deux voyelles distinctes (cf. figure 2).

[1] L'interprétation des effets de répétition pour les mots écrits est l'objet de débats. Les modèles à exemplaires sont en concurrence avec des modèles plus classiques qui supposent l'existence d'un niveau de représentation en « lettres » (voir par ex. Jacoby & Brooks, 1984; Tenpenny, 1995).

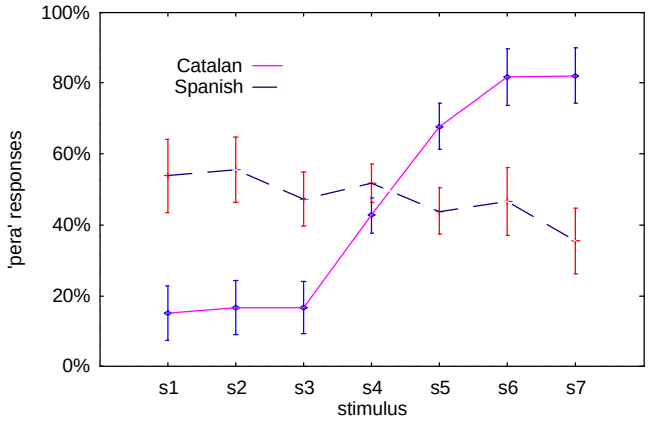

Figure 2: Catégorisation d'un continuum de voyelles [e]-[ɛ] par des sujets de langue maternelle catalane ou espagnole.

Si l'on admet que les tâches de catégorisation et discrimination reflètent des représentations *pré-lexicales*, alors on pourrait conclure que la perception de la parole, et donc la reconnaissance des mots, sont dépendantes de l'expérience linguistique précoce ; les mots ne seraient donc pas enregistrés sous forme acoustique. Cependant un avocat du modèle acoustique pourrait argumenter que ces tâches n'ont rien à voir avec la reconnaissance des mots. Afin de déterminer si les bilingues qui ont appris d'abord l'espagnol ont codé dans leur lexique la différence entre le /e/ et le /ɛ/ catalans, nous avons réalisé l'expérience suivante.

Les participants effectuaient une tâche de décision lexicale dans des listes où apparaissaient des paires minimales de mots utilisant des contrastes catalans pour lesquels les espagnols avaient des difficultés. Quand des stimuli identiques sont répétés dans une liste expérimentale, on observe typiquement un effet de répétition: un stimulus est traité plus rapidement à sa deuxième présentation qu'à sa première. La prédiction du modèle acoustique est que l'effet de répétition doit être identique quelle que soit la langue natale des sujets. Le modèle « prélexical » prédit un effet de répétition entre les membre d'une paire minimale plus important pour les espagnols que pour les catalans.

Les résultats de cette expérience sont indiqués à la figure 3: celle-ci montre les temps de décision lexicale pour des stimuli soit répétés à l'identique ("Same token"), soit formant des paires minimales selon un contraste catalan ("Mini. pair"); Une différence entre "1st" et "2nd" révèle un effet de répétition. On observe une effet de répétition sur les paires minimales seulement pour les sujets espagnols. Cela montre clairement que les représentations des mots en mémoire utilisent un code dépendant de la langue, ou, du moins, que la métrique de comparaison de l'entrée perceptive et des représentations des mots, dépend de la langue dominante.

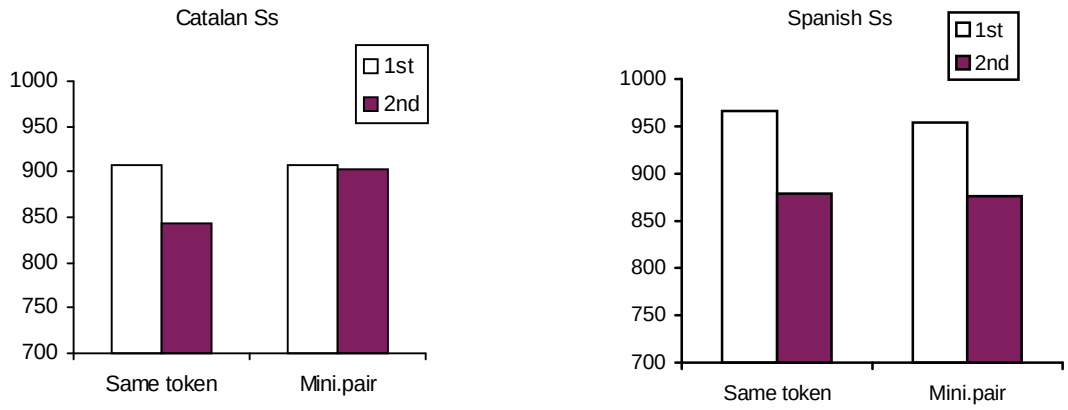

Figure 3: Temps de décision lexicale de sujets catalans et espagnols, sur la première ("1st") et la seconde présentation ("2nd") d'un stimulus répété à l'identique ("Same token") ou avec un changement phonémique catalan ("Mini Pair").

Cette expérience montre que les représentations utilisées pour mémoriser les mots ne sont pas simplement acoustiques : elles dépendent de la langue maternelle des sujets. Ce résultat n'implique pas qu'il s'agisse de représentations symboliques, et permet encore moins d'affirmer que ce sont des représentations phonologiques telles que celles postulées par les linguistes. L'expérience suivante tente d'éclairer cette question.

Expérience 2

Une propriété importante des représentations phonologiques est qu'elles sont structurées hiérarchiquement : dans la plupart de théories phonologiques (e.g. (Goldsmith, 1990)), un mot n'est pas représenté par une simple suite de phonèmes : ceux-ci apparaissent, entre autres, dans une structure syllabique. Par exemple, le mot *caprice* pourra être représenté comme : [k a] [p r i s], et le mot *capture*, comme [k a p] [t u r], les crochets indiquant ici les frontières syllabiques[2]. Dans ces deux mots, /p/ est le troisième phonème, mais il appartient à la première syllabe dans *capture* et à la seconde dans *caprice*.

Le système de perception de la parole est-il sensible à la position des phonèmes dans la structure syllabique? Nous avons réalisé une expérience qui employait un paradigme de détection de phonème avec biais attentionnel (Pallier, Sebastian-Gallés, Felguera, Christophe, & Mehler, 1993). La tâche des sujets était une "détection de phonème" : à chaque essai une lettre représentant un phonème était affichée sur l'écran d'un ordinateur, et était suivie, une seconde plus tard, par un mot prononcé dans des écouteurs. Le participants devaient appuyer aussi vite que possible s'ils entendaient le phonème cible dans le mot.

La manipulation cruciale était que pour un premier groupe de sujets, les phonèmes cibles apparaissaient plus souvent en première syllabe, tandis que pour un second groupe, ils apparaissaient plus souvent en deuxième syllabe. Toutefois, la position phonémique séquentielle des cibles était la même pour les deux groupes. La figure 4 affiche les temps de détection de phonèmes des deux groupes en fonction de la position syllabique de la cible: elle montre clairement que chaque groupe s'est "habitué" à la position syllabique la plus probable. Cela implique que les mots ne sont pas représentés par le système perceptif comme une simple suite de phonèmes, mais que la position syllabique de ceux-ci est indiquée dans la représentation.

[2] Nous ne faisons pas figurer ici les structures sous-syllabiques en attaque et rime.

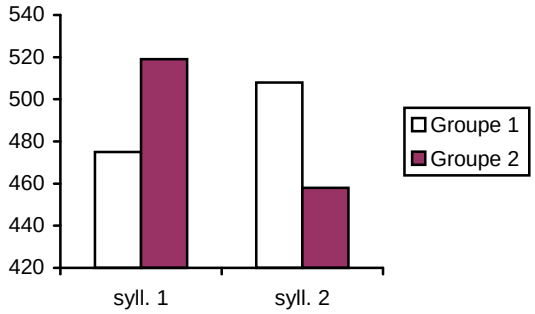

Figure 4: Temps de détection (em msec) d'un phonème en fonction de la syllabe à laquelle il appartient. Groupe 1 = participants "habitués" à des cibles en première syllabe; Groupe 2 = participants "habitués" à des cibles en seconde syllabe.

Il est intéressant de souligner que les sujets n'avaient pas conscience de la "stratégie" qu'ils employaient: interrogés à la fin de l'expérience, ils n'avait pas remarqué que les phonèmes cibles se trouvaient plus souvent dans une syllabe que dans une autre. D'autre part, nous avons répliqué ce résultat en utilisant des stimuli qui étaient des pseudo-mots, plutôt que des mots existants. Cela suggère que les sujets n'utilisent pas une représentation post-lexicale (qui aurait été récupérée *après* la reconnaissance du mot), mais bien une représentation prélexicale, c'est à dire générée *avant* la reconnaissance du mot.

Expérience 3

Les expériences que nous venons de décrire nous ont conduit à l'idée que le système de perception de la parole élaborait une représentation dépendante de la langue qui spécifie la structure syllabique. On peut se demander à quoi peut bien servir la structure syllabique : en effet, elle peut être déduite de la chaîne de phonèmes, et elle est donc, en quelque sorte, une information "redondante". Dans notre thèse (Pallier, 1994), nous avons proposé que la structure syllabique sert de code de correction d'erreur : en imposant que la chaîne de phonèmes perçus soient syllabifiable, le système perceptif dispose ainsi d'une contrainte qui peut lui permettre de résoudre des analyses phonémiques ambiguës ou mal-formées.

En Japonais par exemple, la nombre de structures syllabiques possibles est beaucoup plus restreint qu'en français: la plupart des syllabes japonaises ont une structure CV (consonne-voyelle), ce qui entraîne qu'il n'y a quasiment pas de suites de consonnes dans la parole continue[3]. Si le système perceptif détecte une suite de consonne, il peut "savoir" qu'il y a une erreur. L'expérience suivante (Dupoux et al., in press) montre que ce système "insère" alors une voyelle entre les deux consonnes pour que la représentation qu'il élabore se conforme aux contraintes de la langue japonaise[4].

Dans un premier temps, nous avons créer des pseudo-mots de la forme [VCuCV] (p.ex.*ebuzo*), puis avec un éditeur de sons, nous avons enlevé des portions de plus en plus importantes de la voyelle centrale /u/. Nous avons fait écouter ces stimuli à des locuteurs français et des locuteurs japonais, en leur demandant si les stimuli contenaient un /u/. La

[3] Nous simplifions beaucoup ici; Le lecteur se référera à (Dupoux, Kakehi, Hirose, Pallier, & Mehler, in press) pour plus de détails.

[4] Cela présente aussi l'avantage de permettre l'alignement avec les représentations lexicales pour la phase suivante d'indentification des mots.

figure 5 montrent le taux de réponses 'oui' en fonction de la durée du signal acoustique associé à /u/.

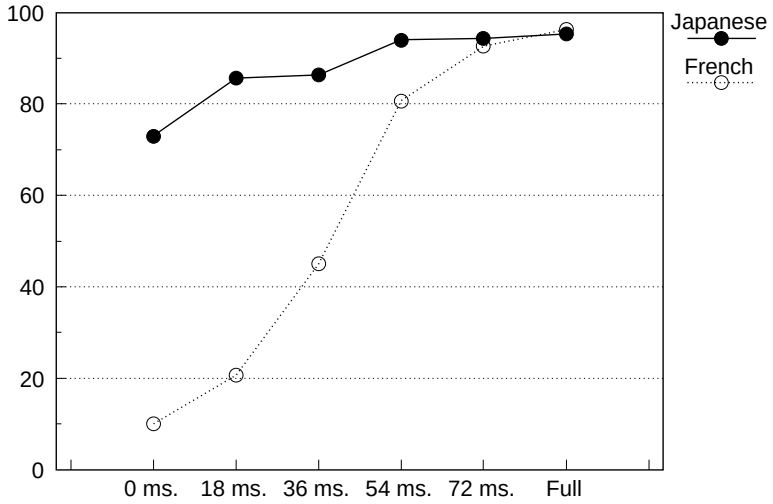

Figure 5: Pourcentage de réponses "/u/-présent" en fonction de la durée de la voyelle /u/ dans des stimuli VCuCV et de la langue des sujets (Japonais et Français).

Dans une seconde expérience, on a présenté des triplets de stimuli aux sujets qui devaient déterminer lequel du premier ou du second était identique au troisième stimulus (tâche dite ABX). Les stimuli pouvait être de la forme VCCV (ex. ebzo), VCuCV (ex. ebuzo), ou VcuuCV (ebuuzo). On constate que les japonais ont plus de difficultés que les français à discriminer des stimuli comme *ebzo* et *ebuzo*. Cela précise le résultat de l'expérience précédente: les japonais percoivent effectivement un son /u/ dans des stimulus comme *ebzo*. Ils ont, par contre, plus de facilité que les français à discriminer des stimuli qui diffèrent par la longueur de la voyelle, comme *ebuzo* et *ebuuzo*. En effet, en Japonais, les suites de voyelles identiques sont admissibles, ce qui n'est pas le cas en Français.

Dans ces expériences, une nette différence de comportement en fonction de la langue apparaît : le système de traitement de la parole des Japonais insère une voyelle "illusoire" à l'intérieur des groupes de consonne illégaux, provoquant chez eux de grandes difficultés à discriminer, par exemple, des stimuli comme *ebzo* et *ebuzo*. Notre interprétation est que le "parseur" phonologique des Japonais "rectifie" l'entrée perceptive pour la rendre conforme à la phonologie de leur langue.

Références

Biederman, I. (1985). Human image understanding: recetn research and a theory. *Computer Vision, GRaphics, and Image Processing, 32*, 29-73.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (in press). Epenthetic vowels in Japanese: a perceptual illusion? *Journal of Experimental Psychology*.

Edelman, S., & Buelthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research, 32*, 2385-2400.

Frauenfelder, U. H. (1992). The interface between acoustic-phonetic and lexical processing. In M. E. H. Schouten (Ed.), *The Auditory Processing of Speech: from Sounds to Words* . New York NY: Mouton de Gruyter.

Fromkin, V. (1973). *Speech errors as linguistic evidence*. The Hague: Monton.

Goldsmith, J. A. (1990). *Autosegmental and metrical phonology*. Cambridge MA: Basil Blackwell.

Halle, M. (1990). Phonology. In D. N. Osherson & H. Lasnik (Eds.), *An Invitation to Cognitive Science: Language. vol.1* (pp. 43-68). Cambridge MA: MIT Press.

Jacoby, L. L., & Brooks, L. R. (1984). Nonanalytic cognition: memory, perception, and concept learning. In G. Bower (Ed.), *The Psychology of Learning and Motivation* (pp. 1-47). New York: Academic Press.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics, 7*, 279-312.

Klatt, D. H. (1981). Lexical Representations for Speech Production and Perception. In T. Myers, J. Laver, & J. Anderson (Eds.), *The Cognitive Representation of Speech* (pp. 11-31). Amsterdam: North-Holland Publishing Company.

Lahiri, A., & Marslen-Wilson, W. D. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition, 38*, 245-294.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society, London, B, 200*, 269-291.

Pallier, C. (1994). *Rôle de la syllabe dans la perception de la parole: études attentionnelles.* , Thèse de doctorat de l'Ecole des hautes études en sciences sociales [available from the author], Paris.

Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition, 64*(3), B9-B17.

Pallier, C., Sebastian-Gallés, N., Felguera, T., Christophe, A., & Mehler, J. (1993). Attentional allocation within the syllabic structure of spoken words. *Journal of Memory and Language, 32*, 373-389.

Pisoni, D. B. (1996). Some thoughts on "Normalization" in Speech Perception. In K. Johnson & J. W. Mullenix (Eds.), *Talker Variability in Speech Processing* . San Diego: Academic Press.

Pisoni, D. B., & Luce, P. A. (1987). Acoustic-phonetic representations in word recognition. *Cognition, 25*, 21-52.

Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychological Bulletin and Review, 2*(3), 339-363.

Ullman, S. (1996). *High-level Vision*. Cambridge, MA: MIT Press.