UNIVERSITE LIBRE DE BRUXELLES
Faculté des Sciences Psychologiques et de l'Education
Unité de Recherche en Neurosciences Cognitives

La perception du voisement en français : investigations comportementales et électrophysiologiques du processus de spécialisation phonologique

Dissertation présentée en vue de l'obtention du grade de Docteur en Sciences Psychologiques,
préparée sous la direction de Madame Monique Radeau et du Docteur Paul Deltenre

par Ingrid Hoonhorst

Mai 2009

Membres du Jury :
Madame Cécile Colin
Docteur Jean-François Démonet
Madame Régine Kolinsky
Madame Jacqueline Leybaert
Monsieur Willy Serniclaes

Remerciements

Un immense merci à tous ceux qui m'ont permis de me lancer dans cette belle aventure.
Merci à Mr Deltenre de m'avoir proposé de me jeter à l'eau, merci pour son soutien constant, sa disponibilité, son enthousiasme scientifique. Merci d'avoir considéré chacune des mes idées…mêmes les plus loufoques !
Merci à Monique Radeau d'avoir accepté de co-diriger cette thèse, merci pour la relecture attentive de mes travaux, pour la justesse et la rigueur de chacune de ses suggestions et surtout merci pour ses nombreux encouragements.
Merci à Willy Serniclaes de m'avoir accompagnée depuis mon mémoire de fin d'études en logopédie. Merci pour ses multiples éclairages, pour nos nombreuses discussions, pour sa disponibilité et pour son engouement communicatif pour la science.
Je ne remercierai jamais assez Cécile Colin. Merci d'avoir traversé avec moi chacune des étapes qui ont mené à la rédaction de cette thèse, merci pour son investissement permanent (et celui de sa famille !), ses relectures, ses commentaires et nos collaborations multiples.

Je tiens à exprimer toute ma reconnaissance aux membres du jury d'avoir accepté de lire et d'évaluer ce travail. Merci aussi aux membres de mon comité d'accompagnement pour leur disponibilité, leurs commentaires et suggestions.

Je ne voudrais pas oublier de remercier tous 'mes' sujets. Pour les expériences Rythme cardiaque, merci aux parents de m'avoir confié leurs enfants de 4 et 8 mois. Pour les expériences comportementales, merci aux enfants de l'école de Capens, à Mme Lacourthiade la directrice de cette école, aux enfants de l'école de Marquefave, à Mme Moreul, leur directrice, merci à Mr Chaussard, inspecteur de l'Académie de Midi-Pyrénées et merci à mon frère Sébastien d'avoir été le relais avec le monde de l'enseignement. Pour les désormais célèbres 'Manips N100', merci à Cécile, Emily, Grégory, Xavier, Hélène, Magali, Alix, Pascale, Laure, Armelle, Léa, Michaël, Florence, Simon, Isabelle et Vincent. Un immense merci aux stagiaires et mémorants qui sont passés par le labo et avec qui j'ai eu beaucoup de plaisir à travailler : merci à Magali, Armelle, Hélène, Grégory, Florence, Simon et Isabelle.

Enfin, merci à tous ceux restés en coulisse. Merci à tous les membres du ‘labo du 10ème’ et en particulier merci à Julie pour ces beaux moments de complicité. Merci au labo de Brugmann et en particulier à Emily pour les parties de badminton et à Mme Beckers pour sa gentillesse et son attention. Merci aux amis toulousains, bruxellois et d’ailleurs. Un merci spécial à Jean-Baptiste pour la relecture de ce travail et à Xavier et Agnès pour le ravitaillement en thé et en chocolat…

Last but not least, merci à Frank et Sébastien et merci à mes parents. Dire qu’il y a 24 ans vous m’accompagniez pour la première fois à l’école… merci pour votre confiance qui traverse les frontières. Et bien sûr merci à Vincent, pour tout mais surtout pour ton soutien, ta curiosité, ton humour et ta bonne humeur.

Table des matières

Géraud de Cordemoy (1668), à propos des enfants :

« Mais toujours il est évident que leur raison est entière dès le commencement, puisqu'ils apprennent parfaitement la langue du païs où ils naissent et même en moins de temps qu'il n'en faudrait à des hommes déjà faits pour apprendre celle d'un païs où ils voyageraient sans y trouver personne qui sût la leur » (p. 26)

Introduction

1. Percevoir

1.1. LA PERCEPTION : UN CONCEPT TRANSDISCIPLINAIRE

En adoptant une perspective cognitiviste pour traiter de la perception de la parole, notre travail se situe au confluent de plusieurs disciplines parmi lesquelles la biologie, la linguistique, la philosophie et la psychologie. Ces diverses approches permettent de rendre compte de la complexité et conséquemment de la richesse du concept de perception.

La question centrale qui réunit phénoménologie, physiologie et psychologie demeure celle de la nature de l'objet perçu : Que nous est-il donné de percevoir ? D'emblée on saisit la dualité inhérente au concept de perception, celle qui oppose une perception objective, réelle, objet d'étude du physiologiste à une perception subjective, représentée, objet de réflexion du philosophe. D'un côté, le physiologiste aborde la perception du point de vue de sa réalité physique : la réception et la transmission d'informations nerveuses provoquées par la présentation d'un stimulus. De l'autre, le phénoménologue s'intéresse à sa réalité vécue, l'expérience consciente de la sensation, sans s'attarder sur son contenu sensoriel. Loin d'être binaires, ces visions se rejoignent dans le concept de *sensation* dont l'étymologie latine 'sensatio' i.e. compréhension, laisse entrevoir le point de convergence entre réalité physique et réalité consciente. La sensation est « le phénomène psychophysiologique par lequel une stimulation externe ou interne a un effet modificateur spécifique sur l'être vivant et conscient » (définition du Nouveau Petit Robert, 2008).

Et c'est en effet avec l'avènement de la psychophysique, discipline dont l'objet d'étude est la relation entre les paramètres du stimulus physique et la sensation concomitante, qu'a été ouverte la voie à une étude multidisciplinaire de la perception. Le philosophe Leibniz sera précurseur en ce domaine en évoquant le moment où « les petites perceptions inconscientes » se regroupent et passent dans le champ du perçu, i.e. le moment où la réalité physique devient réalité consciente. En termes psychophysiques actuels, ce moment correspond au *seuil de détection* d'un stimulus, c'est à dire la limite physique en dessous de laquelle un stimulus n'est plus détecté. A sa suite, le psychologue et philosophe leibnizien Fechner contribua à rationaliser le concept de sensation. Il est à l'origine du concept de *seuil différentiel* ou *seuil de discrimination* qui correspond à la limite en dessous de laquelle deux stimuli ne sont plus

discriminés. Il a par ailleurs établi la loi de Weber-Fechner (1860) qui stipule que la sensation éprouvée par un sujet est proportionnelle au logarithme de la grandeur physique du stimulus. Cette loi fait suite aux travaux de Weber montrant que le seuil différentiel est une fraction constante de l'amplitude du stimulus (relation connue sous le nom de Fraction de Weber). Cette notion de seuil est essentielle en psychophysique.

1.2. THEORIE DE LA DETECTION DU SIGNAL

La notion de seuil a permis de modéliser la prise de décision d'un sujet engagé dans une tâche de détection. Les psychophysiciens ont ensuite repris le modèle mathématique de la théorie de la détection du signal (Tanner & Swets, 1954) qui permet d'analyser la fiabilité d'un processus de décision en présence de bruit. Dans ce cadre théorique, la présentation d'un stimulus cible à une intensité proche du seuil de discrimination équivaut à présenter un stimulus dans un contexte bruité. La figure 1 représente les distributions de la réponse d'un sujet (on parle de *réponse interne*) lorsque la cible est absente (condition bruit seul ; distribution de gauche) et lorsqu'elle est présente (condition signal + bruit ; distribution de droite).

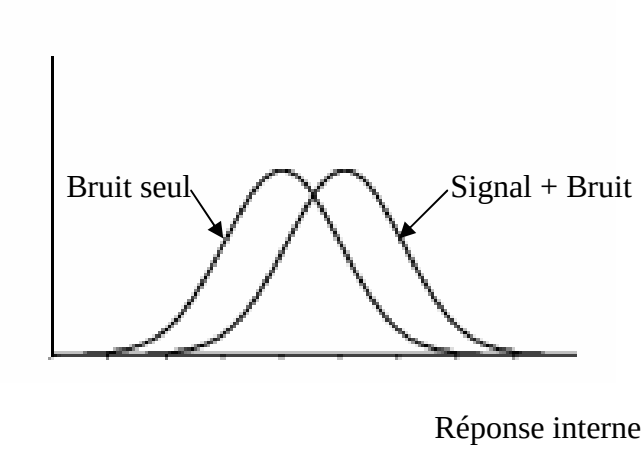

Figure 1 : Distributions gaussiennes de réponse du sujet représentant à gauche, le cas où la cible est absente du stimulus (bruit seul) et à droite, le cas où la cible est présente dans le stimulus (signal + bruit).

La réponse du sujet dépendra non seulement de la présence/absence de la cible dans le stimulus mais aussi du critère de décision du sujet. Ainsi, pour avoir un indice fiable de perception, il faut non seulement prendre en considération les détections correctes (détection de la cible lorsqu'elle est présente) ou leurs compléments (les omissions : pas de détection de

la cible alors qu'elle est présente) mais aussi les réjections correctes (pas de détection de la cible lorsqu'elle est absente) ou leurs compléments (les fausses alarmes : détection de la cible alors qu'elle est absente) (Tableau 1).

		Décision du sujet	
		NON	OUI
Cible présente	NON	Réjections correctes	Fausses Alarmes
	OUI	Omissions	Détections correctes

Tableau 1 : Réponse du sujet en fonction de la présence/absence de la cible dans le signal et du critère de décision du sujet.

Plus le rapport signal/bruit est mauvais, plus les deux distributions sont larges et se recouvrent (figure 2A). Au contraire, moins le signal est bruité, plus les distributions sont étroites et sont séparées (figure 2B). *L'indice d' de détectabilité* correspond au rapport entre le degré de séparation et la largeur de ces distributions. Lorsque les distributions sont séparées et étroites, l'indice d' est élevé : le risque de faire de faire des omissions et des fausses alarmes est réduit. Au contraire et dans le cas extrême où les deux distributions se recouvrent (d' = 0), le nombre de fausses alarmes est équivalent au nombre de détections correctes, i.e. le sujet répond au hasard.

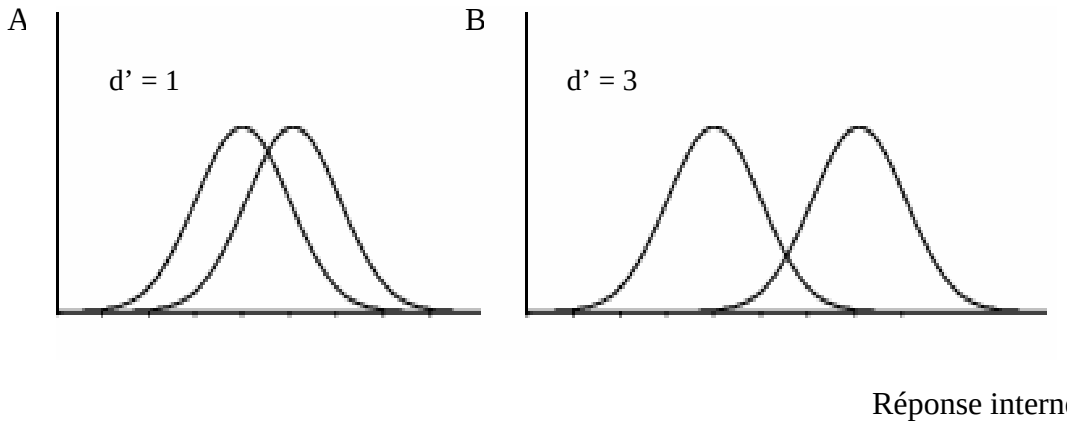

Figure 2A : l'indice d' de détectabilité est faible (d'=1) : les distributions de réponse du sujet se recouvrent largement. **2B** : l'indice d' de détectabilité est élevé (d'=3) : les distributions de réponse du sujet se recouvrent peu.

L'intérêt de cet indice d' est qu'en mesurant le nombre de détections correctes et de fausses alarmes, le critère de décision du sujet (indice β) est lui aussi pris en compte. A discriminabilité constante (d' constant), le nombre de fausses alarmes et le nombre de détections correctes ne sont en effet pas indépendants : l'augmentation du nombre de détections correctes va de pair avec l'augmentation du nombre de fausses alarmes. Ainsi, les performances d'un sujet « libéral » (figure 3A) qui prend le risque de répondre qu'il a perçu la cible sans en être sûr détectera un nombre plus important de cibles qu'un sujet « conservateur » (figure 3B) mais obtiendra aussi un nombre plus important de fausses alarmes. L'indice d' permet donc de comparer les capacités perceptives de sujets utilisant des stratégies de réponse différentes.

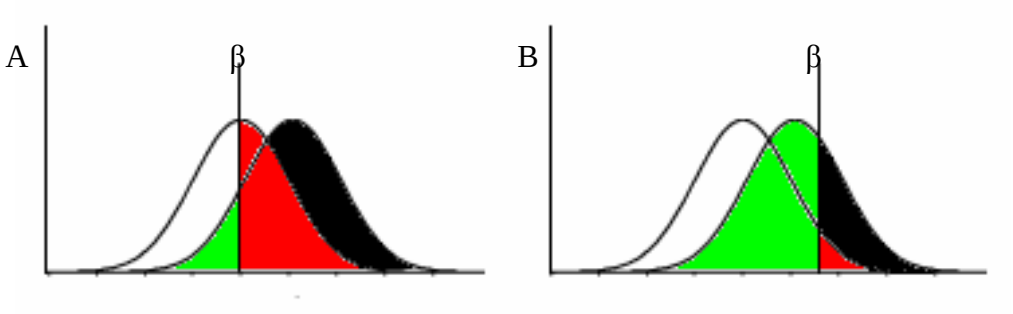

Figure 3A : critère de décision β d'un sujet « libéral » : le sujet détecte souvent la cible, ce qui a pour conséquence d'augmenter le nombre de détections correctes mais aussi le nombre de fausses alarmes. **3B** : critère de décision β d'un sujet « conservateur » : le sujet fait moins de fausses alarmes mais détecte moins souvent la cible qu'un sujet libéral. En blanc : les réjections correctes ; en noir : les détections correctes ; en vert : les omissions ; en rouge : les fausses alarmes.

La théorie de la détection du signal fournit un cadre général pour appréhender les phénomènes perceptifs. Nous verrons plus loin que pour certains chercheurs (e.g. Macmillan, Kaplan, Creelman, 1977) les théories psychoacoustiques peuvent expliquer la perception de n'importe quel signal : du simple clic jusqu'aux signaux complexes de la parole.

2. Percevoir la parole

2.1. LA PAROLE, UN SIGNAL COMPLEXE

La figure 4 représente les spectrogrammes des mots « bar », « par », « dard » et « tard ». Un spectrogramme est une représentation graphique des *indices acoustiques* de la parole. La dimension temporelle est représentée en abscisse, la dimension spectrale en ordonnée et la noirceur du tracé rend compte de la répartition de l'intensité sonore. Les indices acoustiques constituent le premier niveau de l'analyse phonétique : le substrat physique de la parole. Sur la figure 4, on peut observer plusieurs de ces indices :

- Les *formants* (dont les valeurs centrales sont surlignées en jaune) sont des bandes de fréquence qui, par résonance, sont amplifiées lors du passage du son dans les cavités pharyngales (Formant 1 ou F1), buccales (F2) et labiales (F3) (pour une vision plus complète des relations entre formants et cavités, voir Rothenberg, 1981). Les *transitions* de formants désignent les changements de fréquence entre configurations articulatoires successives, comme par exemple celles des consonnes et voyelles. On parle de *locus* du formant pour indiquer le point d'origine de la transition du formant.
- La *barre d'explosion* (en vert) correspond au relâchement de l'air bloqué dans le conduit vocal lors de l'occlusion de la consonne.
- La *barre de voisement* (en rouge) est une bande d'énergie périodique engendrée par la vibration des cordes vocales durant l'occlusion de la consonne.
- L'intervalle de temps entre ces deux derniers événements (barre d'explosion et barre de voisement) correspond au *Délai d'Etablissement du Voisement* (DEV) ou *Voice Onset Time* (VOT). Lisker et Abramson (1964) ont défini cet indice comme le délai entre le relâchement de l'occlusion (signalé par la barre verticale surlignée en vert) et le début de la vibration des cordes vocales (signalé par le début des stries périodiques, erclées en rouge). Le DEV est négatif lorsque la vibration des cordes vocales ccommence avant le relâchement de l'occlusion et positif lorsque elle intervient après.

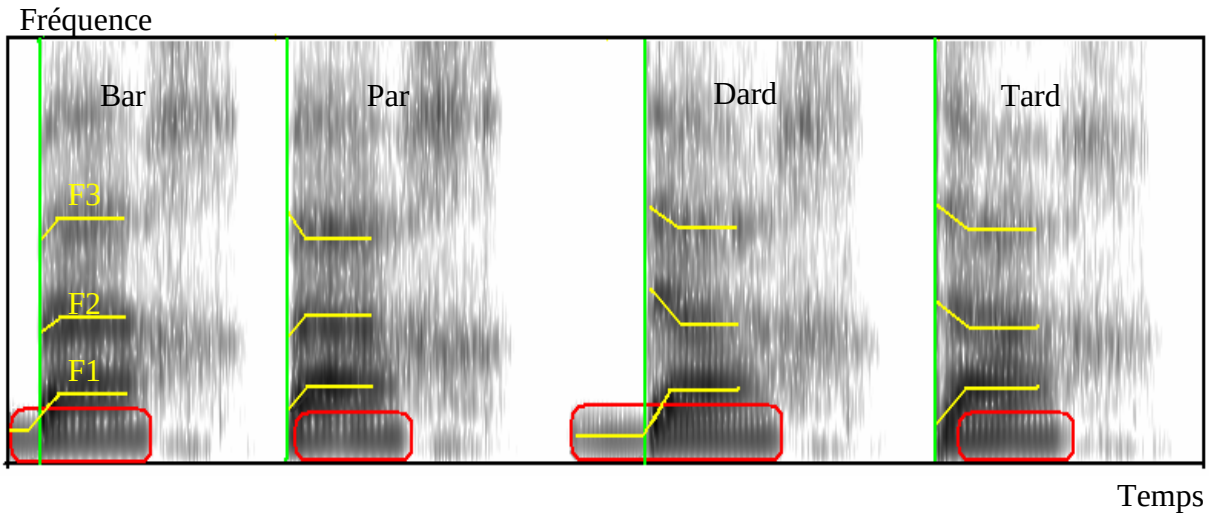

Figure 4 : Spectrogrammes des mots « bar », « par », « dard », « tard » produits par un même sujet. Les barres de voisement sont entourées en rouge ; les barres d'explosion sont représentées par les barres verticales vertes ; les formants sont en jaune.

A partir de cette description du signal de parole, et en tenant compte des indices et des combinaisons d'indices qui permettent de distinguer les mots d'une langue, les linguistes, et notamment les structuralistes, ont postulé l'existence d'unités telles que les traits et les phonèmes (Jakobson, Fant & Halle, 1952). Dans cette approche, le *trait distinctif* correspond à une combinaison d'indices acoustiques et le *phonème* à un ensemble de traits. Jakobson et al. (1952) ont défini un ensemble *universel* de *traits distinctifs binaires* tels que : vocalique vs non vocalique, consonantique vs non consonantique, continu vs discontinu, glottalisé vs non glottalisé, strident vs mat, voisé vs non voisé, nasal vs oral, compact vs diffus, aigu vs grave, bémolisé vs non bémolisé, diésé vs non diésé, tendu vs lâche, complexe vs non complexe. Chaque langue serait caractérisée par un sous-ensemble de ces traits.

Si, dans une approche linguistique, le trait se concrétise lorsqu'il a une valeur distinctive dans une langue, sa réalité psychologique reste une hypothèse (Fromkin, 1979) qui n'a reçu que des confirmations partielles (Morais & Kolinsky, 1994). Le concept de trait est cependant au centre de nombreux travaux sur la perception et la production de la parole qui ont notamment cherché à comprendre comment le sujet isole les traits qui ont une valeur distinctive dans sa langue ou encore à comprendre comment on passe d'une multiplicité d'indices acoustiques à un nombre limité de traits. Pour répondre à ces questions, les psychologues analysent la parole dans ce qu'elle a de plus basique en étudiant par exemple ses corrélats physiologiques

(études 3 et 4 de cette thèse) mais aussi dans ce qu'elle a de plus cognitif en considérant par exemple ses manifestations comportementales (étude 2).

2.1.2. Le trait de voisement

Dans ce travail de thèse, nous nous sommes intéressés à la perception du voisement des consonnes occlusives apicales /d/ et /t/. Le trait de voisement est défini comme la présence de vibrations laryngées durant la production de la consonne (Jakobson et al., 1952). La production du trait de voisement repose dans la majorité des langues sur la relation temporelle entre le début des vibrations laryngées et le relâchement de l'occlusion (explosion), ou *timing laryngé,* dont le corrélat acoustique le plus direct est le DEV (Lisker & Abramson, 1964).
En étudiant les distributions des productions de consonnes occlusives de 11 langues différentes, Lisker et Abramson (1964) ont mis en évidence trois modes différents d'utilisation du timing laryngé : *l'anticipation*, *la simultanéité* et *le retard.* Ces modes correspondent respectivement à la production d'occlusives dont le DEV est négatif (moyenne = -100 ms), positif bref (moyenne = +10 ms) et positif long (moyenne = +75 ms). Le premier mode correspond à l'anticipation de la vibration des cordes vocales sur le relâchement de l'occlusion (figure 5A) ; le second mode, à la simultanéité entre vibration des cordes vocales et relâchement de l'occlusion (figure 5B) ; le troisième mode au retard de la vibration des cordes vocales sur le relâchement de l'occlusion (figure 5C).

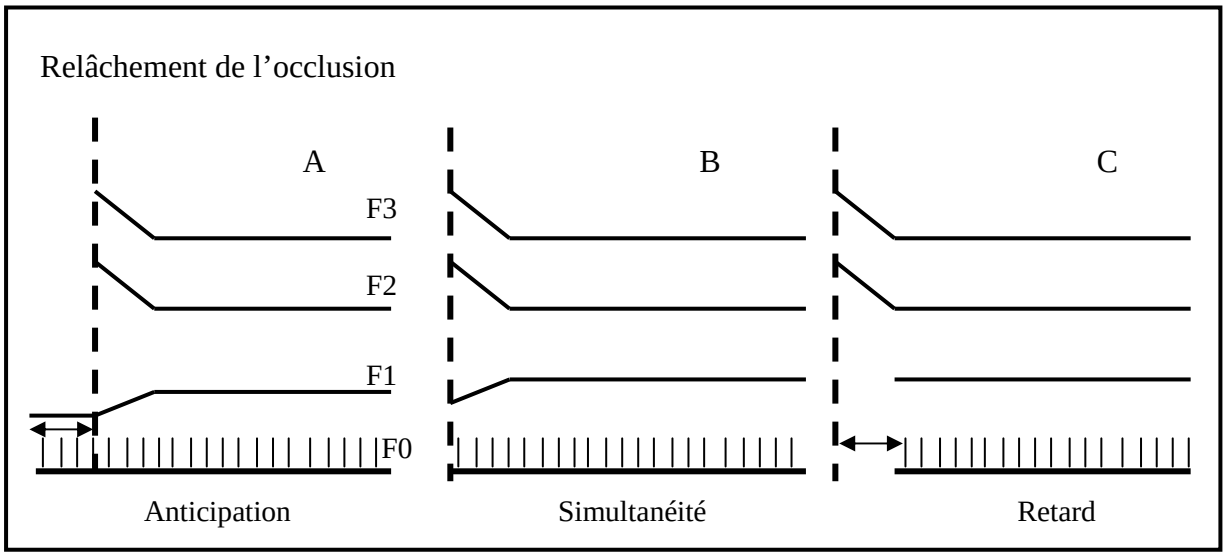

Figure 5: représentation stylisée des 3 modes de timing laryngé mis en évidence par Lisker et Abramson (1964). Dans les cas A et C, le DEV est représenté par une flèche à double sens. Dans le cas B, le DEV est nul. La barre d'explosion correspond à la barre verticale pointillée. La fréquence fondamentale F0 est représentée par les stries verticales.

Lisker et Abramson ont ensuite investigué la correspondance entre les frontières séparant les trois modes de timing laryngé et les frontières perceptives du voisement. Les résultats obtenus tant en identification (tâche dans laquelle les sujets se prononcent sur l'identité du stimulus, Lisker & Abramson, 1967) qu'en discrimination (tâche dans laquelle les sujets déterminent si les stimuli présentés par paire sont identiques ou différents, Abramson & Lisker, 1970) ont validé l'hypothèse d'une correspondance entre catégories de production et de perception. Les trois modes de timing laryngé (anticipation, simultanéité et retard) correspondent en effet à trois catégories perceptives : voisé, non voisé non aspiré et non voisé aspiré. Ces trois catégories sont délimitées par deux frontières universelles[1] : l'une localisée à *-30 ms* et l'autre à *+30 ms* DEV.

Dès l'étude princeps de 1964 sur la production du voisement, Lisker et Abramson mettent en évidence des différences interlinguistiques.

Dans la langue thaï, les 3 catégories perceptives universelles ont une valeur *phonologique*, i.e. les phonèmes voisés, non voisés non aspirés et non voisés aspirés ont une valeur distinctive (Donald, 1978). Dans la majorité des langues du monde cependant, seules deux catégories ont une valeur phonologique. La langue anglaise par exemple ne retient que l'opposition entre phonèmes non voisés non aspirés et non voisés aspirés. Seule la frontière à +30 ms est exploitée dans cette langue (Lisker, Liberman, Erickson & Dechovitz, 1978). La langue anglaise fait cependant quelque peu figure d'exception[2] dans les langues à deux catégories. En français, néerlandais, espagnol, hébreu, arabe, pour ne citer que quelques langues, c'est l'opposition entre phonèmes non voisés et voisés non aspirés qui a une valeur phonologique. La frontière phonologique de voisement, qui dans ces langues se situe à *0 ms DEV*, ne correspond à aucune des deux frontières perceptives universelles (Serniclaes, 1987 ; Maassen, Groenen, Crul, Assman-Hulsmans & Gabreëls, 2001 ; Williams, 1977 ; Laufer, 1998 ; Yeni-Komshian, Caramazza & Preston, 1977). La figure 6 illustre ces différences inter-linguistiques.

1 Le choix du terme 'universel' est motivé par les données de la littérature qui montrent notamment que les animaux sont sensibles aux frontières à -30 et +30 ms DEV (Sinnott & Adams, 1987).

2 En chinois cependant, on retrouve le même type d'opposition phonologique entre phonèmes non voisés aspirés et non aspirés (Liu, Ng, Wan, Wang & Zhang, 2008).

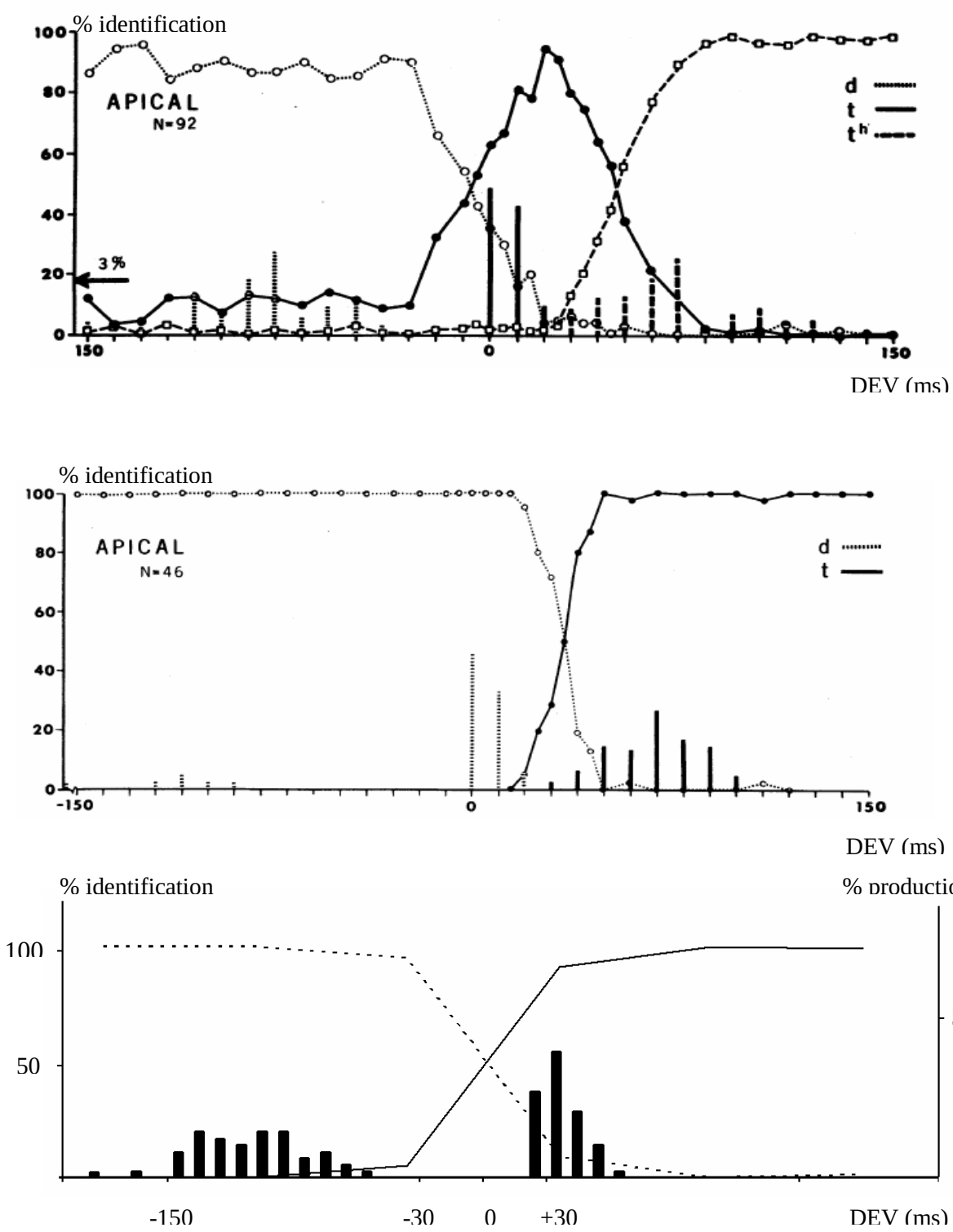

Figure 6 : Fonctions d'identification obtenues sur un continuum de voisement /d-t/ et distributions des productions associées. **En haut** : données pour la langue thaï (Abramson & Lisker, 1968) données ; **au milieu** : pour la langue anglaise (Abramson & Lisker, 1968) ; **en bas** : données pour la langue française (distributions des productions : Serniclaes 1987 ; fonctions d'identification : Hoonhorst et al., 2009).

Si le DEV est l'indice majeur de la perception du voisement, une série d'indices acoustiques secondaires contribuent également à sa perception (Delattre, 1958 ; 1968 ; Wajskop & Sweerts, 1973 ; Lisker, 1978 ; 1985). Pour prendre l'exemple du français, Wajskop et Sweerts (1973) relèvent que cinq indices concourent à la distinction voisé/non voisé pour des occlusives en position intervocalique : la présence ou l'absence de vibrations des cordes vocales pendant l'occlusion, la durée de l'occlusion de la consonne, la longueur et la transition du premier formant de la voyelle dans un contexte [voyelle + consonne], les composantes spectrales et temporelles du relâchement de l'occlusion.

Ces indices contribuent à une perception invariante du trait de voisement dans les différents contextes phonétiques. En français, le DEV varie avec la durée de la transition du premier formant qui elle-même varie avec le lieu d'articulation : plus ce dernier est postérieur, plus le DEV et la transition de F1 s'allongent (Serniclaes, 1987). Or un DEV long signale une consonne *non voisée* tandis qu'une transition de F1 longue signale une consonne *voisée*. L'intégration perceptive du DEV et de la transition de F1 contribue donc à rendre la perception du voisement invariante quel que soit le lieu d'articulation de la consonne.

2. 2. LA PAROLE, UN SIGNAL VARIABLE

Outre sa complexité structurelle, le signal de parole est variable. Citons 3 sources majeures de variabilité :

- *Le contexte phonétique*. Dans la figure 6, les distributions voisée, non voisée non aspirée et non voisée aspirée sont séparées en catégories bien distinctes. Cependant, Lisker et Abramson (1967) ont montré que les distributions de production du voisement pouvaient se superposer dans certaines conditions, rendant la discrimination entre consonnes voisées et non voisées plus difficile.

  Les frontières perceptives se déplacent en effet en fonction du contexte. Par exemple, la frontière perceptive de DEV varie, nous venons de le voir, en fonction du lieu d'articulation. Plus le lieu d'articulation est postérieur, plus la frontière de DEV se déplace vers des valeurs élevées (Lisker et al., 1978). D'autres facteurs tels la nature de l'unité dans laquelle l'indice de DEV est produit (mot isolé ou phrase), la présence ou l'absence d'accentuation au début de la syllabe, la position de la syllabe dans le mot exercent aussi une influence sur la localisation de la frontière perceptive. Spécifiquement, Lisker et Abramson (1970a) ont montré que le DEV des phonèmes /p-t-k/ était plus long à

l'initiale d'une syllabe accentuée d'un mot isolé qu'à l'initiale d'une syllabe non accentuée produite dans une phrase.

De ce fait, les frontières perceptives mentionnées plus haut (-30 et +30 ms pour le thaï, +30 ms pour l'anglais et 0 ms DEV pour le français) ne correspondent pas à des valeurs fixes mais à des *valeurs moyennes* qui fluctuent en fonction de l'environnement phonétique. Pour éviter de tels effets contextuels, nous avons fait le choix dans l'ensemble de nos expériences d'utiliser un contexte neutre (/də/ et /tə/). Dans ce contexte, la frontière perceptive en français est située à une valeur de DEV proche de 0 ms DEV.

- *Le contexte de communication*. Pour ne prendre qu'un exemple, le débit de parole influe sur le degré de superposition des distributions des catégories voisée et non voisée. En thaï, en anglais et en français, la distribution des catégories de voisement caractérisées par des valeurs de DEV négatives ou positives longues (mais pas positives brèves) varie en fonction du débit de parole : plus le débit est élevé, plus les valeurs de DEV sont courtes (Kessinger & Blumstein, 1997).

- *Les locuteurs*. Les caractéristiques physiques du locuteur de même que son état psychologique peuvent faire varier la fréquence fondamentale de vibration des cordes vocales (F0) qui est directement impliquée dans la perception du voisement. En analysant les productions des occlusives /p-b-t-d/ d'enfants de 7 à 11 ans, Whiteside et Marshall (2001) ont notamment relevé un effet du sexe et de l'âge sur les valeurs de DEV en production.

Nous venons de souligner d'une part la complexité du signal de parole qui peut être décomposé en de multiples unités et d'autre part sa variabilité. En partant des notions linguistiques d'*indice* et de *trait* utilisées pour décrire le substrat physique de la parole, nous nous sommes ensuite intéressés à leur réalité psychologique en introduisant les notions de *catégorie de production* et de *frontières perceptives*.

Serniclaes (2005) souligne l'isomorphisme de ces notions. Alors que l'*indice* et la *catégorie* peuvent être définis en termes *absolus*, le *trait* et la *frontière* sont des unités plus abstraites, définies grâce à leur valeur *différentielle* : le trait comme la frontière n'existent que par la valeur d'opposition qu'ils confèrent à l'intérieur d'un système de relations. Serniclaes (2005) cite Jakobson (1973, p.130) en ces termes : « […] il est nécessaire de mettre encore une fois

l'accent sur le fait que tout trait distinctif n'existe que comme terme d'une relation. La définition d'un tel invariant phonologique ne peut se faire en termes absolus : elle ne peut faire référence à une ressemblance métrique mais doit reposer uniquement sur l'équivalence relationnelle ». Il n'existerait donc pas d'invariance absolue (présente dans le signal de parole) mais bien une invariance relative (liée à la valeur distinctive des traits). Les recherches sur la perception catégorielle appuient cette hypothèse.

2.3. LA PERCEPTION CATEGORIELLE

2.3.1. Définitions

2.3.1.1. La perception catégorielle classique

La définition classique de la Perception Catégorielle (PC) implique d'une part que la fonction d'identification soit non-montone et présente une accélération au voisinage de la frontière, et d'autre part que les performances de discrimination puissent être prédites sur base des résultats obtenus en identification. Les premiers chercheurs à parler de PC sont Liberman, Harris, Hoffman et Griffith (1957). Dans leur étude, des auditeurs étaient soumis à une tâche d'identification et de discrimination de stimuli évoluant le long d'un continuum de lieu d'articulation /b-d-g/ créé en modifiant la durée et la direction du deuxième formant. Dans la tâche d'identification, les sujets devaient se prononcer sur l'identité du stimulus (/b/, /d/ ou /g/) tandis que dans la tâche de discrimination, les sujets devaient déterminer si les stimuli présentés par paire étaient identiques ou différents.

De cette première étude et de celles qui suivirent (Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967 ; Studdert-Kennedy, Liberman, Harris & Cooper, 1970), trois grandes conclusions ont été tirées :

(1) la variation *monotone* d'un stimulus physique aboutit à une perception *non monotone*. Plus spécifiquement : à écart acoustique constant, la discrimination de phonèmes situés au sein d'une même catégorie est plus difficile que la discrimination de deux stimuli situés de part et d'autre de la frontière catégorielle. En 1976, Wood insiste sur le caractère non-monotone de la fonction de discrimination et propose le terme d'*effet de frontière phonémique* pour décrire le fait que les scores de discrimination sont plus élevés à proximité de la frontière phonologique qu'à distance de celle-ci.

(2) il est possible de prédire sur base des résultats obtenus en identification ceux obtenus en discrimination. En d'autres termes, le sujet ne serait capable de discriminer que les paires de stimuli qu'il parvient à identifier. La figure 7 (Studdert-Kennedy et al., 1970) illustre cette correspondance parfaite entre frontière d'identification et pic de discrimination.

(3) la parole fait l'objet d'un traitement spécifique réalisé par un décodeur spécialisé. Dans la conclusion de l'article de 1970, Studdert-Kennedy et al. y font explicitement référence : « in our view, some of phonetic perception may be accomplished by a special decoding device available to man as part of his species-specific capacity for language » (p.248).

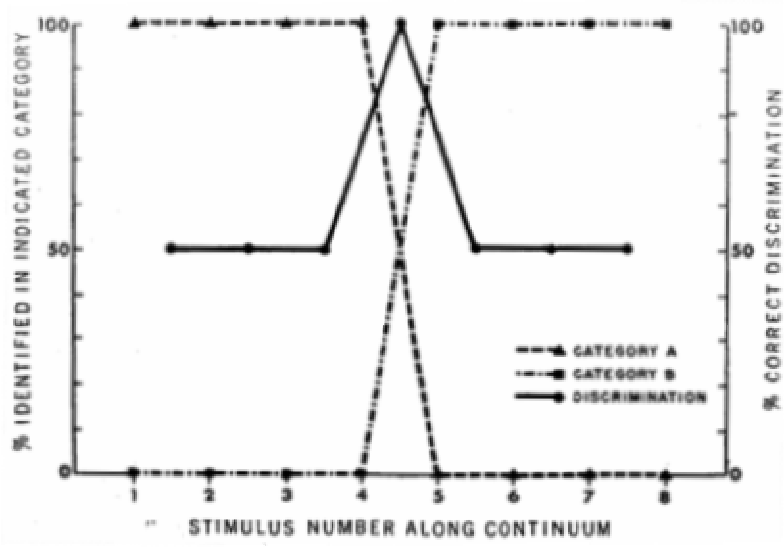

Figure 7: Fonctions d'identification (trait plein) et de discrimination (trait pointillé) « idéales » (Studdert Kennedy et al., 1970). Le pic de discrimination est centré sur la frontière d'identification.

La PC telle que définie dans les conclusions (1) et (2) a été reconnue par la majorité des chercheurs du domaine et de nombreux travaux ont répliqué l'essentiel de ces résultats, tout en mettant en évidence certaines limites (e.g. Schouten, Gerrits & van Hessen, 2003). Cutting et Rosner (1974) ainsi que Snowdon (1987) relèvent à ce propos que la PC est un moyen économique de traiter le flux des informations présentes dans l'environnement. Utiliser un nombre limité de catégories super-ordonnées permettrait d'éviter des traitements cognitifs superflus.

La conclusion (3) a quant à elle fait l'objet de vives critiques. De nombreux chercheurs ont en effet montré que la PC n'était pas limitée aux stimuli linguistiques. En faisant varier de manière régulière le temps de montée du son (rise time) de stimuli issus de 2 continua différents, l'un linguistique et l'autre non linguistique, Cutting et Rosner (1974) ont montré que la perception des sujets testés était non monotone et que la valeur de la frontière perceptive approchait 40 ms quelle que soit la nature du continuum.
Pisoni (1977) a comparé les résultats d'identification et de discrimination de sujets adultes anglophones obtenus en faisant varier un même indice dans deux continua différents. La variation régulière du DEV de stimuli linguistiques ou la variation du délai séparant le début de deux sons purs (de -50 à +50 ms par pas acoustique de 10 ms) de stimuli non linguistiques donnait lieu à des résultats similaires.
Plus largement, les données de la littérature font état d'une PC pour des continua aussi divers que des intervalles musicaux (Burns & Ward, 1978), des couleurs (Bornstein, Kessen & Weiskopf, 1976) et des expressions faciales (Ekman, 1992). Par ailleurs, en montrant que des continua non linguistiques étaient perçus de manière catégorielle par des enfants prélinguistiques, Jusczyk, Pisoni, Walley et Murray (1980) ont donné un argument supplémentaire pour conclure que la PC n'est pas limitée aux sons linguistiques.

D'autres études ont montré que la PC n'est pas l'apanage de l'être humain. Les chinchillas testés dans les expériences de Kuhl et Miller (1975 ; 1978) étaient sensibles aux mêmes variations de DEV que les humains. Après conditionnement avec les stimuli extrêmes d'un continuum /da-ta/ (entre 0 et 80 ms de DEV), la valeur de la frontière d'identification des chinchillas (33.3 ms) ne différait pas significativement de celle d'adultes anglophones (35.2 ms). La similarité entre les modes de perception des animaux humains et non humains allait même au-delà puisque dans leur expérience de 1978, Kuhl et Miller ont montré que l'allongement de la valeur de la frontière perceptive de voisement concomitante de la rétraction du lieu d'articulation (figure 8) ne constituait pas un ajustement perceptif spécifiquement humain. Chez les chinchillas, la valeur de la frontière perceptive obtenue sur un continuum de consonnes bilabiales (/ba-pa/) était plus courte (23.3 ms DEV) que celle obtenue avec des consonnes apicales (/da-ta/ : 33.3 ms DEV) qui elle-même était plus courte que celle obtenue avec des consonnes vélaires (/ga-ka/ : 42.5 ms DEV). Par la suite, on a établi une PC chez la caille japonaise avec un continuum de lieu d'articulation (Kluender,

Diehl & Killeen, 1987) et chez le macaque avec un continuum de DEV (Sinnott & Adams, 1987).

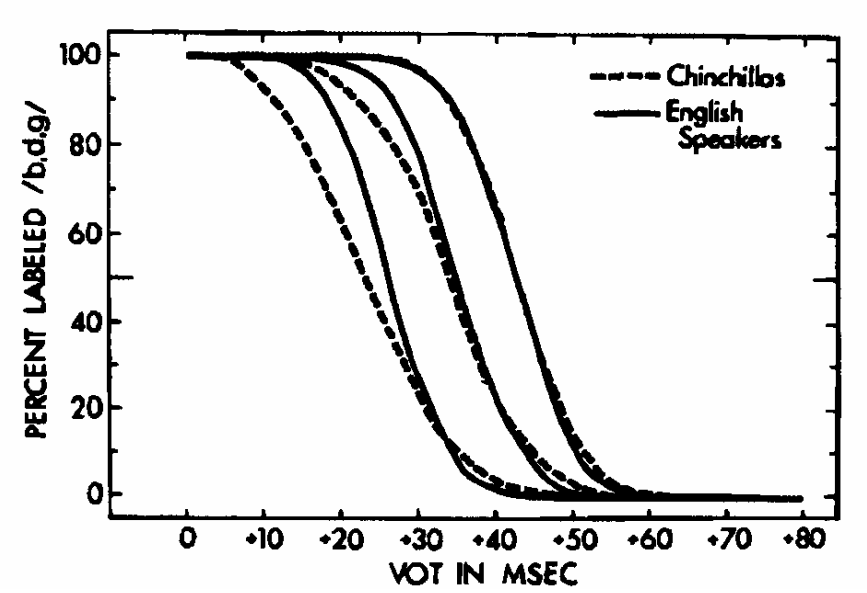

Figure 8: Fonctions d'identifications obtenues par des chinchillas (trait pointillé) et des adultes anglophones (trait plein). A mesure que le lieu d'articulation recule, la frontière d'identification se centre sur des valeurs de DEV plus longues. De gauche à droite : résultats obtenus avec un contraste bilabial /ba-pa/, alvéolaire /da-ta/ et vélaire /ga-ka/.

2.3.1.2. La perception catégorielle revisitée

Le modèle de Haskins (Liberman et al., 1957) tel qu'on vient de le décrire postule que la discrimination d'une paire de stimuli requiert l'identification préalable de chacun des stimuli de la paire. La tâche de discrimination serait donc contrainte par celle d'identification. Cependant, les résultats obtenus en discrimination sont souvent supérieurs à ce qui est prédit sur base de l'identification (Damper & Harnad, 2000). Autrement dit, le sujet humain adulte serait capable de discriminer des paires de stimuli situés dans une même catégorie. Arrêtons nous sur deux facteurs explicatifs de cette capacité à faire des discriminations intracatégorielles : l'*entraînement* préalable aux tâches d'identification et de discrimination et la *nature de la tâche* elle-même.

En présentant des syllabes /ba-pa/ dans différentes tâches de discrimination (tâches même/différent avec et sans entraînement, avec et sans feedback, avec ou sans modification du stimulus standard) à des adultes anglophones, Carney, Widin et Viemeister (1977) ont montré d'une part que l'effet de frontière phonémique peut coexister avec un bon niveau de performance intracatégorielle et d'autre part que l'entraînement a une influence sur les

performances. Récemment les données de Clayards, Aslin, Tanenhaus et Jacobs (2007) ont validé cette dernière suggestion. Dans cette étude, des adultes anglophones étaient soumis à des paires de mots (peach/beach ; peak/beak ; peas/bees) dont le premier phonème variait de -30 à 80 ms DEV. Pendant l'entraînement, les pics des distributions de DEV étaient centrés sur 0 et 50 ms de DEV. Les sujets étaient divisés en deux sous groupes : dans le premier, la largeur des pics des distributions était étroite (8 ms) alors que dans le second elle était large (12 ms). Pendant la tâche d'identification, les sujets devaient choisir entre deux images celle qui correspondait au mot entendu. Les résultats ont mis en évidence un effet de l'entraînement : les pentes d'identification étaient plus raides chez les sujets entraînés avec des pics étroits qu'avec des pics larges. Les performances obtenues en identification étaient donc directement corrélées avec la largeur des distributions présentées pendant l'entraînement.

Concernant la *nature de la tâche*, on peut relever que la durée de l'Intervalle Inter Stimulus (IIS) a une influence sur les performances des sujets en discrimination (Burns & Ward, 1978 ; Werker & Logan, 1985). Dans l'étude de Werker et Logan (1985), les performances obtenues par des sujets adultes anglophones à une tâche de discrimination de contrastes hindi (dental vs rétroflexe) étaient négativement corrélées avec la longueur de l'IIS. Plus l'IIS était long et moins le sujet parvenait à discriminer les paires intracatégorielles. De ces résultats, Werker et Logan ont conclu que l'IIS a une influence sur la nature perceptive du stimulus : *acoustique* avec un IIS de 250 ms, *phonétique* avec un IIS de 500 ms et *phonologique* avec un IIS de 1500 ms.

Cette étude constitue une preuve empirique de l'existence de différents niveaux de représentation de la parole telle que théorisée dans *le modèle à double traitement* (Dual process model) de Fujizaki et Kawashima (1969 ; 1970). Contrairement aux défenseurs du modèle de Haskins, ces auteurs proposent que la discrimination de deux stimuli peut être le résultat d'un traitement *catégoriel* ou *continu*. Si le sujet ne peut pas discriminer les stimuli sur base des informations phonologiques, ce sont les informations acoustiques contenues dans le signal qui lui permettront de juger de la similarité ou de la différence des stimuli[3] (figure 9).

[3] Tout comme Werker et Logan (1985), Fujizaki et Kawashima (1969 ; 1970) postulent qu'il existe différents niveaux de représentation de la parole. Cependant contrairement à Werker et Logan, le modèle à double traitement ne fait état que de deux niveaux : le niveau acoustique et le niveau phonologique.

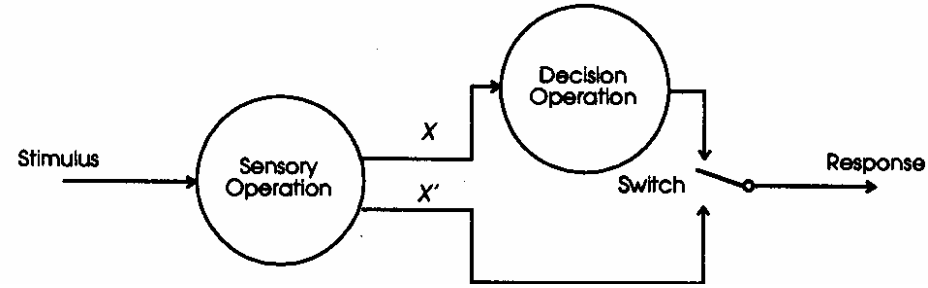

Figure 9 : représentation schématique du modèle à double traitement de Fujizaki et Kawashima (1969 ; 1970). La première voie (X) correspond au traitement catégoriel/phonologique de l'information contrairement au traitement continu/acoustique de la deuxième voie (X').

La discrimination de contrastes intracatégoriels n'est pas limitée aux adultes. Aslin, Pisoni, Hennessy et Perry (1981) ont montré que des enfants âgés de 5 à 11 mois, testés sur un continuum de DEV variant de -70 à +70 ms (par pas acoustique de 10 ms) étaient non seulement sensibles à la frontière phonologique de l'anglais (+30 ms) mais conservaient également une sensibilité autour de l'ancienne frontière universelle située à -30 ms (qui n'a pas de valeur phonologique en anglais). Plus récemment, Rivera-Gaxiola, Silva-Peyrera et Kuhl (2005) ont également montré, sur base de la morphologie de potentiels évoqués auditifs exogènes, que des enfants anglophones de 11 mois testés sur un continuum /da-ta/ discriminaient les contrastes situés de part et d'autre de la frontière phonologique de l'anglais (+30 ms) mais aussi du français (0 ms) alors que cette frontière n'est ni universelle (-30 et +30 ms DEV), ni phonologique dans cette langue. Ainsi, bien que les méthodologies utilisées pour tester les jeunes enfants soient limitées à des mesures de discrimination, ces études mettent en évidence des habiletés de discrimination fines qui ne se limitent pas aux contrastes phonologiques.

Cette absence de relation stricte entre les scores d'identification et de discrimination a conduit à une redéfinition de la PC. Comme de nombreux auteurs, Carney et al. (1977) remettent en cause la définition classique de la PC et limitent celle-ci à la simple correspondance entre frontière d'identification et pic de discrimination : « the important phenomenon is the improved discrimination of stimuli near category boundary » (p. 969). La PC est ainsi réduite à l'*effet de frontière phonémique* c'est à dire à l'augmentation abrupte des performances de discrimination au niveau de la frontière d'identification (Wood, 1976).
Analysons à présent comment la PC a été modélisée.

2.3.2. Modélisations de la PC

2.3.2.1. Modélisation psychoacoustique

Dans une perspective Ockhamienne[4], il serait souhaitable de faire appel à des mécanismes psychoacoustiques généraux pour expliquer le phénomène de PC.
En suivant ce principe, Macmillan (1987) rejette le modèle à double voie de Fujizaki et Kawasjima (1969 ; 1970) à cause notamment de son manque de parcimonie et propose d'expliquer la PC par la ***théorie de la détection du signal***. Pour Macmillan, la réalisation d'une tâche d'identification ou de discrimination peut être ramenée à un processus décisionnel. L'issue de ce processus décisionnel serait influencée par deux facteurs : le *degré d'incertitude* et le *critère de décision* utilisé par le sujet. Le premier facteur est externe puisque directement attribuable au niveau de bruit contenu dans le signal. Par contre le critère de décision est un facteur interne qui dépend de l'expérience du sujet. L'effet de l'expérience sur la décision du sujet a été montré par Eimas et Corbit (1973). Dans leur expérience, Eimas et Corbit montrent que la frontière de voisement varie de 6.1 ms DEV après l'exposition répétée à un même stimulus voisé et de 10 ms DEV après la présentation répétée d'un stimulus non voisé[5]. Pour rendre compte de l'effet de l'expérience sur la décision du sujet, Treisman, Faulkner, Naish et Rosner (1995) ont adapté la théorie de la détection du signal et proposé une nouvelle théorie (***Criterion setting theory)*** qui stipule que le critère de décision du sujet évolue en fonction de la dernière information reçue par le système sensoriel.

Toujours dans une perspective psychoacoustique, Massaro (1987) propose -à l'instar du modèle à deux voies- que la perception nécessite deux étapes, l'une sensorielle l'autre décisionnelle (figure 10).

4 Principe selon lequel « les multiples ne doivent pas être utilisés sans nécessité » (Ockham, XIVe siècle).

5 Cet effet est connu sous le nom d'adaptation. Bien loin d'une explication psychoacoustique, Eimas et Corbit (1973) proposent que la perception du voisement corresponde à l'activation de détecteurs spécialisés dans le traitement du DEV. La stimulation répétée des mêmes détecteurs rehausserait leur seuil d'excitabilité et provoquerait une délocalisation de la frontière perceptive vers des valeurs de DEV traitées par des détecteurs adjacents.

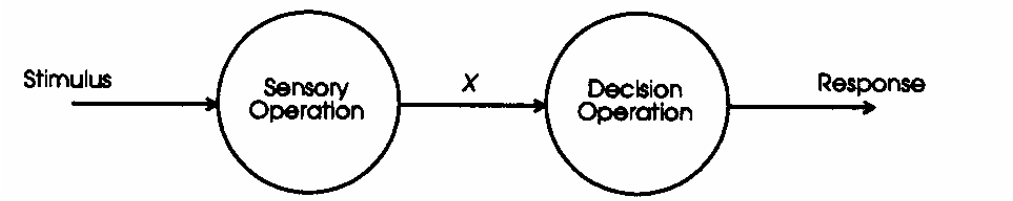

Figure 10 : représentation schématique du modèle de Massaro (1987).

Le modèle de perception à « logique floue » connu sous le nom de FLMP (fuzzy-logical model of perception ; Massaro & Oden, 1980) est un modèle général de la perception applicable à toute modalité sensorielle d'entrée. Le processus de traitement du signal est apparenté à un algorithme comportant trois étapes. Dans un premier temps, le sujet est amené à évaluer de manière indépendante les différentes sources d'informations sensorielles continues. Ces informations permettent au sujet d'extraire des « traits » qui sont affublés d'une valeur logique allant de 0 (radicalement faux) à 1 (absolument vrai). Cette valeur est liée à la présence ou absence de ce trait dans l'information sensorielle. Lors d'une deuxième étape de traitement, les valeurs des traits sont intégrées et comparées aux prototypes mémorisés afin de pouvoir juger de la prototypicité de l'information contenue dans le signal. A l'issue de cette étape, une série de prototypes sont sélectionnés[6]. Lors de la troisième étape de traitement, les valeurs des prototypes sélectionnés sont comparées. Le critère de décision du sujet correspond au choix du prototype qui a la valeur maximale. Massaro insiste sur le fait que cet algorithme de traitement donne lieu à une *division catégorielle* du signal et non pas à une *perception catégorielle* en tant que telle, i.e. le fait que notre vision du monde soit catégorielle n'implique pas que le processus sous-jacent le soit.

Pour d'autres auteurs encore, la simple notion de ***seuil*** permet d'expliquer le passage entre le niveau physique continu et le niveau phonétique discret. Miller, Wier, Pastore, Kelly et Dooling (1976) partent du constat que l'effet de frontière phonémique est une violation de la loi de Weber selon laquelle le seuil différentiel est une fraction constante de l'amplitude du stimulus. Selon cette loi, l'identification d'un stimulus voisé devrait linéairement décroître avec l'augmentation du DEV et la discrimination de paires de stimuli équidistants devrait demeurer constante. Or, la perception de la parole est comme nous l'avons vu non monotone.

[6] Le phénomène d'adaptation pourrait s'expliquer par l'influence de la présentation répétée d'un stimulus sur la trace mnésique du prototype.

Dans leur expérience de 1976, Miller et al. ont comparé les résultats d'identification et de discrimination obtenus en présentant des stimuli simples (bruit **ou** bourdonnement) et complexes (bruit **et** bourdonnement dont le point de départ était décalé de 10 à 80 ms). Alors que la perception était continue avec les stimuli simples, elle était catégorielle avec les stimuli complexes. Selon Miller et al., ces données montrent que la catégorisation phonétique est issue des effets de seuils différentiels masqués : lorsque l'asynchronie temporelle entre les points de départ du bruit et du bourdonnement dépasse le *seuil différentiel*, le bourdonnement est démasqué. Par contre, de part et d'autre de ce seuil de démasquage, la relation entre perception et stimulus obéit à la loi de Weber. Le seuil de démasquage correspond, du point de vue psychoacoustique, à un seuil différentiel et, du point de vue psychologique, à une frontière perceptive.

Appliqué au DEV, on peut penser que les valeurs des frontières universelles de voisement correspondent à ce seuil différentiel. A l'appui de cette hypothèse, Hirsh (1959) a montré avec différents types de stimuli (sons purs ; clics…) que le délai minimal d'asynchronie nécessaire pour juger de l'ordre d'apparition de deux stimuli était de l'ordre de 20 ms à 25 ms.
Il est important ici de noter que dans la littérature les notions de *seuil différentiel* et de *seuil d'ordre temporel* ont été largement confondues. Rappelons que le seuil différentiel correspond à la limite en dessous de laquelle deux stimuli ne sont plus différenciés. Le seuil d'ordre temporel est quant à lui situé à 0 ms et réfère à l'équiprobabilité entre deux ordres possibles (50% de réponses 'anticipation' ou 'retard'). Bien que ce soit le seuil différentiel d'asynchronie que Hirsh ait mesuré à l'aide de la méthode de la Différence Juste Perceptible (DJP ; dans ses résultats, 20 ms correspondait à un écart-type, soit 66% sur une courbe cumulative Normale, 25 ms correspondait à 75%.), de nombreux auteurs (e.g. Stevens & Klatt, 1974 ; Pastore, Ahroon, Buffato, Friedman, Puleo, Fink, 1977 ; Pisoni, 1977 ; Schouten, 1980) ont postulé que les valeurs des frontières universelles correspondaient au délai nécessaire aux sujets pour déterminer *l'ordre* d'apparition des deux indices acoustiques du voisement (le relâchement de l'occlusion et le début de la vibration des cordes vocales).

Le point commun entre ces théories psychoacoustiques (théorie de la détection du signal, Criterion Setting Theory, modèle de perception à « logique floue » ou théorie des seuils) est de ne pas considérer la perception de la parole comme un phénomène *unique* et *catégoriel*. Pastore (1987) souligne l'intérêt de mener des études psychoacoustiques : « An analysis in

terms of difference limens, or in terms of other psychophysical variables that do not assume absolute perceptual discreteness, would allow us to begin to understand more fully the nature of the underlying relationship between perception and the physical stimuli. Such analyses would avoid the implicit attribution of an assumed discreteness and uniqueness to continua that meet the criteria for categorical perception » (p. 48-49). Réduire la PC a un phénomène psychoacoustique permettrait selon lui de mieux comprendre les mécanismes intimes sous-jacents à la catégorisation.

S'il existait une relation biunivoque entre indices acoustiques et traits, cet ensemble d'interprétation suffirait à rendre compte de la perception de la parole. Cependant, l'absence de relation directe entre indices et traits, telle que mise en évidence par l'intégration de divers indices dans la perception d'un même trait, suggère que l'invariance perceptive ne peut pas être limitée à des mécanismes psychoacoustiques. En suivant cette hypothèse, de nombreux théoriciens se sont attachés à chercher ce qui dans le signal de parole pouvait contribuer à l'invariance perceptive.

2.3.2.2. La recherche d'invariants

2.3.2.2.1. Des invariants articulatoires : théorie motrice et théorie de la perception directe réaliste

Pour les partisans de la ***théorie motrice*** de la parole (Liberman et al., 1957 ; Liberman, et al., 1967 ; Liberman & Mattingly, 1985), l'invariant est de nature articulatoire. Malgré les variations acoustiques contextuelles de la parole, l'auditeur aurait accès aux *mouvements articulatoires* que le locuteur met en jeu. La différence acoustique entre deux sons produits par les mêmes articulateurs serait négligée par l'auditeur. Les mouvements articulatoires étant discrets par nature, la PC de la parole (Liberman et al., 1957) constitue une preuve expérimentale à l'appui de la théorie motrice. A l'inverse, les expériences qui ont mis en évidence une PC pour des sons non linguistiques (Cutting & Rosner, 1974) et chez des animaux non humains (Kuhl & Miller, 1975 ; 1978) ont remis en cause les fondements de cette théorie.

Une deuxième version de la théorie motrice a alors été proposée par Liberman et Mattingly (1985). Contrairement à la première version dans laquelle la perception de la parole impliquait l'accès au système moteur, Liberman et Mattingly postulent que ce ne sont pas les commandes motrices qui constituent les invariants articulatoires perçus mais les *gestes*

articulatoires que le locuteur à l'intention de produire. Ce terme de « geste » est emprunté aux travaux de phonologie articulatoire de Browman et Goldstein (1987) et peut être défini comme un ensemble d'actions coordonnées réalisées par les articulateurs du conduit vocal ayant une finalité linguistique (Galantucci, Fowler & Turvey, 2006).

Dans cette même version de la théorie motrice, Liberman et Mattingly (1985) reprennent à leur compte la notion de module (Fodor, 1983) et postulent l'existence d'un *module phonétique* spécialisé dans le décodage des gestes articulatoires. Les travaux de Mann et Liberman (1983) et Whalen et Liberman (1987) sur la *perception duplex* fournissent des preuves expérimentales de l'existence de ce module phonétique. Dans ces travaux, les syllabes /ba/ et /ga/ sont séparées en deux parties : la *base*, constituée des transitions des deux premiers formants et la *modulation de haute fréquence* qui reproduit la transition du troisième formant (modulation montante pour simuler le /ba/ et descendante pour simuler le /ga/). En présentant la base dans une oreille et l'une des deux modulations de haute fréquence dans l'autre oreille, les sujets testés percevaient non seulement les percepts /ba/ et /ga/ dans leur intégralité mais aussi le pépiement associé à la présentation isolée des modulations de haute fréquence. Selon les auteurs, la fusion perceptive de la base et de la modulation relevait du module phonétique tandis que la perception isolée des modulations relevait du système auditif général. Notons cependant que cette interprétation a été remise en cause par Pastore, Schmuckler, Rosenblum et Szczesiul (1983) ainsi que Fowler et Rosenblum (1990) qui ont mis en évidence une perception duplex pour, respectivement, des sons musicaux et le bruit provenant du claquement d'une porte métallique.

La ***théorie de la perception directe réaliste* ou *théorie écologique*** (Fowler, 1986) emprunte l'idée d'invariants articulatoires à la théorie motrice mais s'en détache en arguant que ces invariants sont *directement* perçus sans interface modulaire spécifique et sans inférence cognitive de la part du sujet. La perception directe s'inspire de la théorie de la perception de l'objet de Gibson (1966) en ce qu'elle réfute l'idée que la perception est le résultat d'un processus hiérarchique et sériel de traitement de l'information sensorielle. Pour Gibson comme pour Fowler, le percept n'est pas une *représentation*, il est ce que l'environnement donne à saisir sans intermédiaire cognitif. Dans ce cadre, ce n'est pas le système perceptif qui est le sujet d'étude mais le rapport de l'individu à l'environnement dont il ne peut être soustrait (c'est en ce sens que l'on parle de théorie écologique). Selon Fowler, percevoir la parole passe par la perception directe des gestes articulatoires qui en sont à l'origine (les

«causes distales»). Loin d'être considéré comme l'objet perçu, le signal acoustique n'est que le média par lequel transite l'information. De ce fait, le sujet récepteur ne serait pas sensible aux variations contextuelles acoustiques liées au même geste articulatoire. A ce sujet Fowler (1996) donne l'exemple de la perception du phonème /d/ qui, en fonction du contexte vocalique, varie dans sa structure acoustique. Selon cet auteur, le phonème /d/ est perçu de la même manière dans les occurrences /di/ et /du/ parce que le geste articulatoire (la constriction apicale) est identique dans ces deux syllabes.

2.3.2.2.2. Des invariants acoustiques : les théories auditives

Pour les partisans des théories ***auditives*** (Schouten, 1980 ; Miller et al., 1976), la PC peut être expliquée par des *propriétés générales du système auditif*, ce dernier étant par nature et de manière innée sensible aux propriétés des sons de la parole. Les recherches qui ont montré que la PC n'était ni limitée aux sons linguistiques (Cutting & Rosner, 1974) ni aux humains (e.g. Kuhl & Miller, 1975 ; 1978) ont donné une validité expérimentale à ces théories. De même, les travaux de Eimas, Siqueland, Jusczyk et Vigorito (1971) qui mettent en évidence une PC chez des enfants âgés de 1 à 4 mois laissent à penser qu'elle est basée sur des propriétés auditives générales en place très tôt dans le développement, voire dès la naissance. Bien que séduisantes par leur approche parcimonieuse et écologique, ces théories ne sont pas assez flexibles pour rendre compte de la réalité :

- Comment expliquer que la discrimination intracatégorielle soit possible ?
- Comment expliquer les variations des frontières universelles de voisement avec le lieu d'articulation ?
- Comment expliquer que la frontière phonologique du voisement ne corresponde à aucune des frontières universelles dans certaines langues (i.e. comme en français où la frontière phonologique de voisement est située à 0 ms DEV) ?

2.3.2.2.3. Des invariants relationnels : théorie quantique et théorie du couplage

En faisant un lien entre les dimensions articulatoires et acoustiques, la ***théorie quantique*** de Stevens (1989) dépasse les postulats des théories auditives et propose une invariance relationnelle entre les caractéristiques acoustiques du signal et les mouvements articulatoires. Stevens postule l'existence de zones naturelles de sensibilité et va plus loin en avançant l'idée que ces zones de sensibilité sont basées sur les non-linéarités du passage de l'articulatoire à

l'acoustique. Par non-linéarité, il faut comprendre qu'un même mouvement articulatoire peut avoir des conséquences importantes ou au contraire insignifiantes sur la perception acoustique. L'espace articulatoire serait divisé en états quantiques : des zones stables où la variation d'un mouvement articulatoire n'aurait pas d'incidence sur la perception acoustique (contrastes intracatégoriels) et des zones de transition rapide où la variation d'un mouvement articulatoire aboutirait à une variation abrupte de la perception (contrastes intercatégoriels). Comme réponse à la variabilité du signal acoustique, Stevens propose donc l'invariance des relations entre les mouvements articulatoires du conduit vocal et leurs effets auditifs (Serniclaes, 2000).

L'idée que l'invariance soit relationnelle est reprise par Serniclaes (1987 ; 2005) dans sa ***théorie du couplage*** mais contrairement à Stevens.

La théorie du couplage trouve son origine dans les études qui ont montré que l'intégration d'indices acoustiques est vecteur de stabilité et donc d'invariance. Par exemple, lorsque le DEV s'allonge avec la rétraction du lieu d'articulation, la durée des transitions formantiques augmente aussi (Serniclaes, 2000). De la même manière, la réduction du DEV observée avec l'augmentation du débit de parole est concomitante de la diminution de la durée de la voyelle intrasyllabique (Kessinger & Blusmstein, 1998). Chaque trait serait donc 'codé' par une multiplicité d'indices acoustiques dont les variations sont concordantes (Repp, 1982).

Serniclaes partage avec ces chercheurs l'idée que l'intégration des indices acoustiques et l'ajustement contextuel entre indices participent à l'invariance perceptive, mais il va au-delà. Les résultats d'une étude de Carden, Levitt, Jusczyk et Walley (1981) obtenus avec des stimuli ambigus laissent en effet penser que l'*identité* d'un trait distinctif peut aussi influencer l'identité d'un autre trait. Dans l'étude en question, l'ambiguïté des stimuli venait du mode d'articulation qui pouvait être identifié comme occlusif ou fricatif et ce sans modification du stimulus. Il est apparu que la localisation des frontières de lieu d'articulation variait selon l'information donnée à propos de la nature (occlusive ou fricative) des stimuli. Le fait que la perception d'un trait puisse dépendre de la perception d'un autre trait, évoque les couplages perceptifs ("percept-percept" couplings) mis en évidence pour la vision (Epstein, 1982). De la même manière que la perception de la taille d'un objet dépend de la distance à laquelle cet objet se situe, la *perception* du lieu d'articulation dépend, indépendamment de tout changement acoustique, du mode d'articulation (Carden et al., 1981). Ce qui vaut pour des traits partiellement indépendants comme le mode et le lieu d'articulation devrait également

valoir pour des traits qui sont entièrement corrélés au sein d'une langue, comme par exemple le DEV négatif et le DEV positif.

Appliquée à la perception du voisement, la théorie du couplage (Serniclaes 1987 ; 2000) postule que dans des langues comme le français, l'émergence de la frontière phonologique de voisement située à 0 ms DEV proviendrait d'interactions perceptives entre prédispositions universelles à discriminer le DEV négatif et le DEV positif. Cette théorie part de la constatation que la détection d'asynchronies ne permet pas d'expliquer la perception catégorielle de l'ordre temporel. En effet, pour des stimuli proches du seuil d'ordre temporel (0 ms), un simple processus additif des informations infraliminaires d'anticipation et de retard va engendrer un plateau d'ambiguïté (figure 11). Il est donc exclu que ce processus puisse expliquer la perception du voisement avec une frontière phonologique de 0 ms. Par contre, ce simple processus additif semble être à l'œuvre chez des bilingues franco-anglophones québécois, dont la fonction d'identification se caractérise effectivement par un plateau au voisinage de 0 ms (Caramazza, Yeni-Komshian, Zurif & Carbone, 1973). Pour mettre à l'épreuve cette théorie, Serniclaes (1987) a utilisé soit des stimuli avec un DEV positif soit des stimuli ambigus créés par variation factorielle des DEV négatif et positif (caractérisés par une barre de prévoisement de -120 ms **et** par un retard de voisement d'une durée de 10 à 60 ms). La figure 11 schématise les résultats obtenus en identification. La fonction des francophones ne présente pas de plateau lorsque le DEV est positif. Par contre, en cas de présence conjointe de prévoisement et de retard, leur fonction d'identification atteint un plateau à partir de 20 ms DEV. Ce plateau montre que les informations fournies par le DEV négatif et DEV positif ne sont pas intégrables lorsque les durées de ces deux indices dépassent les seuils d'asynchronie (-30 et +30 ms DEV), ce qui suggère que les auditeurs francophones utilisent effectivement ces mécanismes de seuil pour traiter ces deux indices. A l'appui de cette interprétation, le début du plateau d'identification des francophones correspond à la frontière de DEV des anglophones (20 ms). En cas de conflit supra-liminaire (valeurs de DEV supérieures aux frontières d'asynchronie temporelle), la perception du DEV par les francophones semble donc bien faire appel aux seuils d'asynchronie. Par contre, les conflits infra-liminaires (au voisinage de 0 ms) seraint résolus par des combinaisons entre les processus de détection des deux types d'asynchronie (anticipation et retard), combinaisons qui ont été attribuées à des couplages perceptifs (Serniclaes, 1987, 2000).

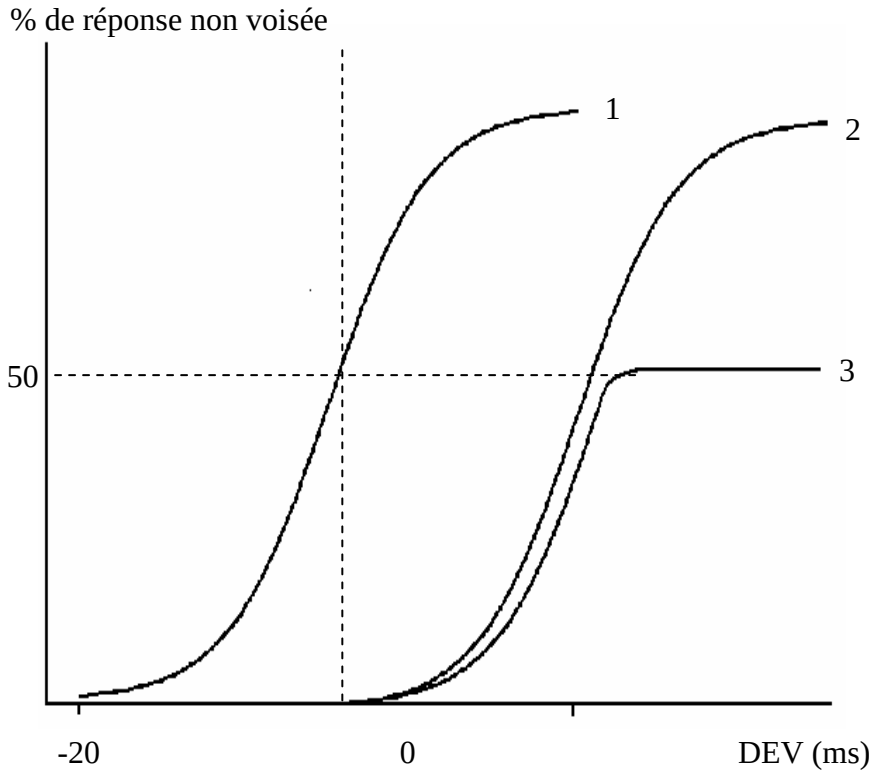

Figure 11 : Fonctions d'identification obtenues par : (1) des sujets adultes francophones sur un continuum de voisement faisant varier le DEV négatif **ou** positif ; (2) des sujets adultes anglophones et (3) francophones sur un continuum de voisement créé par factorisation du DEV négatif **et** positif. Adapté de Caramazza et al. (1973) et Serniclaes (1987).

La figure 12 illustre comment la théorie du couplage modélise le processus de spécialisation phonologique. Dans la partie du haut (A), les frontières universelles divisent l'espace perceptif en 3 catégories (anticipation, simultanéité, retard). L'exposition linguistique va modifier cette organisation : en français, la frontière phonologique émerge par couplage entre frontière universelles (C) tandis qu'en anglais, aucun couplage n'est nécessaire puisque l'une des deux frontières universelles (+30 ms) est conservée (B).

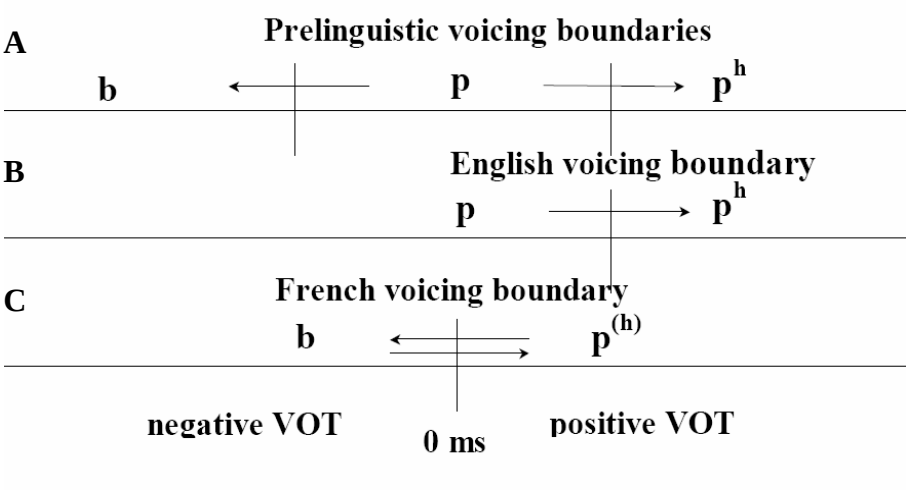

Figure 12 : schématisation du processus de spécialisation phonologique selon Serniclaes (2004). **Partie A :** espace perceptif organisé autour des frontières universelles de voisement localisées à -30 et +30 ms DEV. **Partie B :** réorganisation de l'espace perceptif chez les sujets anglophones de la frontière universelle située à +30 ms DEV. **Partie C :** la frontière phonologique du français émerge à 0 ms DEV par couplage entre prédispositions universelles.

2.3.2.2.4. Des invariants acquis : la théorie de l'apprentissage

D'autres auteurs ont proposé que des habiletés de haut niveau de nature cognitive influencent la perception. On rejoint à cette étape de notre dissertation la pensée de nombreux philosophes. Pour Platon, la perception est une opération intellectuelle que seul l'humain peut exercer puisqu'il est le seul à entretenir un rapport uniquement cognitif avec le monde. Pour Kant, les catégories sont les concepts fondamentaux de la connaissance. Bien que radicale, cette vision purement intellectuelle de la perception amène à considérer l'influence de facteurs cognitifs non linguistiques sur la PC. Arrêtons nous sur l'influence qu'ont l'*apprentissage* et l'*attention* sur la PC.

Lane (1965) propose que les catégories sont le simple produit d'un *apprentissage*. Pour lui, la plasticité de l'espace perceptif est sans limite. La simple exposition intensive à des stimuli peut conduire à une catégorisation. Les expériences menées par Burns et Ward (1978) lui fournissent des preuves expérimentales. Ces auteurs ont en effet montré que sans entraînement, les sujets musiciens percevaient des intervalles musicaux de manière catégorielle contrairement à des sujets non musiciens. De plus, lorsque les sujets musiciens étaient préalablement entraînés avec un continuum dont le pas acoustique était variable, l'effet

de PC disparaissait. La catégorisation ne serait donc pas un phénomène naturel, contraint par des invariants articulatoires ou auditifs mais bien un phénomène *appris et arbitraire*. Tout en souscrivant à cette idée, d'autres auteurs adoptent une position plus nuancée. Miller et al. (1976) par exemple pensent que des frontières perceptives acquises peuvent coexister avec des frontières naturelles. Le caractère plus ou moins catégoriel varierait en fonction de différents facteurs tels que les relations entre stimuli au sein d'une catégorie, la nature du système sensoriel, l'entraînement et les capacités d'attention du sujet.

Sans fournir de preuves empiriques, d'autres auteurs (Aslin & Smith, 1988) proposent que le développement des capacités *attentionnelles* et plus spécifiquement des capacités d'*inhibition de l'attention* permettent au sujet de ne considérer comme pertinentes que les informations acoustiques qui sont contrastives dans sa langue. L'attention jouerait donc un rôle majeur dans le processus de spécialisation phonologique. De manière expérimentale, Lalonde et Werker (1995) montrent que la catégorisation est une habileté qui se développe en synchronie avec d'autres habiletés cognitives non linguistiques. Ces auteurs ont comparé les résultats obtenus par des enfants de 8 à 10 mois à trois tâches : la première de discrimination d'un contraste /ba-da/ natif (anglais : bilabial vs alvéolaire) et d'un contraste non-natif (hindi : rétroflexe vs alvéolaire), la deuxième de catégorisation d'objets et la dernière de recherche d'un objet caché. Les résultats ont montré que les performances obtenues aux tâches de catégorisation phonologique et visuelle étaient corrélées entre elles. Mais au-delà de la simple corrélation entre tâches de catégorisation, les résultats obtenus aux trois tâches étaient corrélés, ce qui amène les auteurs à conclure que le développement des habiletés perceptives est sous-tendu par le développement d'habiletés cognitives générales. D'autres données, obtenues avec des enfants plus âgés, soutiennent l'hypothèse que les habiletés de catégorisation se développent en synchronie avec d'autres habiletés. Par exemple, Gopnik et Meltzoff (1987) ont établi une corrélation entre la capacité à catégoriser des objets et l'explosion lexicale[7]. Plus tard dans le développement, Burnham (2003) a mis en évidence une corrélation entre les résultats de discrimination obtenus par des enfants australiens de 4, 6 et 8 ans sur un contraste /ba-pa/ et leurs résultats en lecture[8].

[7] L'explosion lexicale correspond à l'augmentation significative du nombre de mots produits par un enfant autour de 18 mois.

[8] En Australie, les enfants commencent à apprendre à lire à 5 ans.

Mais, là encore, il semble important de nuancer. Si l'influence de l'attention (et d'autres habiletés cognitives comme l'acquisition de la lecture) sur le développement des habiletés perceptives mérite d'être examinée, le processus de spécialisation phonologique ne peut être limité à ces capacités. Comme souligné par Werker et Curtin (2005), l'attention interagit avec de nombreux autres facteurs tels que les habiletés présentes dès la naissance, le niveau développemental du sujet et la nature de la tâche proposée. Et c'est là le principal écueil rencontré par les théories cognitivistes : comment isoler le facteur cognitif ? Aslin et Smith (1988) relèvent en effet que nombre de variables développementales se confondent avec la maturation d'autres habiletés qui évoluent conjointement avec l'expérience et l'expertise linguistique de l'enfant.

2.3.3. Synthèse

Les travaux des laboratoires Haskins sur la perception catégorielle ont donné lieu à une abondante littérature sur la perception du langage. En toile de fond de toutes ces théories se pose la question de la spécificité du traitement du signal de parole.

Les théories psychoacoustiques et auditives nous invitent à la parcimonie : décomposé en indices acoustiques, le signal de parole n'a rien d'unique. La frontière perceptive ne correspondrait à rien de plus qu'à un seuil différentiel et la perception pourrait être réduite à un algorithme décisionnel. Ce réductionnisme est, nous l'avons vu, imparfait. Demeure cependant l'idée que la perception de la parole se fonde sur *des propriétés auditives générales*, que nous partageons avec les animaux.

La théorie de la perception directe rejette la notion de représentation. Le réel serait directement accessible. Conceptuellement opaque, cette théorie est difficilement testable. Cependant l'idée de prendre en compte *le contexte écologique* est fondamentale : étudier la perception c'est étudier l'interaction du sujet avec son environnement.

La théorie motrice a été l'objet de nombreuses critiques. Les données de PC obtenues avec des sons non linguistiques ont remis en cause ses fondements. Qui plus est, l'idée d'un module spécialisé opérant dès la naissance laisse peu de place aux influences environnementales. Reste que cette théorie souligne le *lien entre perception et production*, une idée à l'origine de nombreux travaux actuellement.

En faisant appel à des habiletés de plus haut niveau, les **théories cognitivistes** pallient les failles des autres théories. Il est néanmoins difficile d'isoler ce qui dans la perception est du ressort de mécanismes bien spécifiques et ce qui est de manière plus générale la conséquence

de la maturation des fonctions cognitives. Les interactions réciproques entre facteurs ne doivent cependant pas interdire leur exploration.

Aucune de ces théories ne rend compte de la complexité de la perception de la parole mais chacune reflète une partie de celle-ci. Aller plus loin dans la compréhension des mécanismes de perception nécessite de changer de perspective. Pour reprendre les mots de Elman, Bates, Johnson, Karmiloff-Smith, Parisi et Plunkett (1996) "studying cells may be useful for understanding life ; but understanding how a cell works will not tell us what it means to be alive" (p.20). Dans ce travail de thèse, nous avons donc choisi d'adopter une *perspective développementale* afin de rendre compte du caractère éminemment *dynamique* de la perception. De toutes les espèces, l'être humain se caractérise par la période de développement la plus longue pendant laquelle le cerveau reste 'plastique', c'est à dire perméable aux interactions avec son environnement. Cette plasticité est source d'adaptation, tant à l'échelle ontogénétique dans la mesure où le cerveau de chaque individu évolue en fonction de la stimulation environnementale, qu'à l'échelle phylogénétique où le développement de chaque individu contribue à la recherche de nouvelles solutions adaptatives bénéficiant à toute l'espèce.

2.4. LA PERCEPTION DE LA PAROLE, UN PROCESSUS DYNAMIQUE

2.4.1. Le potentiel du nourrisson

Partons de la définition des notions de *période sensible* et *période critique* de Knudsen (2004). La période sensible peut être définie comme le laps de temps pendant lequel les effets de l'expérience sur la structuration cérébrale (élaboration et élimination de circuits axonaux et synaptiques) sont majeurs. La période critique correspond à la période pendant laquelle un certain type d'expérience est requis pour permettre à l'individu de se développer normalement ; l'absence de stimulation idoine altère irrémédiablement la suite du développement. Dès la naissance, le cerveau forme un ensemble structuré où les neurones sont interconnectés selon un plan défini. La plasticité du cerveau vient des synapses dont les connexions vont évoluer tout au long de la vie de l'individu. Changeux, Courrège et Danchin (1973) ont montré que l'épigenèse[9] procède par sélection des synapses menant

[9] En biologie, l'épigenèse est une théorie explicative de l'embryogenèse selon laquelle un embryon se développe par différentiation successive de parties nouvelles.

progressivement à la spécialisation fonctionnelle qui caractérise le cerveau adulte. Chez toutes les espèces animales, certaines capacités générales sont observables très tôt : les poussins par exemple lissent leurs plumes, picorent, ouvrent leur bec, suivent un parent, se recroquevillent quand passe un prédateur au-dessus de leur tête (Lorenz, 1970). Ces capacités seraient pré-câblées dans le sens où elles se développeraient sans contact préalable avec l'environnement. Par contre, l'émergence d'habiletés cognitives plus complexes (comme le langage) nécessiterait que l'individu soit soumis à une stimulation spécifique pendant une période dite sensible voire critique. Chez les oiseaux, Immelmann (1969) a montré que, au-delà de l'exposition au chant de l'espèce, c'est le contact avec un tuteur qui constitue la condition nécessaire à l'apprentissage pendant la période critique. Dans son étude, les oisillons nourris par un oiseau d'une autre espèce apprenaient le chant de ce parent nourricier plutôt que le chant de leur espèce. Néanmoins, on peut relever des constantes développementales. Chez l'être humain par exemple, l'âge des premiers mots et des premières phrases varient peu entre individus et ce malgré la diversité des environnements dans lesquels les enfants évoluent. Les effets de l'environnement seraient donc contraints, au moins en partie, par la génétique.

En ce qui concerne le développement de la perception de la parole, DeCasper, Lecanuet, Busnel, Granier-Deferre et Maugeais (1994) ont montré que le bébé humain testé in-utero, préférait la voix de sa mère à celle d'une femme inconnue. De même, les nouveau-nés de 2 jours et de 4 mois issus d'un milieu francophone montrent une préférence pour la prosodie du français plutôt que pour celle du russe (Mehler, Jusczyk, Lambertz, Amiel-Tison & Bertoncini, 1988). Nazzi, Bertoncini et Mehler (1998) ont montré plus récemment que cette capacité à distinguer deux langues était basée sur une différence de rythme entre langues. Toutes ces études font appel à des méthodes d'habituation, sans lesquelles il serait difficile d'étudier le comportement du bébé.

Parmi les méthodes les plus usitées, on trouve *le taux de succion non nutritive*. En 1971, Eimas et al. ont utilisé cette technique pour tester des nourrissons de 1 à 4 mois élevés dans un environnement anglophone. Lors de cette expérience, les enfants étaient soumis à des stimuli voisés et non voisés (/ba-pa/) dont le DEV variait entre -20 et +80 ms par pas acoustique de 20 ms. Après la présentation répétée d'un premier stimulus, l'habituation du bébé se traduisait par une diminution du taux de succion. La variation du taux de succion après la présentation du deuxième stimulus permettait de déterminer si le contraste avait été discriminé. Plus spécifiquement, l'augmentation du taux de succion observée en passant de

+20 à +40 ms DEV a amené Eimas et al. à conclure que les nourrissons étaient capables de manière *innée* de discriminer un contraste acoustique lorsque celui-ci correspondait à une opposition phonologique.

Lasky, Syrdal-Lasky et Klein (1975) et Streeter (1976) ont utilisé une autre technique d'habituation. Dans ces deux études, les auteurs ont évalué les effets de la présentation d'un nouveau stimulus sur le *rythme cardiaque*. Le bébé est d'abord exposé à un premier stimulus jusqu'à la stabilisation du rythme cardiaque avant d'être soumis à un second stimulus. Si le rythme cardiaque diminue, on peut conclure que le contraste a été discriminé. Ces deux études ont été déterminantes pour comprendre la maturation de la perception phonologique. En s'intéressant à la discrimination du contraste /ba-pa/ chez des sujets issus d'un milieu hispanophone, Lasky et al. ont montré que ces enfants, âgés de 4 à 6 mois, étaient capables de discriminer les contrastes -60/-40 ms DEV et +40/+60 ms DEV alors que le contraste phonologique en espagnol -20/+20 ms DEV n'était pas discriminé. On a donc démontré que de jeunes enfants étaient sensibles aux deux frontières universelles de voisement mises en évidence par Lisker et Abramson (1964). Ces résultats ont été répliqués par Streeter (1976) sur des enfants de 2 mois exposés au Kikuyu[10].

Un troisième type de méthodologie a été utilisé par Eilers, Gavin et Wilson (1979). Ces auteurs ont conditionné des bébés à tourner la tête vers le haut parleur lorsqu'ils discriminaient une différence entre un stimulus d'habituation et un nouveau stimulus (*Head Turn Paradigm*). Dans cette étude interlinguistique, Eilers et al. (1979) ont testé l'effet de la langue d'exposition (anglais ou espagnol) sur la discrimination de syllabes /ba-pa/ chez des nourrissons âgés de 6 à 8 mois. Les résultats ont montré que, quelle que soit leur langue d'exposition, tous les bébés étaient sensibles au contraste +10/+40 ms DEV. Par contre, seuls les bébés exposés à la langue espagnole discriminaient le contraste -20/+40 ms.
Ces résultats, comparés à ceux de Lasky et al., peuvent être interprétés comme suit : entre 4 et 6 mois, les nourrissons exposés à l'espagnol ne se sont pas encore spécialisés dans la discrimination des contrastes de leur langue maternelle et sont donc encore sensibles aux frontières universelles de voisement (Lasky et al., 1975). Entre 6 et 8 mois, le processus de

10 Le Kikuyu est une langue bantoue du Kenya dans laquelle la frontière phonologique du voisement est, comme en espagnol, située à 0 ms DEV.

spécialisation phonologique est amorcé et les nourrissons discriminent la paire de stimuli situés de part et d'autre de la frontière phonologique de l'espagnol (Eilers et al., 1979).
Notons cependant que les nourrissons issus d'un milieu hispanophone restent sensibles à la frontière universelle localisée à +30 ms DEV. Dans le cadre de l'hypothèse *'Robuste-Fragile'* (Burnham, 1986), cette sensibilité résiduelle pourrait s'expliquer par la saillance acoustique naturelle de cette frontière. Comme spécifié par Burnham et al. (1991), le masquage de la composante de haute fréquence (relâchement de l'occlusion) par la composante de basse fréquence (vibration des cordes vocales) pourrait expliquer que le DEV négatif soit moins saillant que le DEV positif. Cependant, cette dernière interprétation est sujette à caution. Comment en effet interpréter la discrimination du contraste centré sur +30 ms DEV ? Sans tester la frontière universelle située à -30 ms, il est impossible de savoir si la discrimination du contraste centré sur +30 ms DEV par les anglophones relève d'un processus universel ou d'un processus spécifique à la langue anglaise.

En analysant les réactions d'habituation-déshabituation de nourrissons issus d'environnements linguistiques différents, les études précédemment décrites constituent les premières preuves expérimentales du *processus de spécialisation phonologique*, c'est à dire le passage d'une perception universelle à une perception phonologique, spécifique à la langue du sujet. Les études de Werker et Tees (1984a ; 1984b) marquent un tournant dans la compréhension de ce phénomène. Dans deux expériences, l'une transversale et l'autre longitudinale, Werker et Tees (1984a) présentent à des nourrissons issus d'un milieu anglophone des contrastes de lieu d'articulation du thompson[11] (consonnes glottales non voisées vélaires et uvulaires), de l'hindi (consonnes non aspirées non voisées rétroflexe et dentale) et de l'anglais (consonnes voisées bilabiale et apicale). Quelle que soit la méthodologie utilisée, transversale ou longitudinale, les résultats sont identiques : alors qu'à 6-8 mois, les nourrissons sont capables de discriminer l'ensemble des contrastes, l'habileté à discriminer les contrastes non-natifs thompson et hindi diminue à 8-10 mois jusqu'à devenir quasi nulle à 10-12 mois. La bascule phonologique aurait donc lieu pendant la première année de vie.
Pour expliquer les mécanismes sous jacents à cette spécialisation, Werker et Tees avancent une hypothèse *sélectionniste* : la perception phonologique correspondrait au *maintien* ou à

[11] La langue thompson est une langue indienne parlée en Colombie britannique.

l'*extinction* des frontières universelles. Toutefois, les résultats de discrimination obtenus sur les mêmes contrastes (anglais, thompson, hindi) par des adultes anglophones (Werker & Tees, 1984b) invalident cette dernière conclusion. Bien que les résultats de discrimination de contrastes non-natifs soient moins bons que ceux obtenus avec des contrastes natifs, les adultes anglophones testés étaient capables de faire des discriminations acoustiques fines.

Pour expliquer cette sensibilité résiduelle des sujets adultes, Best (1994) propose dans son modèle d'assimilation perceptive (*Perceptual Assimilation Model),* que la discrimination des contrastes non-natifs est fonction de la capacité d'assimilation de ces contrastes à une opposition catégorielle dans la langue du sujet. Les contrastes non-natifs assimilés à deux catégories phonologiques différentes dans la langue du sujet resteraient discriminables contrairement aux contrastes assimilés à une seule catégorie phonologique. Dans la version évoluée de ce modèle (*Perceptual Assimilation Model Articulatory Organ),* Best et McRoberts (2003) ajoutent l'idée que la spécialisation phonologique est liée à la capacité des nourrissons à détecter les patterns articulatoires de leur langue maternelle. Les contrastes qui font appel à un même articulateur seraient mieux discriminés que ceux produits par des articulateurs différents.

Analysons de plus près, les différents modèles proposés pour rendre compte du processus de spécialisation phonologique.

2.4.2. Modélisation du processus de spécialisation phonologique

2.4.2.1. Cadre général : Aslin et Pisoni

Aslin et Pisoni (1980) décrivent un cadre général dans lequel trois modèles sont proposés pour décrire le processus de spécialisation phonologique (figure 13). La différence majeure entre ces modèles est le degré initial de développement des habiletés de discrimination phonologique : complètement développé, partiellement développé ou sous développé.

- ***Théorie universelle*** *(Universal theory)* : initialement les nouveau-nés seraient capables de discriminer l'ensemble des contrastes phonologiques de toutes les langues. Cette habileté serait maintenue ou perdue après contact avec l'environnement linguistique. Cette théorie reprend donc l'hypothèse sélectionniste de Werker et Tees (1984b).

- ***Théorie de l'harmonisation*** *(Attunement theory) :* à la naissance, les habiletés de discrimination phonologique des nouveau-nés seraient partiellement développées. L'expérience linguistique contribuerait à renforcer, à maintenir ou au contraire à diminuer ces capacités initiales.
- ***Théorie de l'apprentissage perceptif*** *(Perceptual learning theory) :* le nouveau né serait comme pour Skinner (1957) une ardoise vierge. La capacité du nourrisson à discriminer les contrastes phonologiques pertinents dans sa langue serait le résultat d'un apprentissage inductif. Sans contact avec d'autres langues, le répertoire phonologique de l'enfant se limiterait aux phonèmes de sa langue.

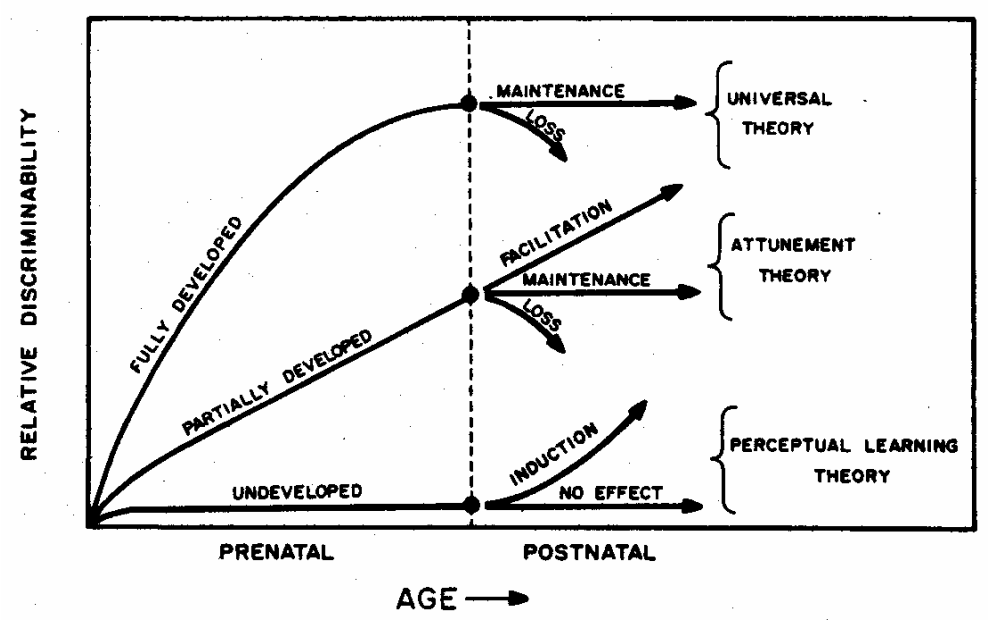

Figure 13 : modèles hypothétiques proposés pour rendre compte du processus de spécialisation phonologique (Aslin & Pisoni, 1980).

Historiquement, les modèles développementaux ont suivi les avancées des modèles de la perception de la parole. La théorie universelle partage avec la théorie motrice l'idée que les capacités de discrimination sont développées dès la naissance et de façon innée. La théorie de l'harmonisation est représentative d'un courant de pensée plus nuancé. Les données expérimentales qui ont montré que la PC n'était ni spécifiquement humaine (Kuhl & Miller, 1975 ; 1978) ni spécifiquement limitée aux sons linguistiques (Cutting & Rosner, 1974 ; Jusczyk et al. 1980 ; Fowler et Rosenblum, 1990) ont amené les chercheurs à limiter l'innéité aux mécanismes auditifs et physiologiques généraux impliqués dans la perception. En ce sens, on peut rapprocher - en partie en tous cas - la théorie de l'harmonisation des théories

auditives. De manière plus évidente, la théorie de l'apprentissage perceptif est apparentée aux théories cognitivistes. Dans cette conception, les capacités de discrimination phonologique se développent en synchronie avec d'autres habiletés cognitives.

Actuellement, la théorie de l'harmonisation est la plus reconnue. Les capacités de discrimination phonologiques seraient grossièrement définies à la naissance et s'affineraient par contact avec la langue environnementale. Dans ce cadre, le processus de spécialisation phonologique relèverait de l'interaction entre un input environnemental et un algorithme d'apprentissage dont l'être humain disposerait à la naissance.

Le modèle PRIMIR (Processing Rich Information from Multidimensional Interactive Representations) élaboré par Werker et Curtin (2005) et le modèle NLM-e (Native Language Magnet model-expanded) dont la dernière version est décrite par Kuhl, Conboy, Coffey-Corina, Padden, Rivera-Gaxiola et Nelson (2008) constituent deux interprétations de la théorie de l'harmonisation.

2.4.2.2. Le modèle PRIMIR

L'objectif du modèle PRIMIR (Werker & Curtin, 2005) est de décrire les processus dynamiques qui sous-tendent l'émergence des représentations lexicales et infra-lexicales. PRIMIR est fondé sur deux grands postulats :

- L'information contenue dans le signal acoustique est *filtrée*. Ces filtres sont au nombre de 3 et correspondent (1) aux contraintes imposées par les mécanismes auditifs, (2) au degré de maturité du sujet et (3) à la nature de la tâche utilisée pour tester les compétences du sujet. Le degré d'activation de chacun de ces filtres fluctue au cours du développement : le filtre (1) est plus actif pendant la première année de vie, le filtre (2) a une plus grande importance chez les enfants tandis que le filtre (3) explique que certains contrastes intracatégoriels soient perçus par des sujets adultes.

- L'information contenue dans le signal acoustique peut faire l'objet de *3 traitements* : (1) Un traitement *général* (General Perceptual plane) qui permet notamment au nourrisson, sur base des régularités statistiques de la langue, de découper le signal de parole en mots. (2) Le deuxième niveau de traitement, de nature *lexicale* (Word Form plane) découle du premier. Le sujet traite directement les mots contenus dans le signal de parole. (3) Le dernier niveau de

traitement est *phonologique* (Phoneme plane). L'une des particularités du modèle PRIMIR est de postuler que la structure phonologique du signal de parole n'est accessible au sujet qu'à partir d'un certain degré d'expertise lexicale.

Bien que ces 3 niveaux de traitement soient liés à des périodes dévelopementales successives, le sujet adulte aurait accès aux trois niveaux et privilégierait un niveau de traitement à un autre en fonction du niveau d'activation des filtres décrits plus haut. La figure 14 schématise le modèle.

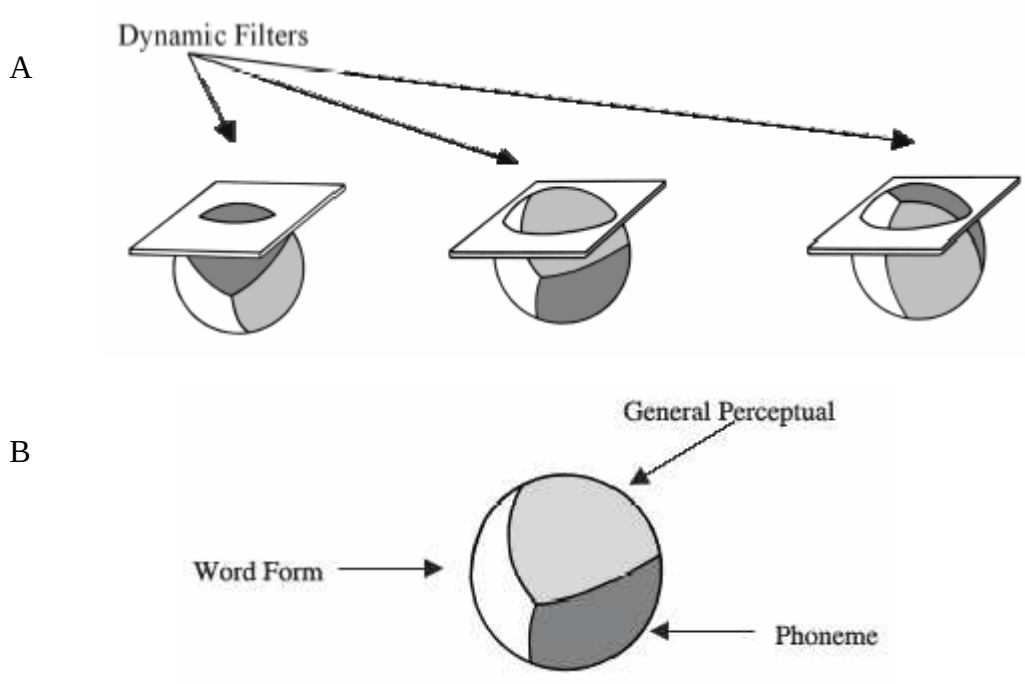

Figure 14: Représentation du modèle PRIMIR de Werker et Curtin (2005). Partie A : les 3 filtres. Partie B : les 3 niveaux de traitement.

2.4.2.3. Le modèle NLM-e

Tout comme PRIMIR, le NLM-e s'attache à modéliser, au-delà du processus de spécialisation phonologique, le développement des habiletés linguistiques en général. La figure 15 schématise les principales phases de développement.

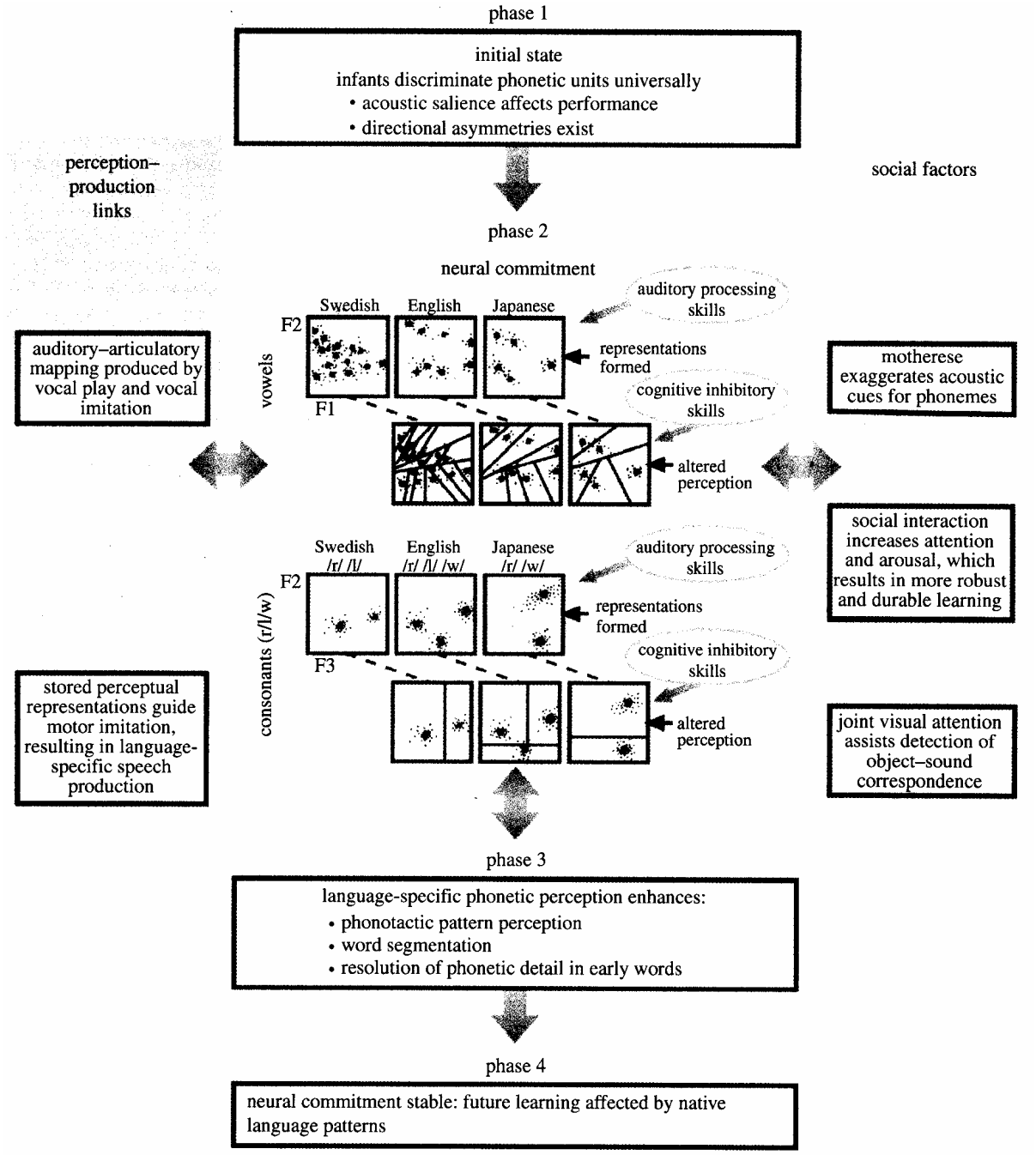

Figure 15 : Représentation du modèle NLM-e de Kuhl et al. (2008) en quatre phases.

La phase 1 correspond à l'état initial. A la naissance, la perception de l'enfant est universelle puisque contrainte par des mécanismes auditifs généraux communs aux animaux humains et

non humains. Dès ce niveau de perception, on peut relever certains effets de saillance acoustique et d'asymétrie directionnelle[12] explicables en termes psychoacoustiques.

La phase 2 constitue le cœur du modèle. Kuhl et al. postulent que deux facteurs sont à l'origine du processus de spécialisation phonologique.

- L'extraction des *régularités statistiques* de la langue conduit à une réorganisation de l'espace perceptif (« warping ») autour des prototypes qui correspondent aux valeurs moyennes de production de ces phonèmes. La rapidité de la réorganisation de l'espace perceptif autour d'un phonème est fonction de la fréquence d'occurrence et des propriétés acoustiques des phonèmes présents dans l'environnement du sujet (Anderson, Morgan & White, 2003 ; Maye, Werker & Gerken, 2002 ; Maye, Weiss & Aslin, 2008). Un phonème largement représenté dans le répertoire phonologique d'une langue et de surcroît saillant acoustiquement fera l'objet d'une réorganisation plus précoce.
- *L'adaptation du discours* des adultes (« motherese » ou « parentese ») adressé aux enfants (modifications de la fréquence fondamentale, de la durée des pauses et de la structure fréquentielle) facilite l'extraction des régularités statistiques de la langue d'exposition. Liu, Kuhl et Tsao (2003) ont établi une corrélation entre la taille de l'espace vocalique de mères parlant le mandarin et les performances de discrimination vocalique de leurs enfants âgés de 6-8 mois et 10-12 mois et ce indépendamment du niveau socio-économique.

En détectant les fréquences distributionnelles des phonèmes de sa langue, les catégories perceptives de l'être humain, contrairement à celles des autres animaux, sont réorganisées autour de phonèmes prototypiques qui *aimantent* les productions allophoniques[13] (Kuhl, 1991). Cet effet d'aimantation provoque l'élargissement de l'espace perceptif au niveau des frontières et le rétrécissement au niveau des prototypes (Grieser & Kuhl, 1989). Ainsi, à écart acoustique constant, les stimuli non prototypiques sont perçus comme plus proches du stimulus prototypique qu'entre eux (Kuhl, Williams, Lacerda, Stevens & Lindblom, 1992).

[12] On parle d'asymétrie directionnelle lorsque la discrimination d'un contraste est plus aisée dans un sens que dans l'autre.

[13] Les productions allophoniques sont les variantes non prototypiques d'un même phonème.

D'autres facteurs de nature physiologique et sociale participent également à cette réorganisation phonologique.

Concernant les *facteurs physiologiques*, Kuhl et al. (2008) précisent que cette spécialisation phonologique a lieu pendant une période critique (entre 6 mois et 3 ans) durant laquelle la réorganisation de l'espace phonologique modifie en profondeur les connexions neuronales. Guenther, Nieto-Castanon, Ghosh et Tourville (2004) ont notamment montré que l'effet d'aimantation provoque la diminution du nombre de cellules nerveuses impliquées dans le traitement des stimuli situés au centre d'une catégorie. Cette diminution explique que la discrimination intracatégorielle est moins aisée que la discrimination intercatégorielle pour laquelle un plus grand nombre de cellules sont recrutées. Au-delà des modifications induites par le développement des habiletés de perception, le NLM-e prévoit - contrairement à PRIMIR - une interaction entre les habiletés perceptives et de production. Par exposition linguistique, le sujet mémoriserait les patterns de production de sa langue, ce qui en retour renforcerait le lien établi entre pattern perçu et pattern produit. Ce lien entre perception et production a été mis en évidence par une technique d'imagerie fonctionnelle. Dehaene-Lambertz, Hertz-Pannier, Dubois, Mérieux, Roche, Sigman et Dehaene (2006) ont montré chez des enfants de 11 à 17 semaines issus d'un milieu francophone que la simple écoute de phrases de 2 secondes provoquait l'activation de l'aire de Broca avant même que ces enfants n'aient l'âge de babiller.

Concernant les *facteurs sociaux*, Kuhl, Tsao et Liu (2003) ont montré que le processus de spécialisation phonologique nécessitait une interaction sociale entre l'enfant et les membres de la communauté linguistique. Dans cette étude comparative, des enfants mandarins de 9 mois écoutaient une histoire présentée en modalité auditive seule (bande son) ou en modalité audiovisuelle (DVD) ou encore en modalité auditive mais via une personne présente dans la salle d'expérimentation. Les résultats obtenus à une tâche de discrimination d'un contraste mandarin qui revenait fréquemment dans l'histoire écoutée étaient meilleurs dans le groupe qui avait bénéficié d'une interaction sociale. Les deux autres groupes (bande son ou DVD) présentaient des résultats similaires. L'exposition linguistique est donc une condition nécessaire mais non suffisante pour amorcer le processus de spécialisation phonologique. Ces données amènent Kuhl à rejeter les modèles connexionistes qui selon elle ne prennent pas en compte la composante sociale : « in natural complex language learning situations, infants may

require a social tutor to learn, i.e. they are not computational automatons » (Kuhl et al., 2008, p.984).

La phase 3 du modèle est prédictive. Le processus de spécialisation phonologique entraîne une amélioration de diverses habiletés : la détection des régularités phonotactiques (Mattys, Jusczyk, Luce & Morgan, 1999), des probabilités transitionnelles entre segments et syllabes (Saffran, Aslin & Newport, 1996) et l'analyse de la structure phonétique des mots (Werker, Fennell, Corcoran & Stager, 2002). Récemment, Kuhl, Conboy, Padden, Nelson et Pruitt (2005) ont mis en évidence une corrélation positive entre les capacités de discrimination d'un contraste natif anglais (/ta-pa/) d'enfants de 7 mois et le niveau général de langage (Echelle McArthur : niveau de vocabulaire, de la complexité syntaxique…) de ces mêmes enfants à 14, 18 et 24 mois. Par ailleurs, et dans la même étude, les auteurs ont établi une corrélation négative entre les capacités de discrimination d'un contraste non-natif mandarin (fricatif vs. affriqué) et le niveau de langage. Ces deux corrélations *inversées* permettent d'écarter l'idée d'une simple supériorité auditive ou cognitive des enfants ayant un bon niveau de langage.

La phase 4 correspond à la clôture de la période critique. A ce point du développement, la plasticité cérébrale est bien moindre et la facilité avec laquelle une nouvelle langue est acquise diminue. Lors de l'apprentissage de la langue maternelle, la détection des régularités statistiques a profondément modifié les structures cérébrales dédicacées au traitement des informations linguistiques. Bien que possible, l'apprentissage d'une nouvelle langue est contraint par le degré de parenté entre la langue seconde et la langue native. Plus ces deux langues sont proches, plus l'apprentissage est aisé.

2.4.3. Synthèse

Les deux points communs majeurs entre PRIMIR (Werker & Curtin, 2005) et le NLM-e (Kuhl et al., 2008) sont d'une part de postuler que ce ne sont pas des détecteurs spécialisés ou des modules innés qui président au développement linguistique mais bien des prédispositions auditives de bas niveau que nous partageons avec nombre d'animaux non humains. D'autre part, la réorganisation de l'espace perceptif ne serait pas préprogrammée mais naîtrait des transactions constantes entre un substrat cérébral plastique et un environnement linguistique dans lequel l'être humain est capable de détecter des régularités.

L'idée d'un apprentissage statistique n'est pas neuve (Saffran et al., 1996) mais son application à la phonologie l'est davantage. Récemment, Maye et al. (2002) ont montré que la familiarisation d'enfants de 6 et 8 mois à des distributions de voisement unimodale ou bimodale avait une incidence sur la discrimination. Contrairement aux enfants familiarisés avec une distribution unimodale, les enfants habitués à écouter des stimuli appartenant à deux catégories montraient un effet de préférence lorsque dans la phase test les stimuli des deux catégories étaient présentés en alternance. Maye et al. (2008) ont répliqué ces résultats et montré que cet effet se généralisait à des contrastes voisés vs non voisés dont le lieu d'articulation différait des stimuli utilisés pendant la phase de familiarisation.

3. Problématique

Cette revue de littérature nous amène à considérer le développement du langage comme un processus *contraint* et *dynamique*. Les contraintes qui pèsent sur l'ontogenèse du langage sont, nous l'avons vu, multiples : génétique, auditive, articulatoire et cognitive. On peut à ce titre considérer le développement comme un processus universel et général. Le seul fait que l'être humain partage la quasi totalité de ses gènes avec des animaux qui ne parlent pas légitime cette approche. Ces contraintes n'empêchent toutefois pas d'appréhender le développement du langage comme un processus dynamique résultant d'une transaction constante entre ce qui est possible (les prédispositions universelles) et ce qui est disponible (les ressources environnementales).

S'agissant du voisement, nous nous sommes efforcés dans les trois études principales qui constituent la partie expérimentale de cette thèse, de considérer la perception à la fois comme une habileté contrainte par des prédispositions universelles mais aussi comme un processus évolutif émergeant par différentiation successive, i.e. comme une *épigenèse* pour reprendre la terminologie de Werker et Tees (1999).
Les études sur la perception du voisement sont légion mais rares sont celles qui considèrent une autre langue que l'anglais et qui investiguent la perception des indices de DEV positif et négatif. Pour ces raisons, nous avons étudié la perception de syllabes constituées d'une consonne occlusive apicale voisée ou non voisée et d'une voyelle neutre (/də/ et /tə/) chez des sujets d'âges différents (nourrissons, enfants, adultes) et de langue maternelle française. Dans cette langue en effet la frontière de voisement (0 ms DEV) ne correspond à aucune des

frontières universelles (-30 et +30 ms DEV), ce qui rend plus manifeste le processus de spécialisation phonologique. Partant de cette observation, nous avons cherché à comprendre pourquoi la frontière phonologique de voisement en français était située à mi-distance entre les frontières universelles et c'est pourquoi nous avons testé l'ensemble des sujets avec un continuum dont les variations de DEV permettaient d'investiguer la sensibilité aux frontières universelles et à la frontière phonologique du français.

L'objectif de la première étude était de mettre en évidence le processus de spécialisation phonologique en analysant les variations du rythme cardiaque de nourrissons âgés de 4 et 8 mois. Sur base des résultats obtenus par Lasky et al. (1975) et Eilers et al. (1979), notre hypothèse de travail était que les nourrissons de 4 mois seraient sensibles aux contrastes situés de part et d'autre des frontières universelles de voisement (-30 et +30 ms DEV) tandis que les nourrissons plus âgés discrimineraient mieux les paires de stimuli situés de part et d'autre de la frontière phonologique du français (0 ms). D'un point de vue méthodologique, la discrimination des différents contrastes devait se traduire par une décélération du rythme cardiaque.

Mais comment se poursuit le développement des habiletés perceptives de l'enfant une fois que celui-ci est devenu un spécialiste de sa langue ?

Dans une deuxième étude, nous avons évalué chez des enfants âgés de 5 à 8 ans les capacités d'identification et de discrimination de stimuli linguistiques (/də/ et /tə/) et non linguistiques (couleurs et expressions faciales). Deux questions ont motivé cette étude : l'apprentissage de la lecture a-t-il une influence sur la perception du voisement comme le suggère Burnham (2003) ? Les habiletés d'identification et de discrimination du voisement se développent-elles en synchronie avec d'autres habiletés comme pourraient le postuler les cognitivistes ?

Mais, le processus de spécialisation phonologique est-il irrémédiable ? Ne peut-on pas penser que le sujet adulte mature puisse rester sensible à des contrastes de voisement qui ne sont pas distinctifs dans sa langue ? Et encore : cette éventuelle sensibilité 'archaïque' aux frontières universelles de voisement ne nous informe-t-elle pas sur les mécanismes sous-jacents au processus de spécialisation phonologique ?

La troisième étude est présentée en deux parties : la partie 3.1, qui a été publiée dans un chapitre de livre, expose les fondements théoriques et les résultats préliminaires de nos travaux sur les corrélats neurophysiologiques du voisement. En testant un plus grand nombre de sujets et en affinant notre méthodologie, nous avons dans la partie 3.2 poursuivi ces recherches et enregistré les potentiels auditifs évoqués par les stimuli /də/ et /tə/ chez un groupe d'adultes francophones. Notre hypothèse était qu'il devait être possible d'objectiver chez des sujets francophones une sensibilité résiduelle aux frontières universelles de voisement en étudiant la morphologie (simple vs double pic) de la composante N100, un potentiel évoqué cortical qui a la caractéristique d'être évoquée par l'apparition d'un stimulus ou par un changement au sein d'un stimulus continu.

Etude 1

French native speakers in the making: from language-general to language-specific voicing boundaries[1]

By examining VOT discrimination in four and eight-month-old infants raised in a French-speaking environment, the present study addresses the question of the role played by linguistic experience in the reshaping of the initial perceptual abilities. Results showed that the language-general -30 and +30 ms VOT boundaries are better discriminated than the 0 ms boundary in four-month-old infants whereas eight-month-olds better discriminate the 0 ms boundary. These data support explanations of speech development stressing the effects of both language-general boundaries and linguistic environment (*Attunement theory*: Aslin & Pisoni, 1980; *Coupling theory*: Serniclaes, 2000). Results also suggest that the acquisition of the adult voicing boundary (at 0 ms VOT in French vs. +30 ms in English) is faster and more linear in French vs. English. This latter aspect of the results might be related to differences in the consistency of VOT distributions of voiced and voiceless stops between languages.

[1] Hoonhorst I, Colin C, Markessis E, Radeau M, Deltenre P, Serniclaes, W. French native speakers in the making: from language-general to language-specific voicing boundaries (in revision). Journal of Experimental Child Psychology.

Introduction

Much of the research on the early development of speech has focused on the change from a 'language-general' to a 'language-specific' pattern of perception (Kuhl, Williams, Lacerda, Stevens & Lindblom, 1992; for a review: Vihman, 1996). Throughout this article, we will use the term 'language-general' to refer to the basic acoustic ability shared by human and non-human animals to discriminate phonetic contrasts and 'language-specific' to refer to the phonological mode of perception.

Theoretical perspectives have considerably changed over the years, from fairly strong innate assumptions in early studies to a much greater emphasis on learning in recent work (Burns, Yoshida, Hill & Werker, 2007; Maye, Werker & Gerken, 2002; Maye, Weiss & Aslin, 2008; for a review: Kuhl, Conboy, Coffrey-Carina, Padden, Rivera-Gaxiola, & Nelson, 2008). However, there still remains a basic concern about the adaptation of the initial abilities to discriminate all the phonetic contrasts present in the world languages (Werker & Tees, 1984) to the categories present in the native language. This problem is central to the present study.

NATURE OF LANGUAGE-GENERAL PROPERTIES

There is a definite trend in current studies on speech development for viewing categorization processes in terms of prototype formation rather than in the emergence of boundaries. However, neuroimaging data obtained after training suggest that prototypes on the one hand and category boundaries on the other hand are represented in different brain regions (Guenther, Nieto-Castanon, Ghosh & Tourville, 2004). Further, training studies (Guenther, Husain, Cohen & Shinn-Cunningham, 1999) suggest that exposure to modal category values leads to a reduced discrimination of within-category differences, i.e. to prototype formation. On the contrary, exposure to stimuli at the limits of different categories leads to an increased discrimination of between-category differences, i.e. boundary emergence. Both mechanisms probably contribute to the build up of phonological categories and the adaptation of language-general settings to the native language can be viewed as a matter of using language-general boundaries for separating language-specific prototypes (in the perspective settled by Kuhl, 1993).

FROM LANGUAGE-GENERAL TO LANGUAGE-SPECIFIC SPEECH PERCEPTION

The phonological remapping mechanism, i.e. the change from a 'language-general' to a 'language-specific' pattern of perception, has been theorized in a lot of studies. In the

conceptual framework proposed by Aslin and Pisoni (1980) and Aslin, Werker and Morgan (2002), different theories have been described to explain the role played by linguistic exposure in the maturation of speech perception. In the frame of the *Universal theory*, perceptual ability is fully developed at birth and is either maintained or lost through native-language experience. On the contrary, the tenants of the *Attunement theory* postulate that perceptual ability is only partially developed at birth and that linguistic experience entails the facilitation of native-language perception on the one hand and the reduction of non-native phonetic contrasts discrimination on the other hand. In the *Perceptual learning* approach there is no initial perceptual ability, discrimination of native phonetic contrasts emerged only through induction based on early linguistic experience.

DEVELOPMENT OF VOICING PERCEPTION

The voicing distinction between stop consonants in initial position depends on Voice Onset Time (VOT), i.e. the delay between voicing onset and closure release (Lisker & Abramson, 1967). Concerning language-general to language-specific remapping in the perception of voicing categories, the most widespread theory is the *Attunement theory*. First, contrary to what has been proposed by the tenants of the *Universal theory*, there is growing evidence that the ability to discriminate non-native phonetic contrasts is not lost. Rather, non-native contrasts remain perceptible even though less better perceived than the native ones (Rivera-Gaxiola, Silva-Peyrera & Kuhl, 2005; Burns et al., 2007). Further, contrary to what was suggested by tenants of the *Perceptual learning theory*, several studies conducted with both infants and non-human animals on VOT continuum (e.g. Aslin, Pisoni, Hennessy & Perey, 1981 for data with infants; Sinnott & Adams, 1987 for data with monkeys) evidenced peaks of discrimination centered on negative and positive language-general boundaries located around -30 and +30 ms VOT (as first claimed by Lisker & Abramson, 1970). Data obtained with non-linguistic Tone Onset Time continuum (e.g. Jusczyk, Pisoni, Walley & Murray, 1980 for data with infants; Steinschneider, Volkov, Fishman, Oya, Arezzo, Howard, 2005 for data with animals) reinforced the conclusion drawn with pre-verbal infants that there exists some natural sensitivity to categorically perceive VOT before any linguistic influence. This sensitivity is thought to be grounded in some general auditory capacities shared by human and non-human animals as evidenced in electrophysiological data recorded either with positive (Steinschneider et al., 2005) or negative VOT values (Liégeois-Chauvel, de Graaf, Laguitton & Chauvel, 1999; Hoonhorst, Colin, Markessis, Radeau, Deltenre & Serniclaes, 2009).

These data suggest that rather than being uniform the initial perceptual space is already "warped" and an important challenge is to understand how this initial warping is later modified after contact with the native language. One account for this language-specific warping is the ability of infants to perceive the statistical regularities present in their native language. In recent models of development, e.g. the PRIMIR model (Processing Rich Information from Multidimensional Interactive Representations) of Werker and Curtin (2005) and the NLM-e model (Native Langue Magnet theory-expanded) of Kuhl et al. (2008) the language-general to language-specific remapping is linked to the sensitivity of infants to the distributional properties of speech sounds in their environmental language. Recently, Maye et al. (2002) and Maye et al. (2008) have provided evidence of this distributional sensitivity. Tested on the endpoints of a continuum, six and eight-month-old infants who had first been familiarized with a bimodal frequency distribution with stimuli concentrated at both ends of the continuum discriminated between the two stimuli presented whereas those who were familiarized with a unimodal distribution, with stimuli positioned in the middle of the continuum, did not.

CROSS-LINGUISTIC DIFFERENCES IN VOICING DISTRIBUTIONS

Even if most of the world languages are characterized by a bimodal VOT distribution, there are cross-linguistic differences. In English, VOT distributions are characterized by an opposition between short lag and long lag VOT (Kessinger & Blumstein, 1997) whereas in French or in Spanish VOT distributions oppose long lead vs. short lag (in e.g. French: Caramazza & Yeni-Komshian, 1974; Serniclaes, 1987; in Spanish: Williams, 1977). Differences in perception and/or production evidenced in cross-linguistic studies raised the question of differential timing in the development of perception in English vs. Spanish/French (see Table 1 for a summary). Lasky, Syrdal-Lasky and Klein (1975) showed that four-to-six-and-a-half-month infants raised in a Spanish-speaking environment were more sensitive to the negative and positive VOT language-general boundaries than to the 0 ms VOT boundary to which Spanish-speaking adults are sensitive. Using the same /ba-pa/ continuum as Lasky et al. (1975), Eilers, Gavin and Wilson (1979), compared six-to-eight-month-old infants raised in a Spanish vs. English environment and showed that infants of the first group yielded both a discrimination peak for the +10/+40 ms contrast which corresponds to the English phonological boundary and to the -20/+10 ms VOT contrast that straddles the adult Spanish boundary whereas infants of the English-speaking group only provided evidence of a

sensitivity for the +10/+40 ms contrast. Integration of the data from the two studies cited above leads to the first conclusion that the phonological shift occurs earlier in Spanish -between six and eight months of age- than in English where it occurs between ten and 12 months of age (as also demonstrated by e.g. Werker & Tees, 1984; Kuhl, Stevens, Hayashi, Degushi, Kiritani & Iverson, 2006). Secondly, these results showed that the phonological voicing boundary in Spanish (at 0 ms VOT) is not part of the initial language-general perceptual settings: "Linguistic listening experience may be a necessary prerequisite for the acquisition of lead VOT contrasts in infants" (Eilers et al., 1979, p 17).

However, this last conclusion was challenged by results obtained recently by Rivera-Gaxiola et al. (2005) and Burns et al. (2007). Rivera-Gaxiola et al. (2005) conducted a longitudinal study on English-speaking infants tested at seven months and at 11 months of age. Examining the morphology of the exogenous event related-potentials recorded while infants were stimulated with /da-ta/ syllables, they concluded that native voicing contrasts (straddling the English phonological boundary) and non-native ones (straddling the Spanish/French phonological boundary) were equally well discriminated by seven-month-olds. For 11-month-olds, the perception of non-native contrast worsened whereas the perception of the English contrast improved. In the same way but with a behavioral method (visual fixation length), Burns et al. (2007) tested six-to-eight-, 10-to-12- and 14-to-20-month-old infants raised in a monolingual English-speaking environment and in a bilingual French/English-speaking environment with words beginning with /ba/ and /pa/ natural syllables (with VOT values of eight, 28 and 48 ms). Whereas the linguistic status (i.e. monolingual vs. bilingual) did not influence the discrimination of six-to-eight-month-olds (all of them discriminated the French phonological boundary), results varied according to the linguistic status for both the 10-to-12- and the 14-to-20-month-olds, i.e. the monolinguals discriminated the French phonological voicing boundary and the bilinguals both the French and the English phonological boundaries.

Study	Stimuli	Method	Subjects				Discriminated Contrasts (ms)	
			Env. language			Age (months)	-VOT	- vs + VOT	+VOT
			E	S	F				
Eimas et al. 1971)	/ba/-/pa/	High Amplitude sucking	X			1; 4			20/40
Lasky et al. 1975)	/ba/-/pa/	Heart rate		X		4 to 6.6	-20/-60		20/60
Eilers et al. 1979)	/ba/;/pa/	Head Turning	X			6			10/40
				X		6		-20/10	10/40
Aslin et al. 1981)	/b/-/p/	Head Turning	X			5.6 to 11.6	-70/0		0/70
Rivera-Gaxiola et al. (2005)	/da/; /ta/	ERP	X			7		-24/12	12/46
			X			11		Idem but -24/12 less better discriminated than 12/46	
Burns et al. 2007)	/ba/; pa/; /pha/	Visual Fixation	X			6 to 8		-8/28	
			X			10 to 12			28/48
			X			14 to 20			28/48
			X		X	6 to 8		-8/28	28/48
			X		X	10 to 12		-8/28	28/48
			X		X	14 to 20		-8/28	28/48

Table 1: Summary of the main studies dealing with the perceptual development of infants raised in English- (E), Spanish- (S), and French (F) speaking environment (env.). Results corresponding to the discriminated contrasts within the negative (-VOT) or the positive (+VOT) part of the continuum or across the 0 ms VOT boundary (- vs + VOT) are expressed in ms.

The present study

The purpose of the current study was to collect data in French, a two-category language with a 0 ms VOT boundary, using a within-subject design so as to get additional insight into the mechanisms underlying the language-general to language-specific remapping in voicing perception. Our aim was to evidence the shift from initial voicing language-general boundaries to the phonological French adult boundary and to compare these results with those obtained for English in the literature. Following previous studies, our working assumptions

were that language-general VOT boundaries are located around -30 and +30 ms, and that the French VOT is located around 0 ms. Accordingly, our experimental design was aimed at testing changes in sensitivity to these contrasts between four and eight months of age. In order to test the sensitivities to the -30, 0 and +30 ms values in a relative way and to have some control over the assumption that these values are indeed better discriminated than adjacent contrasts on the continuum, we also used three other VOT contrasts (centered on -50, -20 and +50 ms). Finally, given that VOT boundaries depend on the phonetic context, we used syllables with fairly neutral stop and vowel places of articulation (/də-tə/ continuum), as the VOT boundary for these syllables is located around 0 ms for French listeners (Medina & Serniclaes, 2005).

Method

PARTICIPANTS

The initial sample consisted of 29 infants of the sleep unit of the Hôpital Universitaire des Enfants Reine Fabiola and those attending the nursery of Brugmann Hospital, both located in Brussels (Belgium). Infants were recruited after acceptance of the study by the ethical committee of both hospitals. Parents were informed of the procedures and signed a consent form. Data from 11 infants had to be rejected, seven due to exposure to a second language (parents bilingualism), seven due to fussiness, three due to crying and one due to experimenter's error. Finally, complete data were available for 18 infants, eight infants (four girls and four boys) with an average age of four months and nine days (SD=15 days) and ten infants (four girls and six boys) with an average age of eight months and seven days (SD=17 days). Anamnestic data attested that all infants were exclusively exposed to French at home, in their close environment and in the nursery.

STIMULI AND APPARATUS

We used stimuli varying along a /də-tə/ VOT continuum (similar to the syllables used in Medina & Serniclaes, 2005). With this continuum, the stop and vowel places of articulation are fairly neutral, the apical (/d-t/) place of articulation being midway between the labial and velar articulations and the /ə/ vowel being produced by the vocal tract in neutral position. Stimuli were generated by a parallel formant synthesizer (Klatt, 1980) provided by Carré (http://www.tsi.enst.fr/~carre/). The onsets of the initial frequency transitions of F1, F2 and

F3 were respectively 200, 2100 and 3100 Hz. F0 value was 120 Hz, transitions lasted 24 ms and the steady part 180 ms (figure 1). Stimuli were synthesized with nine different VOT values: -60, -40, -30, -20, -10, +10, +20, +40, +60 ms. The duration of the post release part of the stimuli was constant and the overall duration depended on the amount of the negative VOT. Negative VOT was synthesized with periodic energy (60 dB), F1 bandwidth at 50 Hz, and F2 and F3 bandwidths both at 600 Hz. Positive VOT was synthesized with aperiodic energy (30 dB), with F1 bandwidth at 600 Hz, and F2 and F3 bandwidths at 70 and 100 Hz, respectively. The voiced vocalic segment was synthesized with periodic energy (60 dB) and with F1, F2 and F3 bandwidths at 50, 70 and 100 Hz, respectively.

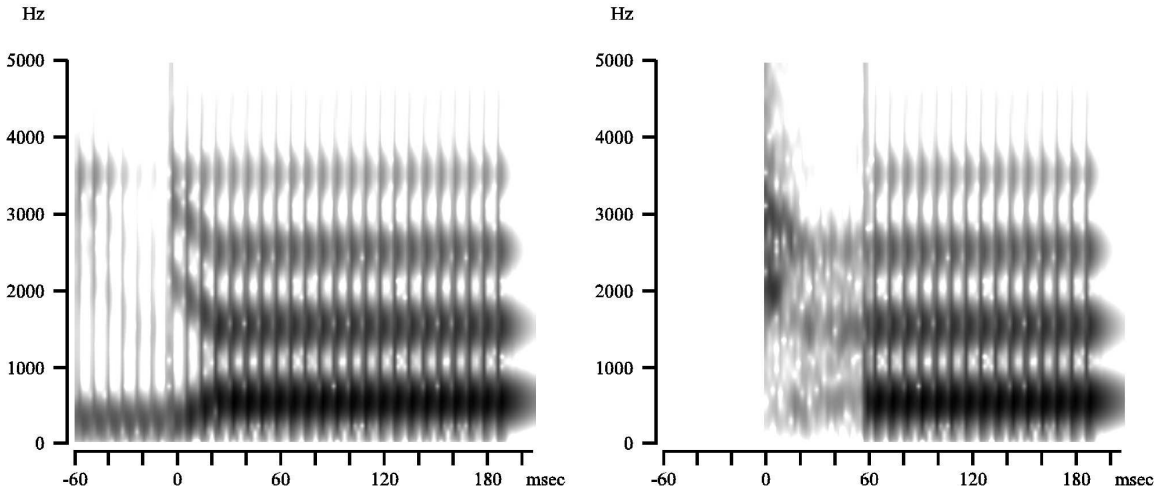

Figure 1: /də/ and /tə/ endpoint stimuli with -60 ms VOT (left) and +60 ms VOT (right).

Six sets corresponding to six stimulus pairs separated by 20 ms VOT (-60 /-40; -40/-20; -30/-10; -10/+10; +20/+40; +40/+60 ms centered on -50, -30, -20, 0, +30 and +50 ms VOT) were generated. Each set was composed of ten trials themselves consisting of 10 stimuli. The inter-trial interval lasted eight seconds and the inter-stimulus interval lasted 800 ms (figure 2). For each set, the subject was exposed to the habituation stimulus of the pairs described above during the first eight trials (i.e. 80 stimuli) and then to the dishabituation stimulus of the pairs during the last two trials (i.e. 20 stimuli). Each set lasted three minutes and the whole experiment 18 minutes.

To take an example, set A (figure 2) was composed of eight trials including the same -60 ms VOT habituation stimulus followed by two trials with the same -40 ms VOT dishabituation

stimulus. To take another example relative to set B, subjects were exposed to eight trials of -40 ms VOT habituation stimuli and two trials of -20 ms VOT dishabituation stimuli.

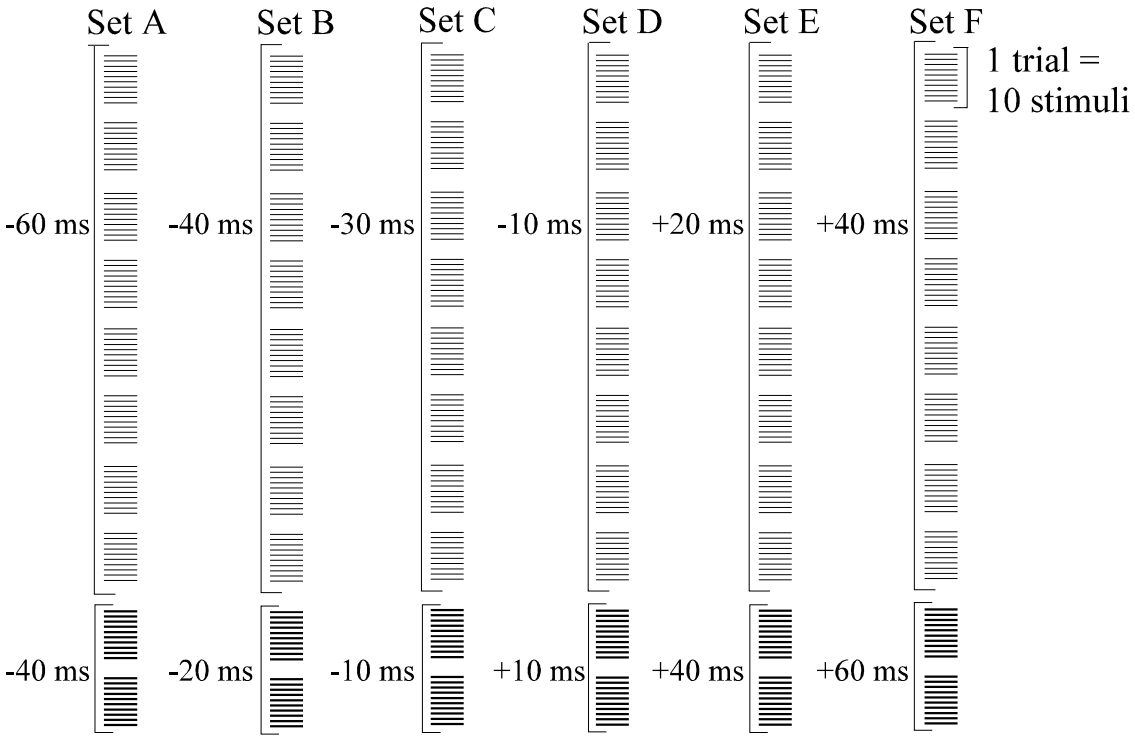

Figure 2: Stimulation paradigm of the experiment. Six sets (A, B, C, D, E, F) composed of ten trials were presented in a randomized order. Each trial consisted of ten stimuli.

Sensitivity to the -30 and +30 ms VOT contrasts was assessed with -40/-20 and +20/+60 ms stimulus pairs (corresponding to sets B and E on figure 2). Two control pairs centered on -50 and +50 ms , i.e. the -60/-40 and +40/+60 ms pairs corresponding to sets A and F (figure 2), were used for testing these sensitivities in a relative way. Sensitivity to the 0 ms VOT contrast was assessed with a -10/+10 ms stimulus pair (corresponding to set D on figure 2) and a -30/ -10 ms control pair (corresponding to set B in figure 2).

PROCEDURE

Infants were seated on the laps of one of their parents in a sound-attenuated booth. An experimenter was inside the booth to hold the infant's attention thanks to silent toys. Another experimenter was outside the booth and managed the presentation of stimuli. Stimuli were presented binaurally at 70 dB SPL via a loudspeaker (Amplaid MK3) situated 80 cm in front of the infant. Intensity was calibrated with a Larson-Davis sound level meter (800b) with the artificial ear placed at the level of the baby's head. Stimuli were produced by the sound card

of a laptop running i-shell v.2.5 software (www.tribeworks.com) and amplified with a Stereo Amplifier (3225 PE Power Envelope NAD). Heart rate was monitored with a set of Bluesky electrodes placed on the left wrist, the right and the left ankle (Bartoschuk, 1962) connected to an electrophysiological acquisition system (Ampli Portilab 5-16/ASD, ANT Software). Electrocardiogram was amplified 20 times, band-pass filtered (0.5-35 Hz) and digitized at 128 Hz by a data acquisition software (Portilab, ANT Software). The computer software identified the R-wave of the ECG and computed the heart rate in beats-per-minute (bpm).

DATA PROCESSING

Thanks to careful visual screening, transitory events having much shorter duration than the QRS complexes were identified as movement artifacts. The erroneous and abnormally short bpm values -due to these transitory events- were replaced by the preceding bpm value.

Two scores were then computed. The *Variation Score* was calculated for each trial by taking the difference between the mean of the longest two bpm values recorded between the fourth and the sixth second after trial onset (Moffit, 1971; Leavitt, Brown, Morse & Graham, 1976) and the mean of the longest two bpm values recorded during the five second before trial onset. The variation score was meant to provide a titration of the habituation effect known to reduce the heart rate deceleration response triggered by a stimulus sequence as the sequence went on (Lacey & Lacey, 1958). The *Dishabituation Score* was then computed. It corresponded to the difference between variation scores on trials nine and eight. As trial eight was the last trial containing the standard stimulus and trial nine the first containing the deviant stimulus, the dishabituation score was used as an index of discrimination.

DATA ANALYSIS

Dishabituation scores were submitted to an ANOVA with Stimulus pair (six pairs centered on -50, -30, -20, 0, +30 and +50 ms) as the within subject factor and Group (two levels: four or eight months) as the between-subject factor.

Table 2 shows the planned contrasts used to test three couples of stimulus pairs: -60/-40 vs. -40/-20; -30/-10 vs. -10/+10; +20/+40 vs. +40/+60 ms, each couple corresponding to the difference between sensitivity to one of the boundaries under scope (-30, 0, +30 ms) and its within-category control pair (-50, -20, +50 ms respectively). A fourth planned contrast was used to test the difference between the 0 ms boundary and the -30 and +30 ms boundaries taken together.

Results

Figure 3 shows the dishabituation results obtained by the four- and eight-month-olds for each of the six contrasts. Graphically, it is striking to note that the pattern of results obtained by the eight-month-olds mirrors the pattern of results obtained by the four-month-olds. As expected, two dishabituation peaks at -30 and +30 ms were found for the four-month-old group and one dishabituation peak was found at 0 ms for the eight-month-old group. More surprisingly, there were also two dishabituation peaks at -50 and +50 ms for this latter group.

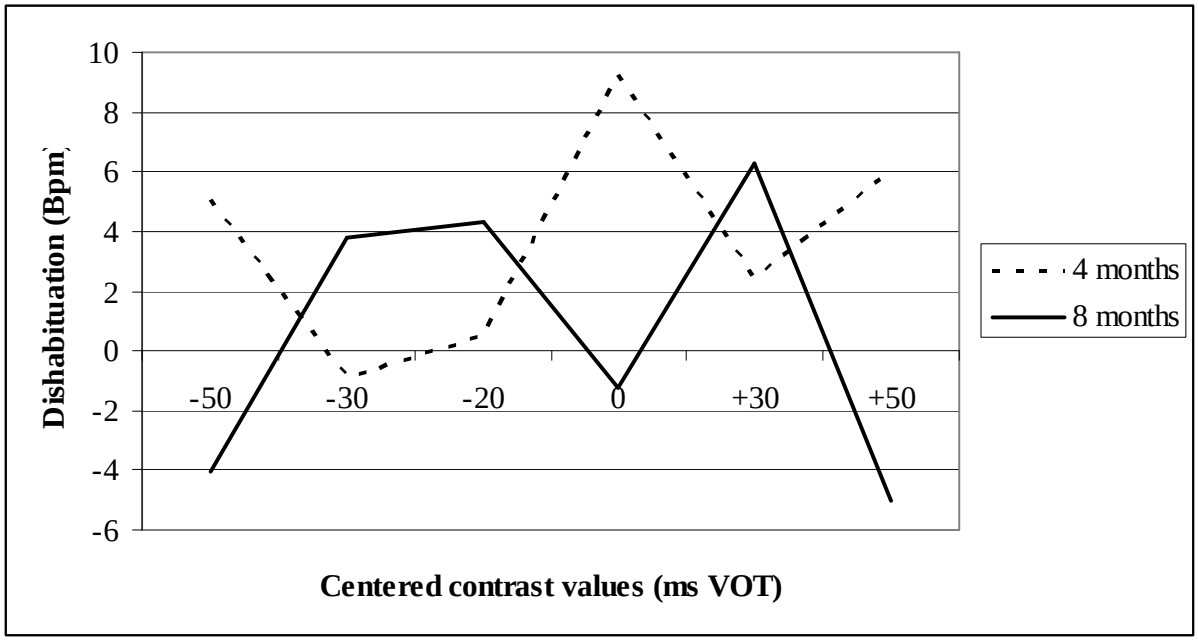

Figure 3: Dishabituation scores (expressed in beats per minute) for the six tested pairs (centered on -50, -30, -20, 0, +30, +50 ms VOT).

Results of the ANOVA showed that effects of group and pair were not significant (both $F<1$). However, the Pair X Group interaction was significant ($F(5,90) = 2.47$, $p < .05$, $\eta^2 = .13$). Table 2 shows that there were significant differences between groups for the contrasts testing 0 and +30 ms boundaries whereas the contrast testing the -30 ms boundary was not significant. Further, the planned contrast opposing the VOT pairs centered on -30, 0 and +30 ms VOT relative to their control pairs was also significant, hence indicating a difference in sensitivity to the language-general (-30 and +30 ms VOT) and language-specific boundary (0 ms VOT).

Planned Contrast	F ratio F(1,16)	p value	Effect size (η^2)
-50/-30	1,65	> .05	.09
-20/0	4.32	= .05 *	.21
50/30	7.98	= .01 *	.33
-20/0 vs. -50/-30 & 50/30	13.82	< .01 *	.46

Table 2: Results (F ratio, p value and size effect) of the four planned contrasts used to determine which stimulus pairs were differently perceived by four-month- and eight-month-old infants. Significant results are flagged by a star.

Discussion

The present research evidenced important changes in VOT perception between four and eight months of age for infants raised in a French-speaking environment. From four to eight months, there was a *decrease* in sensitivity to the language-general boundaries, at -30 and +30 ms, although this trend was only significant for the latter boundary, and there was also a significant *increase* in sensitivity to the French 0 ms VOT boundary. These results provide a straightforward confirmation to what could be inferred from a set of previous results (Lasky et al., 1975 and Eilers et al., 1979) which suggested on the whole that there are initial language-independent sensitivities to negative and positive VOT boundaries, that the 0 ms VOT boundary is not part of these initial settings and that this sensitivity occurs later in the course of perceptual development (see Introduction, Table 1). However, while these complex changes were up to now separately observed in different studies, they were evidenced here within the same study, using a common methodology and the same contrasts.

Changes in sensitivity to the -30 and +30 ms VOT values vs. the 0 ms one were expected from earlier studies. Our working hypothesis was that language-general boundaries were most probably located around -30 and +30 ms and that the French boundary was located around 0 ms. Accordingly, we expected changes from language-general to language-specific to occur around these VOT values, and this was confirmed by the results. More surprisingly, the present results also revealed better discrimination of the contrasts centered on -50 and +50 ms

VOT in eight-month vs. four-month-old infants. What we see (figure 3) is a general change in discrimination along the VOT continuum, from discrimination peaks at -30/-20 and +30 ms at four months to -50, 0 and +50 ms peaks at eight months. These changes are more complex than those expected at the start. It would seem that the sensitivity to negative and positive VOT contrasts remains present at eight months although located at longer VOT values (50 instead of 30 ms). While the exact nature of these changes remains to be understood, what is already clear is that the early development of VOT perception does not merely consist in the replacement of the -30 and +30 ms boundaries by the 0 ms VOT boundary. Instead, what seems to occur is a global change in the "warping" of the VOT dimension characterized by the emergence of a central boundary (at 0 ms) and a shift of negative and positive boundaries towards more extreme values (from -30 and +30 ms to -50 and +50 ms).

Considering our results, one might wonder why the language-general to language-specific remapping seems to occur earlier in French (between four and eight months) than in English (after eight months, Werker & Tees, 1984). The present results suggest that infants raised in French are sensitive to the 0 ms boundary between four and eight months, and similar results were obtained both for infants raised in Spanish (Eilers et al., 1979) and in English (Rivera-Gaxiola et al., 2005; Burns et al., 2007). However, while the 0 ms boundary corresponds to the adult one in French and Spanish, this is not the case in English where the adult boundary (discriminated after eight months) is located around +30 ms.

One possible account for this difference in the timing of development might be the differences in voicing distributions. As already mentioned in the Introduction, French and English do not share the same distributional properties. The voiced and voiceless categories are adjacent in English (e.g. Lisker & Abramson, 1967) vs. distant in French (figure 4). As suggested by Hay (2005), this difference does not favor English-speaking subjects, from whom greater temporal perceptual acuity is requested.

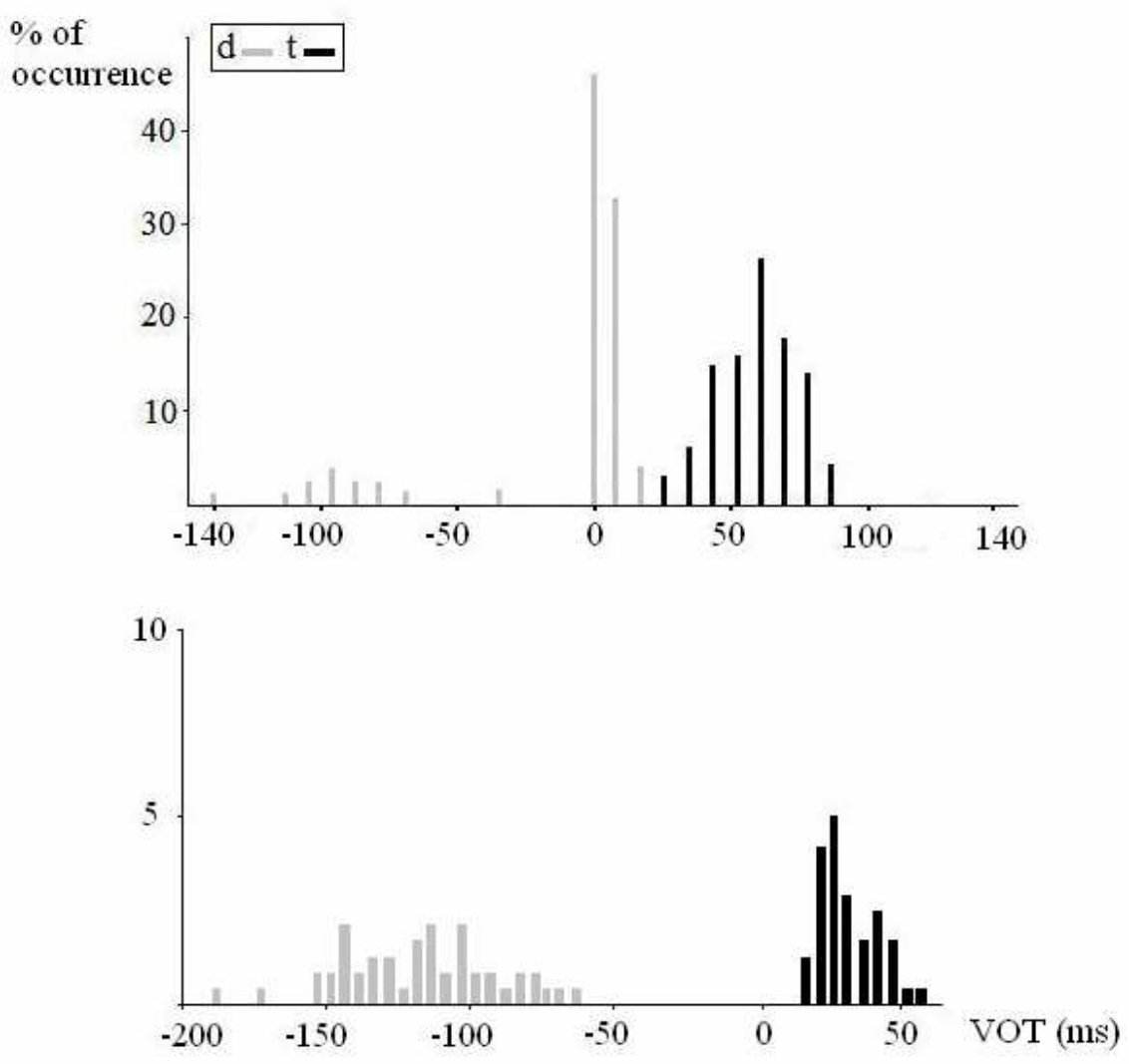

Figure 4: **Top**: Alveolar /d-t/ distributions in English (500 words uttered as isolated items by four speakers ; redrawn and adapted with permission from Lisker & Abramson, 1967). **Down**: Alveolar /d-t/ distributions in French (three vowel contexts, /aiu/, 16 speakers ; adapted from Serniclaes, 1987).

A further obstacle to the acquisition of the English boundary might arise from the fact that English adult productions are not completely located on the positive side of the VOT continuum. As shown by Lisker and Abramson (1967), a small but non negligible proportion of adult English voiced stops are produced with negative VOTs. The exposure to stops with either negative or positive VOTs, even though the latter are much more frequent, might direct the infant towards the 0 ms VOT boundary. This trend would be prompted by the search of a boundary fitting with the VOT categories easier to discriminate. The search for the "easiest perceptual solution" would guide infants raised in an English environment towards the 0 ms boundary as a first choice. In support of this explanation, VOT distributions of stops produced

by children raised in an English environment are transitorily more biased towards negative VOT values during the early childhood (see Zlatin & Koenigsknecht, 1975 for prevoiced productions in two and six year-old English-speaking children). However, at a later stage, the fairly low frequency of negative VOT in adult speech and/or the weak relevance of negative VOT for operating lexical distinctions would lead English children to opt for the positive VOT boundary.

At a theoretical level, the fact that we did not evidence any enhanced discrimination around the 0 ms phonological boundary at four months of age prevents us from supporting the *Universal theory* described in Aslin and Pisoni (1980), which is characterized by the initial ability of infants to discriminate all the possible phonetic contrasts. On the contrary, the present results are better explained by language-specific *attunements* of boundaries, i.e. infant's perception sharpened by linguistic experience. However, the emergence of a language-specific boundary which is not present in the universal repertoire, e.g. the 0 ms VOT boundary, cannot be explained by the mere attunement of a single boundary. Such an entirely new language-specific boundary might simply emerge after exposure to the stimulus distributions in the linguistic environment. However, data collected with adults suggest that the 0 ms VOT boundary (as well other similar language-specific boundaries) arises from combinations between pairs of language-general boundaries (Serniclaes, 2000 ; Serniclaes & Geng, in press). These combinations are fairly complex because language-general boundaries have to be adapted to new purposes. For instance, simply adding the percepts generated by the negative and positive boundaries for perceiving the French VOT categories would give rise to weak categorical precision around 0 ms. This problem is solved by *coupling* the negative and positive VOT percepts, i.e. by relating the perception of presence vs. absence of negative VOT to the perception of presence vs. absence of positive VOT (Serniclaes, 2000). The consequence of this "percept-percept coupling", which is similar to those evidenced in visual perception (Epstein, 1982), is an increase in the accuracy of both negative and positive VOT perception -instead of requiring a bit more than 30 ms for perceiving each cue, only a bit more than 0 ms is sufficient- as well as a concomitant increase in categorical precision around the 0 ms boundary.

In the future, it would be of interest to further explore the complex changes from the initial warping of the perceptual space to the language-specific ones, in particular the further development of perceptual sensitivities along the VOT dimension between eight and 12 months of age in languages with different VOT distributions. It would also be of interest to

examine the sensitivity of French-speaking adults to language-general boundaries and to the French adult phonological boundary with electrophysiological methods, e.g auditory evoked-potentials, so as to get an insight into the neural code of voicing perception.

Conclusion

The change in the "warping" of the VOT dimension characterized by the emergence of a central boundary (at 0 ms) and a shift of negative and positive boundaries towards more extreme values in eight vs. four-month-old infants raised in a French speaking environment supports explanations of speech development which stress the joint effects of both language-general boundaries and linguistic environment (*Attunement theory*: Aslin & Pisoni, 1980; *Coupling theory*: Serniclaes, 2000). Further, this shift from a language-general to a phonological mode of perception occurs at the same time (i.e. between four and eight months) for infants raised in a French or a Spanish environment and occurs earlier than for infants raised in an English-speaking environment (after eight months). The explanation we propose for this difference in developmental timing for voicing perception stems from significant differences in the distributional properties of the language-specific productions, reinforcing the idea that distributional properties play an important role in infants' perceptive specialization.

Etude 2

Development of categorical perception: comparisons between voicing, colors and facial expressions[1]

The aim of the present paper was to compare the development of perceptual categorization of voicing, colors and facial expressions in French-speaking children (aged from six to eight years) and adults. Differences in both Categorical Perception, i.e. the correspondence between identification and discrimination performances, and in Boundary Precision, indexed by the steepness of the identification slope, were investigated. Whereas there was no significant effect of age on Categorical Perception, Boundary Precision increased with age, both for voicing and facial expressions though not for colors. Further, the precision of the voicing boundary was correlated with reading abilities. These results are discussed in the light of two developmental theories: the reading hypothesis and the general cognitive hypothesis.

[1] Hoonhorst I, Medina V, Colin C, Markessis E, Radeau M, Deltenre P, Serniclaes W. Development of categorical perception: comparisons between voicing, colors and facial expressions (in revision). *Journal of Experimental Child Psychology*.

Introduction

CATEGORICAL PERCEPTION

Categorical perception is commonly defined by the mismatch between the monotonic variation of a set of stimuli regularly varying along a physical continuum and its resulting non-monotonic perception. Harnad (1987) emphasized two characteristics associated with categorical perception: "1/ a set of stimuli ranging along a physical continuum is given one label on one side of a category boundary and another label on the other side and 2/ the subject can discriminate smaller physical differences between pairs of stimuli that straddle that boundary than between pairs that are entirely within one category or the other".

This definition was shown to suit both non linguistic and linguistic continua. Cutting and Rosner (1974) evidenced that auditory rise-time variations are perceived categorically either as presented in a linguistic or in a non-linguistic continuum. Bornstein, Kessen and Weiskopf (1976) evidenced the same categorical mode of perception with color hues and Ekman (1992) with facial expressions. Concerning linguistic continua, Liberman, Harris, Hoffman and Griffith (1957) demonstrated that stimuli varying along a /b–d–g/ continuum, i.e. stimuli varying in place of articulation, are perceived categorically. The same conclusion was reached with continua where Voice Onset Time (VOT), i.e. the delay between voicing onset and closure release (Lisker & Abramson, 1967), was modified. Specifically, the categorical perception of VOT was evidenced in human adults (Abramson & Lisker, 1970), human infants as young as one month of age (Eimas, Siqueland, Jusczyk & Vigorito, 1971) and even in non-human animals such as chinchillas (Kuhl & Miller, 1975; 1978). It is now accepted that categorical perception is neither speech- nor human-specific.

THE DEVELOPMENT OF CATEGORICAL PERCEPTION

The issue of the development of categorical perception during infancy has been largely documented. As far as VOT is concerned, it is now accepted that during the first year of life, infants move from a language-general to a language-specific mode of perception, i.e. whereas all babies discriminate voicing contrasts according to the three voicing categories separated by two universal boundaries located at -30 and +30 ms VOT (Aslin, Pisoni, Hennessy & Perey, 1981) during the first months of life, they soon become specialists in their mother tongue by adopting the phonological boundary/ies relevant in the language spoken in their

environment (0 ms VOT in French: Serniclaes, 1987). Far less is known about the following steps of maturation in voicing perception. Zlatin and Koeningsknecht (1975) tested two- and six-year-old English-speaking children with words beginning with a voiced or a voiceless phoneme and showed that from two to six years of age the identification function undergoes a development leading to greater Boundary Precision. Hazan and Barrett (2000) spanned these results to different phonetic contrasts and to English-speaking children from six to 12 years of age. Concerning discrimination scores, Elliott, Busse, Partridge, Rupert and de Graaf (1986) evidenced a negative correlation between age and the size of the just noticeable difference discriminated by English-speaking children from eight to 11 years when tested on a /ba-pa/ continuum.

This study was designed as a continuation of these researches. We aimed at documenting 1/ the development of categorical perception and 2/ the determinants involved in this development, the difficulty being the interacting influences between the different factors of maturation. This study is composed of three experiments, one pertaining to the development of voicing perception and the last two pertaining to the development of colour and facial expressions perception. Specifically, the comparison between these three experiments enabled us to test two specific hypotheses: The reading hypothesis (Burnham, Earnshaw & Clark, 1991; Burnham, 2003) according to which "the intensification of language speech perception between two and six years is related to the onset of reading instruction" (pp. 573) and the general cognitive hypothesis (Karmiloff Smith, 1991; Lalonde & Werker, 1995) according to which categorical perception of speech sounds evolves in synchronicity with other abilities through the influence of cognitive maturation. According to the reading hypothesis, there should be no difference between age groups for either color or facial expression perception.

Since most of the studies in the field have used either identification or discrimination tasks, it seemed important for the purpose of the present study to use both identification and discrimination results and to assess the correspondence between the two as explained in the following paragraph.

METHODOLOGICAL CONSIDERATIONS

Even if the methodology differs from one study to the other, the two golden standard tasks used to assess categorical perception remain the identification task in which the subject is requested to label the stimulus and the discrimination task where the subject is asked to determine if the items of a stimuli pair are identical or not. The resulting identification

function is used to determine the *Boundary Precision*, i.e. the steeper the identification function, the greater the precision (Simon & Fourcin, 1978) and to predict the discrimination scores thanks to a probabilistic formula (Pollack & Pisoni, 1971). The comparison between the discrimination scores expected from identification scores and the observed scores obtained through the discrimination task represents the classical *Categorical Perception* test. The stronger the correspondence between the expected and the observed discrimination scores, the stronger the Categorical Perception (Damper & Harnad, 2000). Throughout the paper, we will distinguish categorical perception as the general phenomenon by which a monotonic variation leads to a non–monotonic perception and Categorical Perception (written with capital first letters) as the test that compares expected and observed discrimination scores.

Although either the Boundary Precision or the Categorical Perception tests have been extensively used to assess categorical perception, few studies have integrated them in a single study. Yet, these tests are complementary rather than redundant. Figure 1 presents the possible combinations between Boundary Precision and Categorical Perception measures. In example a/ both Boundary Precision and Categorical Perception are strong, a situation that corresponds to perfect categorical perception. In example b/ Boundary Precision is strong but Categorical Perception is weak indicating a mismatch between expected and observed discrimination scores (described in Damper & Harnad, 2000 ; Medina, Hoonhorst, Bogliotti, Sprenger-Charolles & Serniclaes, submitted). In example c/ Boundary Precision is weak but Categorical Perception is strong reflecting the pattern of results obtained by illiterate adults (Serniclaes, Ventura, Morais & Kolinsky, 2005). In example d/ Boundary Precision and Categorical Perception are weak reflecting both poor labeling abilities and a mismatch between expected and observed discrimination scores.

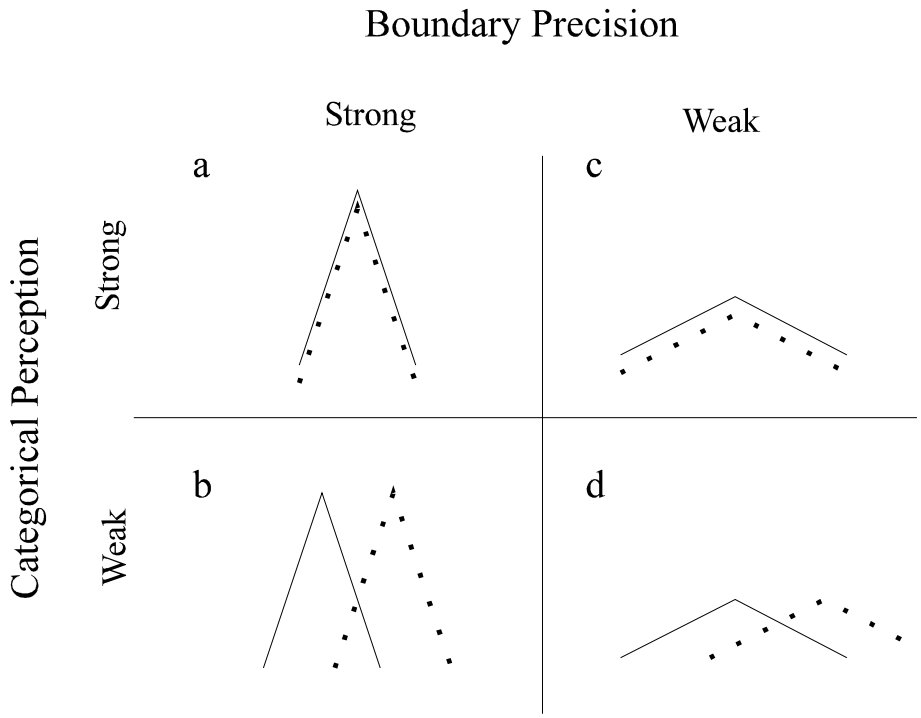

Figure 1 : Categorical model of perception integrating both Boundary Precision and Categorical Perception indexes as co-factors. Boundary Precision is assessed by the magnitude of the discrimination peak, as illustrated by the differences between the right-hand and left-hand columns of the figure. Degree of Categorical Perception is inversely related to the difference between observed discrimination scores (continuous lines) and those expected from identification (dotted lines), as illustrated by the difference between upper and lower lines of the figure.

EXPERIMENT 1

Method

PARTICIPANTS

Four age groups composed of 30 children and 21 university students participated in this experiment. The university students were chosen among a group of 72 monolingual and native speakers of French by selecting those who had the lowest level of English vocabulary (Peabody Picture Vocabulary Test revised, Dunn & Dunn, 1981). The 21 students (three men and 18 women) attended the first year of University (mean age: 19 years four months, SD: 13 months). The children were also monolingual native speakers of French and belonged to three

different age groups: 11 children (six boys and five girls) attended last year of kindergarten (mean age: five years 10 months, SD: two months), eight children (three boys and five girls) attended the first grade of primary school (mean age: six years 10 months, SD: five months), and 11 children (five boys and six girls) attended the second grade of primary school (mean age: seven years nine months, SD: seven months). All subjects reported no history of auditory and language disorders. For children, parents gave their written consents.

STIMULI

Speech stimuli were synthesized syllables composed of the apical stop consonant /d/ or /t/ followed by the neutral vowel /ə/. Six syllables with VOTs varying from -75 ms to +75 ms with a 30 ms acoustic step were created. Stimuli were generated by a parallel formant synthesizer (Klatt, 1980) provided by R. Carré (CNRS, France). The onsets of the initial frequency transitions of F1, F2 and F3 were 200, 2200 and 3100 Hz, respectively and the formant transitions lasted 20 ms. Steady state formant frequencies were 500, 1500 and 2500 Hz, respectively. The F0 value was constant at 120 Hz. The overall duration of all the syllables was 200 ms. Negative VOT was synthesized with periodic energy (60 dB), F1 bandwidth at 50 Hz, and F2 and F3 bandwidths both at 600 Hz. Positive VOT was synthesized with aperiodic energy (30 dB), with F1 bandwidth at 600 Hz, and F2 and F3 bandwidths at 70 and 100 Hz, respectively. The voiced vocalic segment was synthesized with periodic energy (60 dB) and with F1, F2 and F3 bandwidths at 50, 70 and 100 Hz, respectively. Figure 2 presents the spectrogram of the voicing continuum.

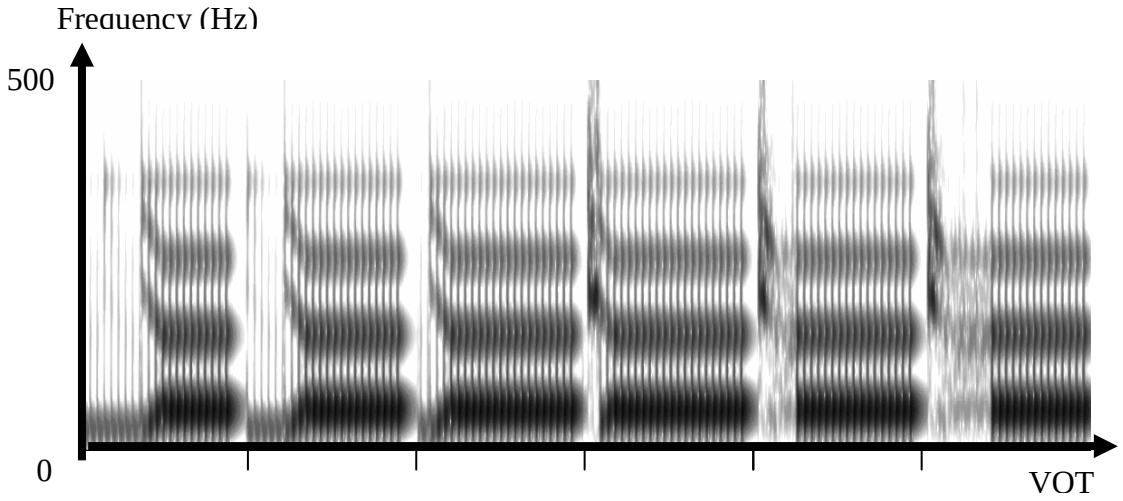

Figure 2 : /də-tə/ continuum spectrogram. The synthetic stimuli varied along a VOT ontinuum from -75 to +75 ms in 30 ms step (from left to right: -75, -45, -15, +15, +45, +75 ms).

PROCEDURE

Children were individually tested in a quiet room in their school during two sessions (lasting together one hour) whereas adults were tested in a single 30 minutes session. The familiarization and language screening tests described below were only administrated to children.

Familiarization:

Identification. Children were first familiarized with the task. Two characters of familiar children's books who were renamed Dom and Tom so that the initial phonemes corresponded to the tested ones were presented to the children. The correspondence between the character and the phoneme was explicitly formulated through a small computer game in which the child was requested to repeat the sound produced by the character. The sounds corresponded to the continuum endpoints, i.e. -75 ms VOT for the /də/ syllable and +75 ms VOT for the /tə/ syllable. The endpoints stimuli were then randomly presented (10 times each) in a preliminary identification session in order to familiarize the child with the procedure. Presentation was controlled by E-prime 1.1. (Psychology Software Tools). Children were asked to associate the stimulus with the picture of one of the two characters pasted on a button box whereas adults were asked to respond by answering either /də/ or /tə/ by pressing a button on the computer keyboard. A correct response was reinforced by a visual feedback. When 70% of the responses were correct, subjects were admitted to the identification test. During the identification test, each of the six stimuli of the continuum (-75, -45, -15, +15, +45, +75 ms VOT) was presented 10 times. The inter-stimulus interval was controlled by the subject since his response entailed the next stimulus presentation. Stimuli were presented in a randomized order and at a comfortable level through Sennheiser HD650 headphones.

Discrimination: the discrimination task was preceded by a practice phase in which the endpoints pairs (-75/-75 ms; -75/+75 ms ; +75/-75 ms ; +75/+75 ms) were presented five times. Children were asked to associate the pair of stimuli with one of the two pictures pasted on a button box. The picture associated to the "same" response represented a pair of identical characters (Dom next to Dom and Tom next to Tom) whereas the picture associated to the "different" response represented two different characters (Dom next to Tom and Tom next to Dom). Contrary to children, adults were asked to respond by pressing a button on the computer keyboard. For the discrimination test, each stimulus was presented paired with itself

(six identical pairs: -75/-75 ; -45/-45 ; -15/-15 ; +15/+15 ; +45/+45 ; +75/+75) and with the adjacent stimulus (ten different pairs, i.e. five stimulus combinations (-75/-45 ; -45/-15 ; -15/+15 ; +15/+45 ; +45/+75) x two orders (e.g. -75/-45 and -45/-75). Each pair was presented five times. The subject's response entailed the next pair presentation and the inter-stimulus interval between two stimuli within each pair lasted 100 ms.

Language screening: all children performed the French version of the Peabody Picture Vocabulary Test (Echelle de Vocabulaire en Images (EVIP), Dunn, Thériault-Whalen & Dunn, 1993). Behalf the youngest group, children achieved scores on the Odédys reading test (Laboratoire Cogni-sciences, IUFM de Grenoble, 2002) to assess the ability to read 20 regular words, 20 irregular words and 20 pseudo-words. Both correctness (number of words well read /20) and rapidity were evaluated.

DATA PROCESSING

Boundary Precision assessment was based on the slope of the identification curve, a shallower slope indicating lesser precision. The slope was measured separately for each subject using logistic regression (see Equation 1, Mc Cullagh & Nelder, 1983) with the Labeling Response as the dependent variable and VOT stimulus as the independent variable. The logistic function is fairly simple and it has been frequently used for fitting identification curves in the studies on speech perception (e.g. Nearey, 1990), though other functions such as the Cumulative Normal (Finney, 1971) were also used (e.g. Hazan & Barrett, 2000). Equation 1 gives the most general form of the logistic function. Individual assessments of slopes were then used for testing the difference between groups with ANOVA.

$$P(\text{response}) = e^{y} / (e^{y} + 1)$$

Equation 1 : Formula of the logistic function. Where y = Logit (P) = log (P/(1-P)) = I + S*VOT ; I stands for the intercept and S corresponds to the slope of the identification function. The boundary, which corresponds to P = 0.5 or to Logit(P) = 0, is obtained by taking –I/S.

Categorical Perception was assessed by comparing the size of the observed discrimination d' scores with those expected from the identification data the latter being computed with elementary probability formulas (adapted from Pollack & Pisoni, 1971). For each VOT pair

(e.g. for the +45/+75 ms VOT pair), the observed discrimination responses and those derived from the identification data (expected scores) were converted into d' scores by taking the difference between the normal-deviate (z value) corresponding to the rate of "same" responses to the "same" pairs (e.g., the "same" responses to the +45/+45 ms pair and to the +75/+75 ms pair) and the rate of "different" responses to the "different" pairs (e.g., the "different" responses to the +45/+75 ms and +75/+45 ms pairs). Before conversion into z values, response scores above 50% were reduced by 2.5% (with 20 responses per couple of pairs, 2.5% corresponds to one half of the precision of the response scale, i.e. one half of 1/20), and those below 50% were increased by 2.5%.

Results were analyzed by repeated measures ANOVAs. Whenever Mauchly's sphericity test was significant, Greenhouse-Geisser corrected degrees of freedom values were used.

Results

IDENTIFICATION

The mean function slope for the four groups of subjects is presented on figure 3.

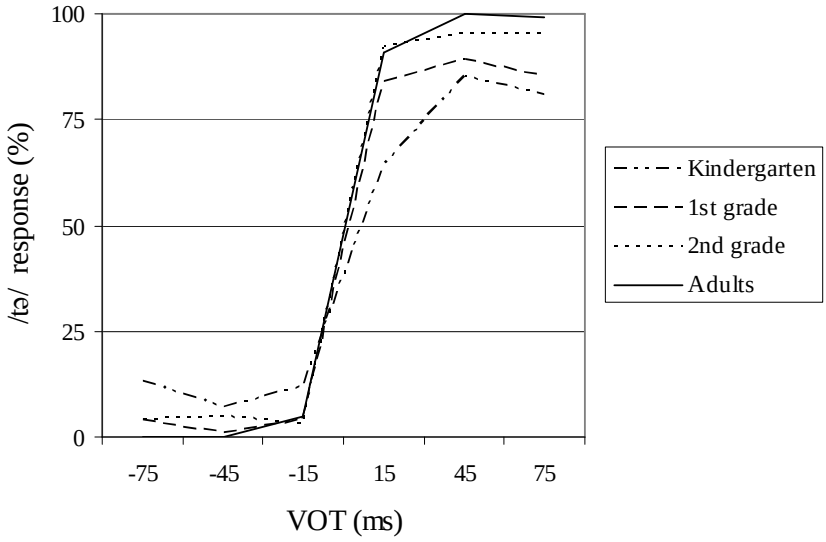

Figure 3 : Identification function on the /də-tə/ VOT continuum for children attending kindergarten, first and second grade of primary school and adults.

Differences in mean boundaries and slopes between the groups were tested by univariate ANOVAs. The mean boundary was not significantly related to the groups ($F(3,50) = 2.44$, $p > .05$; $\eta^2 = 0.13$) and was located at 6.6 ms (SD: 16.5 ms) for the four groups taken together.

Boundary Precision was assessed by the Group x Slope interaction. The slope was significantly related to the groups ($F(3,50) = 5.55$, $p < .01$, $\eta^2 = 0.26$), the mean slope being steeper for the adults (0.77 logit/ms) than for second grade (0.38 logit/ms), than for first grade (0.29 logit/ms) and for kindergarten (0.15 logit/ms). Posthocs comparisons revealed that the difference in Slope between kindergarten and adults was significant ($F(1,47) = 13.73$, $p < .01$) whereas the differences between the other groups were not (Kindergarten/1st grade: $F(1,47) = 0.44$, $p = 3.07$; Kindergarten/2nd grade: $F(1,47) = 1.47$, $p = 1.39$).

DISCRIMINATION

The expected and observed discrimination scores for the four groups are presented in figures 4 and 5. Graphically, it can be noted that maximum scores were obtained for the VOT value centered on 0 ms whatever the group. For observed d' scores, a second peak was located on the VOT value centered on +30 ms.

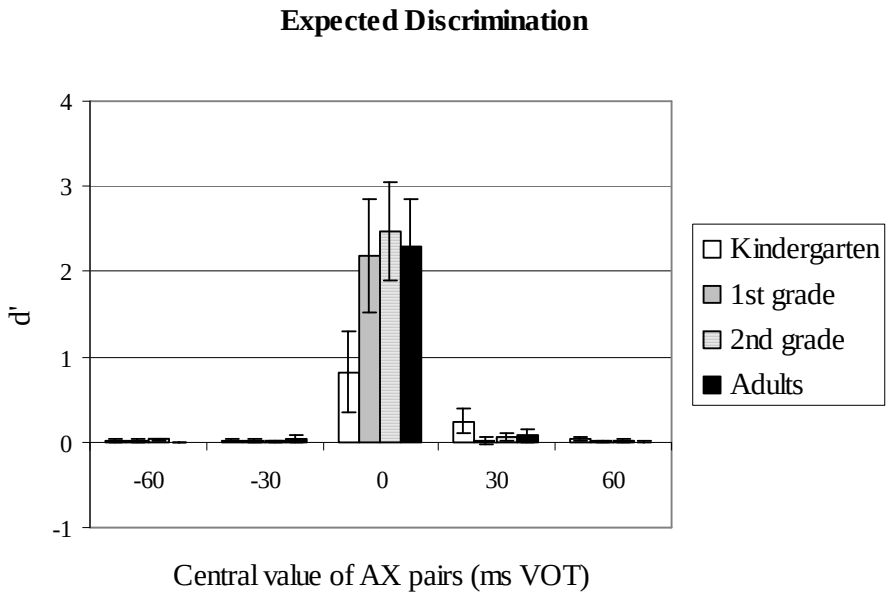

Figure 4 : Expected d' scores on the /də-tə/ VOT continuum for children attending kindergarten, first, second grade of primary school and adults. Error bars represent the standard deviations. The x-axis values represent the central values of AX stimulus pairs (e.g. -60 corresponds to the -75 vs. -45 ms VOT pair).

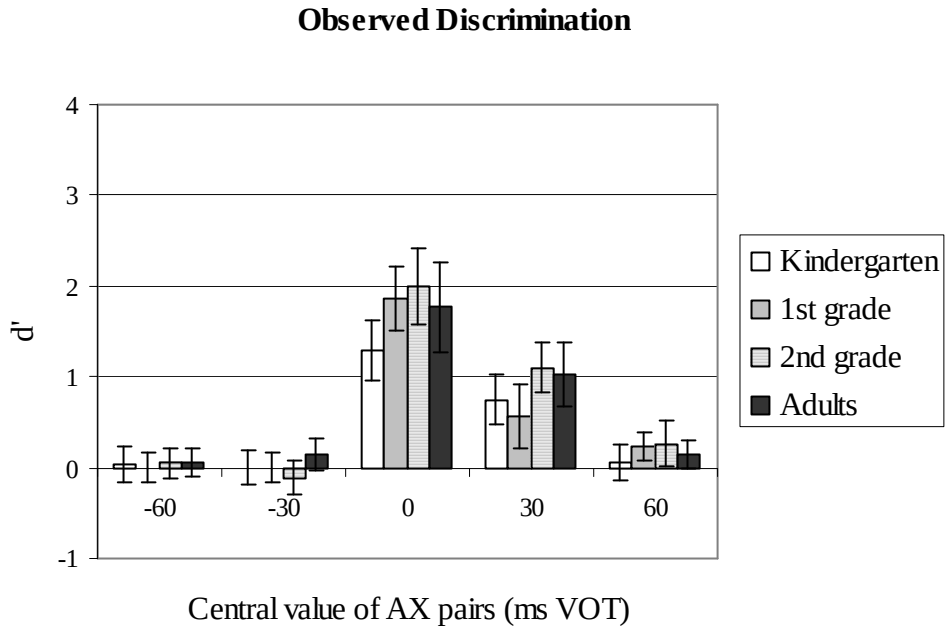

Figure 5: Observed d' scores on the /də-tə/ VOT continuum for children attending kindergarten, first, second grade of primary school and adults. Error bars represent the standard deviations. The x-axis values represent the central values of AX stimulus pairs.

Differences in Categorical Perception between groups were tested in a Task (two levels: observed vs. expected d' scores) x VOT Pairs (five levels: centered on -60, -30, 0, +30, +60 ms VOT) x Group (four levels: kindergarten, 1st grade, 2nd grade, adults) repeated measures ANOVA. Results revealed a main effect of Task ($F(1,47) = 8.34$, $p < .01$; $\eta^2 = 0.15$) which was not modified by Group (Task x Group: $F(3,47) = .18$, $p > .05$; $\eta^2 = 0.01$) but differed according to the VOT Pair (Task x VOT Pair: $F(3,188) = 16.01$, $p < .000$; $\eta^2 = 0.25$). Paired-sample t-tests with Bonferroni corrected p values of significance ($p = .01$, i.e. the 0.05 p value of significance was divided by five, the number of contrasts) between expected and observed d' scores revealed that the difference between the expected and observed d' scores was significant for the VOT pair centered on +30 ms ($t(50) = -8.92$, $p < .001$) and to a lesser extent for the VOT pair centered on +60 ms ($t(50) = -2.76$, $p < .01$). Differences between expected and observed d' scores were not significant for the other pairs ($t(50) = -.54$, $p = .59$; $t(50) = -.12$, $p = .9$ and $t(50) = 1.65$, $p = .11$, for the pairs centered on -60, -30 and 0 ms VOT respectively. The consistent differences between the expected and observed d' scores showed that categorical perception was not perfect. Moreover, the fact that these differences were

only significant for the pairs located in the positive VOT region indicated an asymmetry in perceptual processing between the two sides of the VOT continuum.

CORRELATIONS BETWEEN CP AND READING LEVEL

Table 1 shows Pearson coefficients obtained when correlating scores at the EVIP vocabulary test, correctness and rapidity scores at the Odédys reading test, Boundary Precision and Categorical Perception measures. Boundary Precision measure was indexed by the steepness of the slope computed for each subjects (in logit/ms). Categorical Perception measures corresponded to the matching between d' expected and observed scores, i.e. the difference between expected d' scores and observed d' scores for each of the five VOT pairs.

	EVIP	Odédys					
		Regular Words		Irregular Words		Non Words	
		Corr.	Speed	Corr.	Speed	Corr.	Speed
Boundary Precision	-.33	-.31	.42	-.30	.60*	-.36	.46*
Categorical Perception (d' Exp - d' Obs)							
-60	.17	-.03	-.02	.10	-.11	.14	.01
-30	.16	-.17	.09	-.05	.09	.02	.11
0	.31	.14	-.09	.11	-.24	.21	-.20
+30	-.35	-.24	.19	-.19	.18	-.31	.34
+60	-.04	-.27	-.20	.31	-.22	.41	-.18

Table 1: Pearson correlations between EVIP vocabulary test, Odédys reading test (correctness and rapidity) and categorical perception measures (Boundary Precision, i.e. steepness of the slope and Categorical Perception, i.e. difference between expected and obtained d' for the -60, -30, 0, +30 and +60 ms central VOT pairs). Significant correlations are flagged by a star.

From these results, it appeared that neither the vocabulary measures, nor the reading measures, were correlated to Categorical Perception index. However, reading measures were significantly correlated to Boundary Precision at least for irregular (r = .60) and non words (r = .46) reading speed. These positive correlations meant that the quicker the subject read the

steeper the identification slope was. The absence of correlation between Boundary Precision and reading correctness might come from the strategy privileged by the majority of the subject, i.e. to read quickly rather than correctly. From these results it appeared that reading is correlated with some measures of Boundary Precision but not with Categorical Perception.

EXPERIMENT 2

Method

PARTICIPANTS

Four age groups composed of 40 children and 10 adults participated in this experiment. None of these children were part of the first experiment sample. Due to poor instructions comprehension, data of five subjects, one attending kindergarten and four attending the second grade of primary school were excluded. The final set comprised 13 children (six boys and seven girls) attending last year of kindergarten (mean age: six years two months, SD: five months), nine children (three boys and six girls) attending first grade of primary school (mean age: seven years one month, SD: three months), 13 children (six boys and seven girls) attending second grade of primary school (mean age: seven years 11 months, SD: five months) and 10 adults (three boys and seven girls ; mean age: 28 years five months, SD: five months). All subjects were native speakers of French, non bilingual and reported no history of auditory and language disorders. For children, parents gave their written consents.

STIMULI

Colors: Six Munsell color samples of two hues, yellow and green, equated for brightness (Value = 9) and saturation (Chroma = 10) were used. The Munsell color system was chosen because values of hue and chroma are perceptually evenly spaced. The six stimuli designated in Munsell color nomenclature were 3.75Y ; 8.75Y ; 3.75GY ; 8.75GY ; 3.75G and 8.75G. The color samples were presented in a 13 x 22 square at a 50 cm distance on a computer screen (Philips Screen, 107S, 15 inch controlled by a Laptop, Dell, Inspiron 1300 for children and Sony screen, Multiscan E230, 15 inch controlled by an Intel Pentium 4 for adults) calibrated with a colorimeter (Spyder2 Express Color Vision, Datacolor, 2006 and Eye-one for children, XRirte for adults). Stimuli were presented at a 50 cm distance.

*Facial expression*s: two black and white identities expressing happiness and sadness (Montreal Set of Facial Displays of Emotion, Beaupré & Hess, 2005) were taken as the endpoints of the continuum (100% Happy (H) – 100% Sad (S)). These identities represented a male full face oriented. The four intermediate stimuli were created by morphing, i.e. by interpolation between happiness and sadness (Morph Man 2000, STOIK software) with a 20% gap between the six stimuli (100% H ; 80%H-20%S ; 60%H-40%S ; 40%H-60%S ; 20%H-80%S ; 100%S).

Similar to the /də-tə/ VOT continuum, six stimuli were created for the green-yellow and happy-sad continua. Results of an unpublished pilot study, conducted with green-yellow and happy-sad stimuli, showed that adults' identification boundary was located between the third and the fourth stimulus for both continua and that adults AX discrimination scores peaked for the contrast straddling this boundary.

PROCEDURE

Children were individually tested in a quiet and dark room in their school during two sessions (lasting together one hour) whereas adults were tested in a single 30 minutes session.

Visual screening: all subjects were screened for color deficiency with the Ishiara test (1968).

Identification: A practice phase using the continuum endpoints (3.75Y/8.75G for colors and 100% happy/100% sad for facial expressions, all presented 10 times) was played in order to familiarize the child with the procedure. Presentation was controlled by E-prime 1.1. (Psychology Software Tools). Subjects were asked to associate the stimulus with one of the stickers (green/yellow for colors and happy/sad smiley for facial expressions) pasted on the computer keyboard. A correct response was reinforced by visual feedback. When 70% of the responses were correct, subjects were admitted to the identification test. During the identification test, each of the six stimuli of the continuum was presented 10 times. So as to attract the attention of the subject to the center of the screen, stimulus presentation began with the apparition of a cross which lasted 500 ms. Then, a black screen appeared during 400 ms, followed by the stimulus lasting 400 ms (figure 6). The inter-stimulus interval was controlled by the subject, his response entailing the next stimulus presentation. The order of presentation

between the color and the facial expression identification tasks was counterbalanced across subjects. Data analysis was identical to the one described for the /də-tə/ VOT continuum.

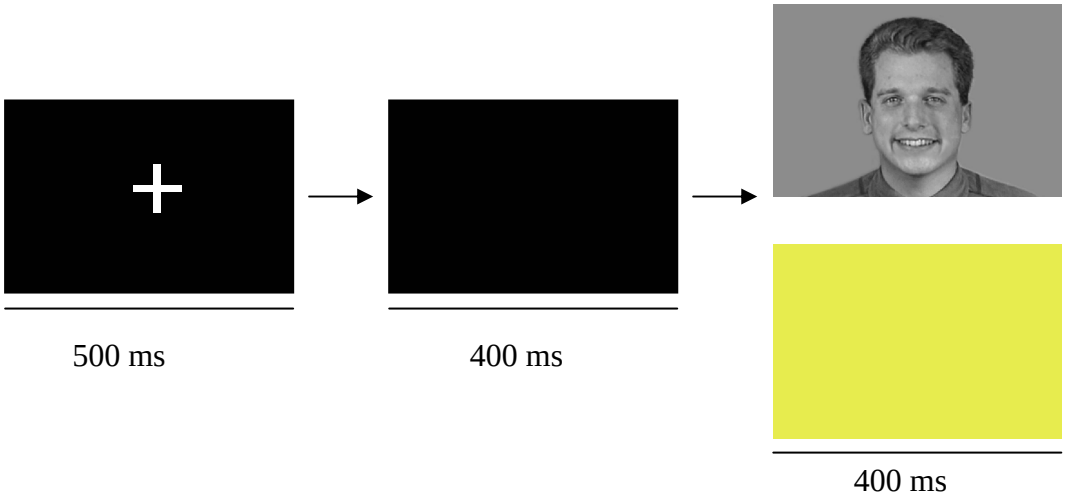

Figure 6: Identification task in Experiment 2 (with an example for color and for facial expression).

Discrimination: the discrimination task was preceded by a practice phase in which the endpoints pairs were presented five times. Subjects were asked to associate the pair of stimuli with one of the two pictures pasted on the keyboard. The picture associated to the "same" response represented the same colors or facial expressions whereas the picture associated to the "different" response represented colors or facial expressions belonging to different categories. For the discrimination test, each stimulus was presented paired with itself (i.e. six identical color pairs (C): C1/C1 ; C2/C2 ; C3/C3 ; C4/C4 ; C5/C5 ; C6/C6 or six identical facial expression pairs (F): F1/F1 ; F2/F2 ; F3/F3 ; F4/F4 ; F5/F5 ; F6/F6) and with the adjacent stimulus (ten different pairs, i.e. five stimulus combinations (C1/C2 ; C2/C3 ; C3/C4 ; C4/C5 ; C5/C6) x two orders for colors and five stimulus combinations (F1/F2 ; F2/F3 ; F3/F4 ; F4/F5 ; F5/F6) x two orders for facial expressions). The experimental protocol for the discrimination task is presented in figure 7. The order of presentation between the color and the facial expression discrimination tasks was counterbalanced across subjects.

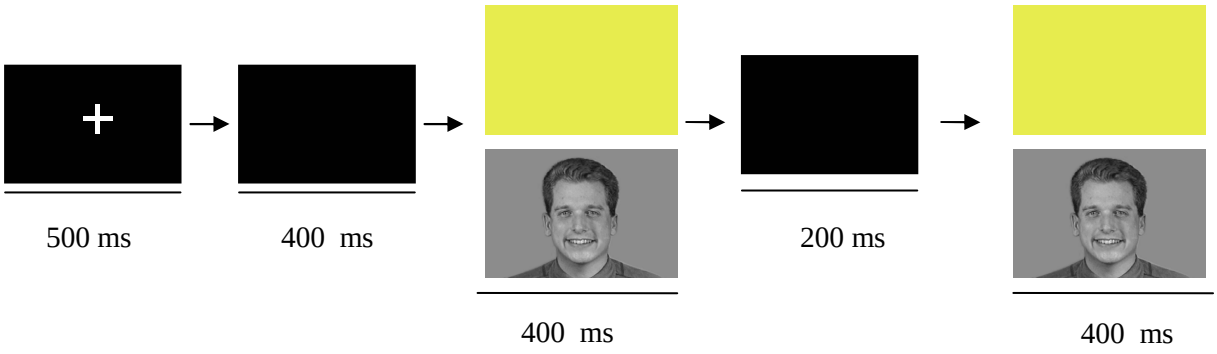

Figure 7: Discrimination task in Experiment 2 (with an example for color and for facial expression).

Results

IDENTIFICATION

Colors: The mean function slope for the four groups of subjects is presented in figure 8. As it can be seen on the figure, identification slopes were quite similar for the different groups and centered between the third and the fourth stimulus.

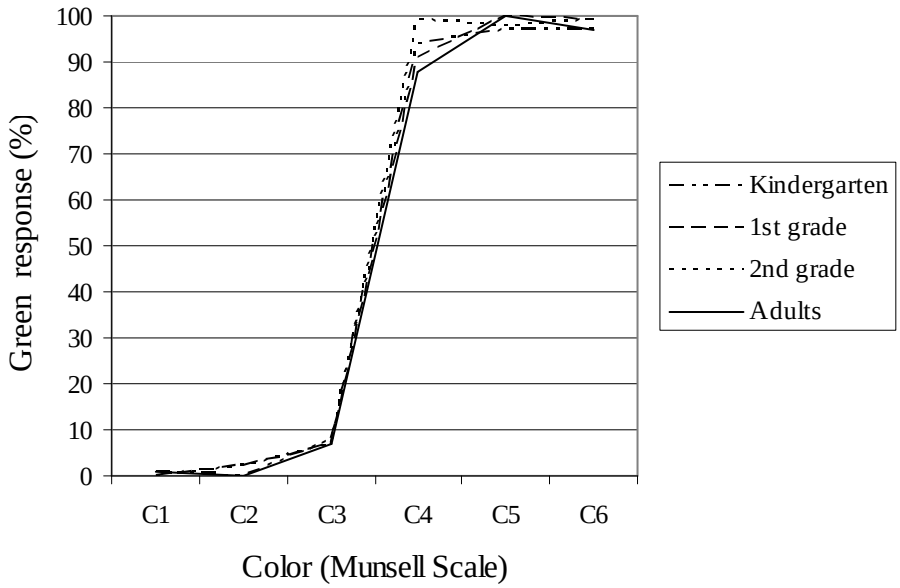

Figure 8: Identification function on the yellow-green continuum for children attending kindergarten, first, second of primary school and adults.

Differences in mean boundaries and slopes between the groups were tested by univariate ANOVAs. The mean boundary was not significantly related to the groups ($F(3,44) = .58$, $p = .64$; $\eta^2 = 0.04$) and was located just between the third and the fourth stimulus when taken the five groups together (mean = 3.46 ; SD = 0.29).

The Slope x Group interaction was not significant ($F < 1$), showing this way no improvement in Boundary Precision across years.

Facial Expressions: The mean function slope for the four groups of subjects is presented on figure 9.

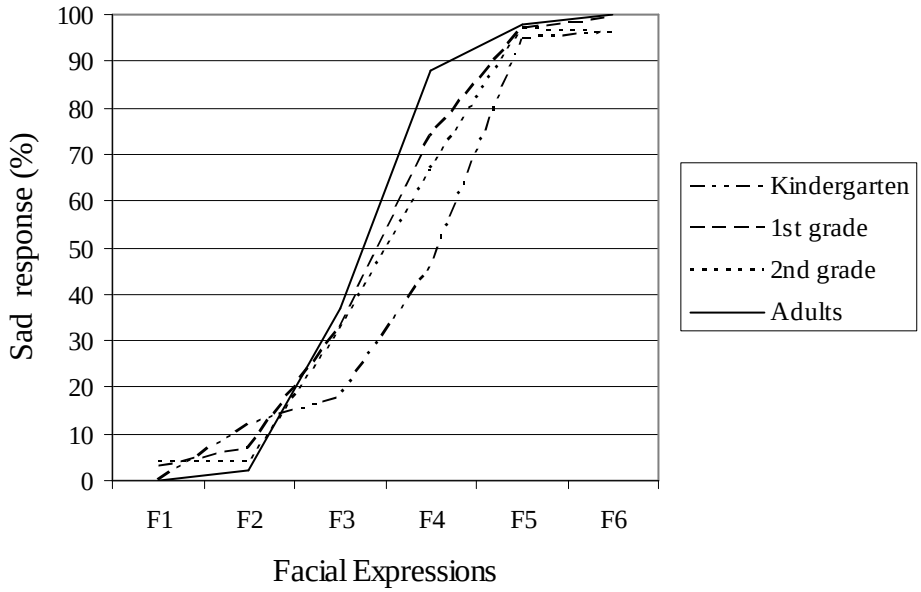

Figure 9: Identification function on the happy-sad continuum for children attending kindergarten, first, second grade of primary school and adults.

Differences in mean boundaries and slopes between the groups were tested by univariate ANOVAs. The mean boundary did not differ according to the Group ($F(3,44) = 1.72$, $p = .18$; $\eta^2 = 0.11$) and was located just between the third and the fourth stimulus when taken the five groups together (mean = 3.51 ; SD = 0.7).

Boundary Precision was assessed by the Slope x Group interaction. The slope was significantly related to the groups ($F(3,44) = 6.34$, $p < .01$; $\eta^2 = 0.32$), the mean slope being steeper for the adults (0.39 logit/ms) than for the second grade (0.17 logit/ms) and for the first grade (0.08 logit/ms), this latter being identical for kindergarten children. Posthocs comparisons revealed that the differences in slope between the adult group and each of the other groups were significant (adult/kindergarten: $F(1,41) = 15.64$, $p < .01$; adult/1st grade: $F(1,41) = 13.04$, $p < .01$ and adult/2nd grade: $F(1,41) = 7.97$, $p < .05$). These results showed that Boundary Precision for the happy-sad continuum remains stable during childhood (at least until the 2nd grade) and improves later on.

DISCRIMINATION

Colors: The expected and observed discrimination scores for the five groups are presented in figures 10 and 11. Graphically, we observed that the maximal d' scores were reached for the AX inter-categorical color pair (C3C4).

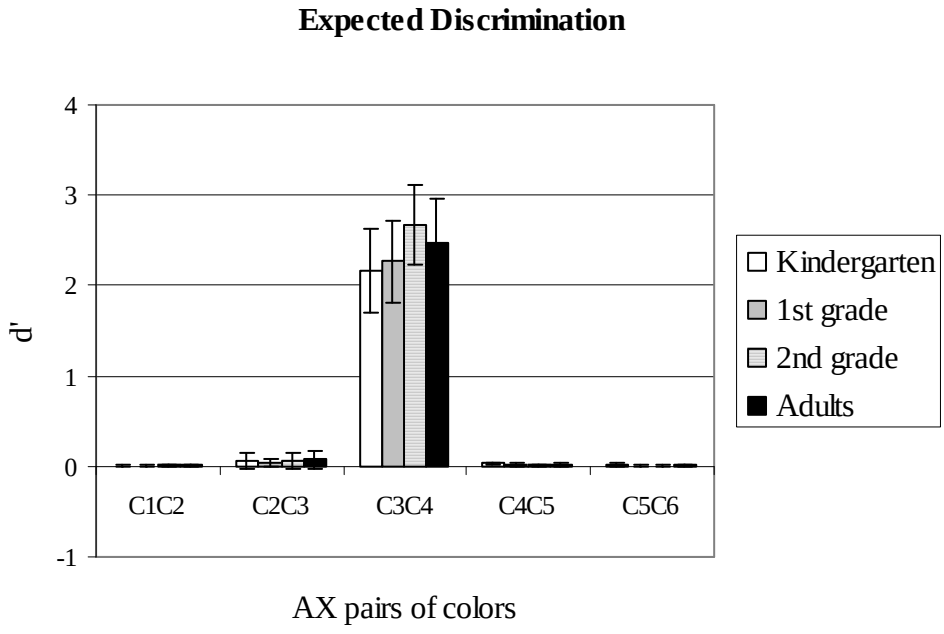

Figure 10: Expected d' scores on the yellow-green continuum for children attending kindergarten, first, second grade of primary school and adults. Error bars represent the standard deviations. The x-axis values represent the AX pairs of colors.

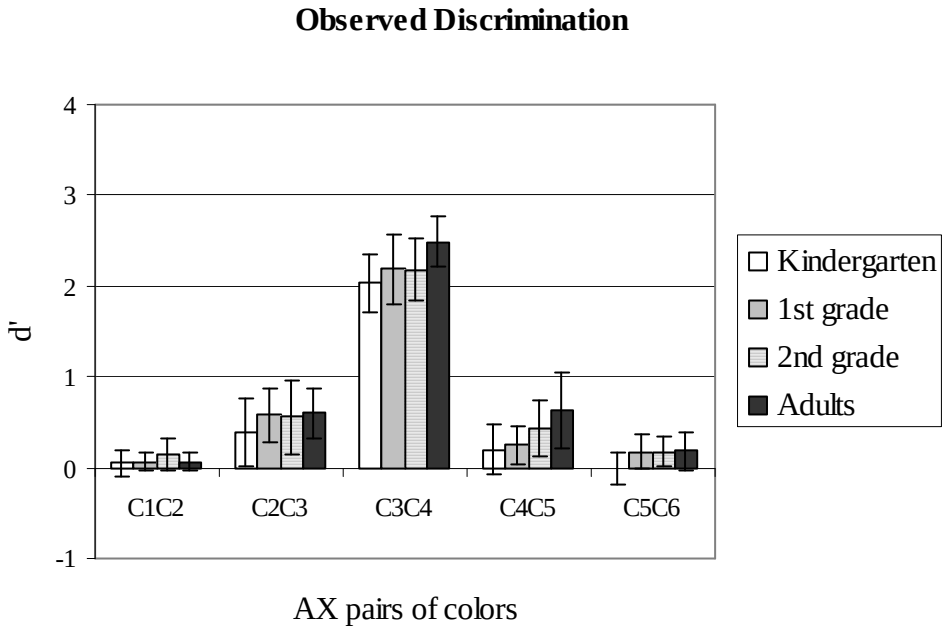

Figure 11: Observed d' scores on the yellow-green continuum for children attending kindergarten, first, second grade of primary school and adults. Error bars represent the standard deviations. The x-axis values represent the AX pairs of colors.

Categorical Perception was tested in a Task (two levels: observed vs. expected d' scores) x Color Pairs (five levels: C1C2, C2C3, C3C4, C4C5, C5C6) x Group (four levels: kindergarten, 1st grade, 2nd grade, adults) repeated measures ANOVA. Results revealed a main effect of Task ($F(1,41) = 19.72$, $p < .001$; $\eta^2 = 0.33$) which was not modified by Group (Task x Group: $F(3,41) = 1.21$, $p > .05$; $\eta^2 = 0.08$) but differs according to the Color Pair (Task x Color Pair: $F(4,164) = 7.55$, $p < .001$; $\eta^2 = 0.16$). Paired-sample t-tests with Bonferroni corrected p values of significance (p threshold of significance = .01) revealed that the difference between expected and observed d' scores was significant for the C2C3 and C4C5 intra-categorical pairs ($t(44) = 3.76$, $p < .001$; 4.80, $p < .001$, respectively) but not for the C3C4 inter-categorical pair ($t(44) = 1.51$, $p > .01$).

Facial Expressions: The expected and observed discrimination scores for the four groups are presented in figures 12 and 13. Graphically, one can note a pattern of response different from the two preceding continua with much more variability.

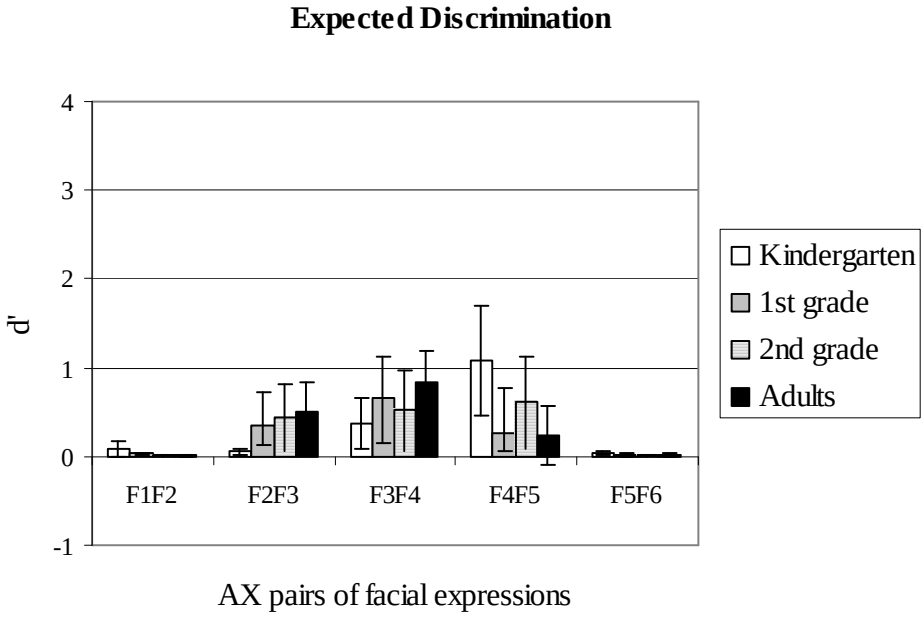

Figure 12: Expected d' scores on the happy-sad continuum for children attending kindergarten, first, second grade of primary school and adults. Error bars represent the standard deviations. The x-axis values represent the AX pairs of facial expressions.

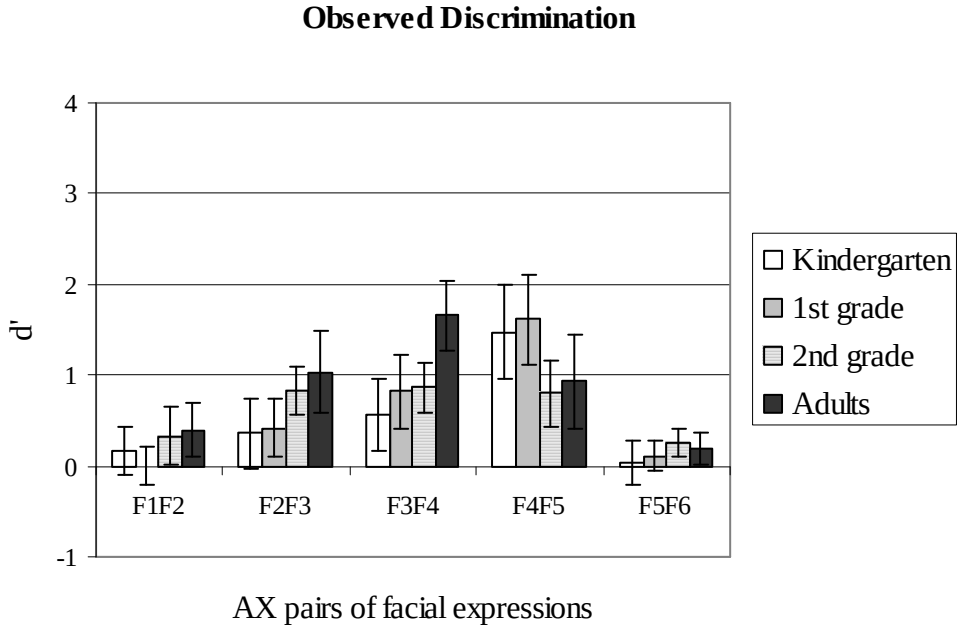

Figure 13: Observed d' scores on the happy-sad continuum for children attending kindergarten, first, second grade of primary school and adults. Error bars represent the standard deviations. The x-axis values represent the AX pairs of facial expressions.

Categorical Perception differences between groups were tested in a Task (two levels: observed vs. expected d' scores) x Facial expression Pairs (five levels: F1F2, F2F3, F3F4, F4F5, F5F6) x Group (four levels: kindergarten, 1st grade, 2nd grade, adults) repeated measures ANOVA. Results revealed a main effect of Task ($F(1,41) = 37.56$, $p < .001$; $\eta^2 = 0.48$) which was not modified by Group (Task x Group: $F(4,41) = 1.59$, $p > .05$; $\eta^2 = 0.10$) but differs according to the Facial expression Pair (Task x Facial expression Pair: $F(4,164) = 4.08$, $p < .01$; $\eta^2 = 0.09$). Paired-sample t-tests with Bonferroni corrected p values of significance (p threshold of significance = .01) revealed that the difference between the expected and observed d' scores were significant for F2F3 and F4F5 intra-category pairs ($t(44) = 3.39$, $p = .001$; 3.74, $p = .001$, respectively), and for F3F4 inter-category pair ($t(44) = -3.11$, $p < .01$).

Discussion

The aim of the present study was to investigate the development of categorical perception in French-speaking adults and children from the last year of kindergarten to the second grade of primary school. More particularly, our aim was to shed some light on the determinants of categorical perception development. One objective was to see whether perceptual development around six years of age arises from improvements of Categorical Perception per se, i.e. the dependency of discrimination on identification, or instead from a greater Boundary Precision. Another objective was to see whether changes in perceptual categorization during this period were related to reading, as suggested by Burnham (Burnham, Earnshaw & Clark, 1991; Burnham, 2003), rather than to the general development of cognitive abilities (Karmiloff-Smith, 1991; Lalonde & Werker, 1995).

THE DEVELOPMENT OF CATEGORICAL PERCEPTION

For sake of clarity, Table 2 summarizes the results obtained for speech and visual continua.

	Identification		Discrimination
	Boundary location	Boundary Precision	Categorical Perception
VOT continuum	Between the third and the fourth stimulus for all groups	Becoming steeper between kindergarden and adulthood	Difference between expected and observed discrimination, especially for the +30 central pair
Colors continuum	Between the third and the fourth stimulus for all groups	Mature in kindergarden	Difference between expected and observed discrimination, for C2C3 and C4C5 color pairs
Facial expressions continuum	Between the third and the fourth stimulus for all groups	Becoming steeper between kindergarden and adulthood	Difference between expected and observed discrimination, for F2F3, F3F4 and F4F5 facial expression pairs

Table 2: Main identification and discrimination results obtained for the VOT continuum (Experiment 1) and the color and facial expression continua (Experiment 2).

Results of Experiment 1, reinforced by those of a companion study (Medina, Hoonhorst, Bogliotti, Sprenger-Charolles and Serniclaes, submitted) suggested that Categorical Perception does not change after some six years of age. Performances of six to eight-year-old children were similar to those of adults. These results contrast with important changes in Boundary Precision between children and adults. This last point, i.e. the late maturation of Boundary Precision, has been emphasized by Zlatin and Koenigsknecht (1975) who noticed that “the magnitude of VOT difference required for distinguishing between prevocalic stop cognates decreases as a function of the age of the listeners” (pp. 541). Raskin, Maital and Bornstein (1983) reached the same conclusion after analyzing the development of Boundary Precision in three-to-four-year-old children and adults presented with color continua. They stated that “older observers judge category boundaries to be sharper than do younger observers, and older observers judge the limits of color categories with greater circumscription” (pp. 140). Taken together, these results provide different examples of two independent phenomena: the stability of the effect of categorization on discrimination (i.e. Categorical Perception) on the one hand, and the gradual maturation of identification performances on the other hand. In our view, the fact that the identification of categories get more precise with age would stem from a gradual maturation of low-level processing whereas

higher-order integrative mechanisms responsible for the matching between identification and discrimination, i.e. the control of language-specific categories, would be mature earlier in time (not later than six years of age as suggested by the results of the present study).

Results of Experiment 2 evidenced both similitudes and differences in the developmental patterns of categorical perception maturation across the different continua. Common to all the continua is the fact that Categorical Perception does not significantly change as a function of age. Only differences in Boundary Precision were found for two among the three continua under scope, i.e. for the voicing and facial expression continua but not for the color continuum. Although the two first continua pertain to different sensory modalities, auditory or visual, an important common point is that both are fairly complex when compared to the color continuum. Voicing perception depends on several different acoustic cues (e.g. Stevens & Klatt, 1974), and at least one of these cues varied along the synthetic VOT continuum used in the present study (F1 frequency at voicing onset). Similarly, the perception of facial expressions depends on different visual cues (Bruyer, Granato & Van Gansberghe, 2007), which also covaried along the continuum under use which was generated by morphing between two different pictures of natural facial expressions. Instead, the color continuum was much simpler as it was generated only by changes in hue, being acknowledged that color perception depends on several other cues in natural conditions (Franklin & Davies, 2004). The fact that age affected the perception of the two complex continua (voicing and facial expressions) and not the simple one (color) suggests that the late development of Boundary Precision is somehow related to the integration of different physical cues. We propose that the complexity of voicing and facial expression contrasts arises from the fact they are multiple-cued.

Learning to perceive such contrasts means integrating different cues each having a specific boundary, distinct from the one characterizing the categorical contrast. The challenge for the learning processes is to consolidate the categorical boundary, in spite of the persisting discrimination of individual cues. The struggle between progress in the consolidation of the labeling boundary and the discrimination of individual cue boundaries is seemingly at the origin of the deficit in Categorical Perception in the present results. For the /də-tə/ continuum the mismatch was particularly evident for the +30 ms VOT contrast. Discrimination of the +30 ms positive VOT contrast was fairly large, better than what was expected from the

identification data, even though this contrast was clearly apart from the French VOT boundary which is located around 0 ms VOT (at some +7 ms in the present study). The increased discrimination of VOT differences around +30 ms might be due to concomitant changes in another voicing cue, namely F1 transition (Stevens & Klatt, 1974). Examination of figure 2 shows an important qualitative difference between the +15 and +45 ms VOT stimuli: the F1 transition is still present at +15 ms but absent at +45 ms.

The difference between identification and discrimination tasks was also observed for the happy-sad continuum. Whereas the categorical boundary was located between the third and the fourth stimulus, the discrimination peak was centered between the fourth and the fifth stimulus until the second grade of primary school. As for voicing, multiple cues participate to facial expression perception. Since children rely mostly on facial features (e.g. mouth, nose) to perceive faces, their perception could be influenced by the interaction between the whole set of facial features whereas adults would be less influenced because they rely mostly on configural processing (Mondloch, Le Grand & Maurer, 2002)[2].

The next question is whether the late development of speech perception is piloted by reading instruction or is instead a specific instance of a general improvement.

THE READING HYPOTHESIS

Results of Experiment 1 evidenced a correlation between the rapidity to read irregular and non-words and Boundary Precision but no correlation between reading abilities and Categorical Perception. These data reinforce the conclusion drawn by Serniclaes et al. (2005) who showed no link between Categorical Perception and literacy but rather a link between Boundary Precision and literacy when comparing literate and illiterate voicing perception performances. Therefore and contrary to Burnham et al. (1991) and Burnham (2003), our data did not evidence a direct relationship between the onset of reading and Categorical Perception, since there was no correlation between reading abilities and the correspondence between identification and discrimination scores. Rather, our results demonstrate a link between the onset of reading and the precision to label voiced and voiceless phonemes. While the correlation between Boundary Precision and literacy suggests some link between reading

2 As explained by Schwarzer (2000, pp. 391-392), the term ‘configural’ refers to “the specific spatial relationships among the individual facial features, especially the second-order relationships which include the spatial locations of the facial parts relative to the prototypical arrangement of the eyes, nose and mouth".

instruction and the development of speech categorization, the present results for the non-linguistic continua entailing the visual modality are crucial to confirm the reading hypothesis. The fact that facial expression perception develops in much the same way as voicing perception suggests a general determination, independent of reading. In this view, the correlation between reading instruction and Boundary Precision evidenced in Experiment 1 would not arise from an effect of reading instruction on speech perception development but rather from an effect of speech perception on reading acquisition.

THE GENERAL COGNITIVE HYPOTHESIS

The similar development of categorization for a speech and a non-speech continuum is reminiscent of Lalonde and Werker (1995) results. Since they evidenced in eight-to-10-month-old babies a synchronous change in a native vs. non native linguistic discrimination task, a visual categorization task and an object search task, they concluded that speech perception development is underlined by general cognitive competences. Although the present results concern a much later period of age, between six years of age and adulthood, they also point to a general cognitive development and they give some indications on its nature. The present study suggests that the late development of categorization is characterized by improvements in the precision of perceptual boundaries whereas Categorical Perception (i.e. the relationship between identification and discrimination) remains fairly constant. Further, the changes in Boundary Precision which take place during this period are seemingly due to a better integration of the cues which contribute to same categorical distinction. Similar considerations might also apply to the early development of categorization performances, before one year of age. Different studies indicate that the early exposure to the native language gives rise to the integration between different acoustic cues pertaining to the same phonological feature (positive VOT and F1: Miller & Eimas, 1983; negative and positive VOT: Hoonhorst, Colin, Markessis, Radeau, Deltenre, & Serniclaes, 2009).

Taken together, the present results on the late development of perceptual categorization, between six years of age and adulthood, and those collected before one year of age suggest a long-standing integrative process of physical cues which are basically different in nature. While these cues are integrated in different ways for recognizing different categories, they are still processed independently up to some level, as evidenced by the within-category discrimination peaks in the present results. This might well explain why the buid-up of

complex categories, such as those involved in speech and face perception, extends over such a long period of time.

Conclusion

These experiments were designed to get some insight into the determinants of categorical perception development. Results obtained with the same methodology and for three different continua did not evidence a single pattern of development but rather a differential development for the color continuum compared to the voicing and facial expression continua. For the former, both Boundary Precision and Categorical Perception were mature at the end of kindergarten whereas for the two latter Boundary Precision continues to develop between kindergarten and adulthood. We propose that this differential pattern of development comes from a difference in the structural complexity of the categories, the perception of facial expressions and voicing involving the integration of multiple cues. We also propose that the change across years in Boundary Precision and not in Categorical Perception stem from a gradual maturation of low-level processes.

Etude 3.1

The N100 component:

An electrophysiological cue of voicing perception[1]

In many languages the voicing distinction between stop consonants in initial position depends on laryngeal timing. The time interval between closure release and voicing onset (Voice Onset Time) has been extensively investigated since it is the major acoustic correlate of voicing. This chapter reviews the literature on the neurophysiological correlates of voicing perception with an emphasis on the basic auditory sensitivity common to human newborns and animals. Since it has been shown that the acoustic level of perception is indexed by late cortical evoked-potentials, we recorded the N100 component in French-speaking subjects, for whom the phonological boundary corresponds to none of the acoustic ones. We hypothesized that in French, the basic acoustic boundaries should be revealed by the N100 characteristics without any contamination by the phonological one. This hypothesis was confirmed by the results, which open the way to a better understanding of the link between the acoustic and the phonological levels of speech perception.

[1] Hoonhorst I, Colin C, Markessis E, Radeau M, Deltenre P, Serniclaes, W. N100 component: an electrophysiological cue of voicing perception. In: Fuchs S, Loevenbruck H, Pape D, Perrier P, editors. Some aspects of speech and the brain. Bern: Perter Lang Verlagsgruppe, 2009:5-34.

Introduction

VOICING

In 1977, Abramson opened the way to a wide field of research by defining Voice Onset Time (VOT) as the *"temporal relation between the onset of glottal pulsing and the release of the initial stop consonant" (p. 296)* and by underlining the importance of the temporal order between these two events, i.e. the temporal order between voicing onset and closure release. Voicing lead was used to characterize productions in which periodic energy precedes closure release, while voicing lag described the inverse pattern.

The huge number of subsequent studies conducted on VOT can certainly be explained by the peculiarly straightforward relation between the perception of voicing and VOT, which is the most important cue for voicing perception, at least in initial position (Lisker et al., 1978). In medial position, it is combined with other temporal cues related to laryngeal timing (Saerens et al., 1989). These other acoustic cues, such as the value of the fundamental frequency, formant transition duration and the frequency value of F1, only affect voicing perception when the VOT is ambiguous and they are therefore considered as secondary cues (Summerfield & Haggard, 1977; Hillenbrand, 1984).

There are multiple reports showing that the VOT perceptual boundary shifts as a function of place of articulation (Abramson & Lisker, 1973; Sharma et al., 2000) and vocalic context (Summerfield & Haggard, 1974). This provides supplementary degrees of freedom for the design of experiments on voicing perception in various contexts.

TWO OR THREE LEVELS OF SPEECH PERCEPTION?

Based on their empirical findings showing an effect of inter-stimulus-interval (ISI) on discrimination performances, Werker and Logan (1985), followed by other authors (Burnham, 1986; Flege, 1988; 1992), developed a three-factor model of speech perception in which the acoustic, phonetic and phonological levels of speech processing are dissociated. They showed a correlation between the ISI and the complexity of the level of perception reached: ISIs of 250, 500 and 1500 ms were respectively associated with the acoustic, phonetic and phonological levels of perception. The acoustic level was related to the ability to discriminate acoustic contrasts that are not phonological in any language of the world; the phonetic level was tied to the ability to discriminate contrasts which are phonological in languages not

spoken by the subject, and the phonological level was linked to the capacity to discriminate phonemes belonging to his/her own language. The key finding of this study was the demonstration that the experimental setting may influence the subject's discrimination and lead him/her to perceive contrasts that are phonological neither in his/her language nor in any of the other languages.

This study raised two main issues in the field of the development of perceptual abilities. First, by underlining the impact of experimental conditions on the ability to discriminate fine-grained contrasts, one must accept the existence of successive levels of speech processing which we may expect to correspond to a hierarchical functional organization of the perceptual system. Secondly, even though Werker and Logan (ibid) postulated a phonetic level of perception, they did not specify its nature and cautiously concluded *"it is not clear whether phonetically relevant perception is a function of a specific linguistic processor or the result of second-order auditory factors […]" (p. 43).*

The suggestion of an intermediary phonetic mode of perception comes from the observation that human perceptual boundaries are built up as a combination of trade-off between acoustic cues, which Werker and Logan (1985) called *"second-order auditory factors"*. Indeed, some authors have shown that categories of voicing come from the combination of VOT and first formant (F1) transition (Simon & Fourcin, 1978) or that categories for place of articulation arise from the second and third formant (F2 and F3) combination. Since the resolution of this debate remains unclear, we will use a simplified scheme with two levels of speech perception: an acoustical and a phonological one. The acoustic level is characterized as non-speech specific and is common to non-human animals and human babies before six months of age. As for the phonological level of perception, it is both speech- and language-specific.

Combined results gathered with diverse methods (Multi-Unit-Activity, Current Source Density, intra-cortical and scalp recordings of cortical activity) have provided evidence of the neurophysiological correlates of these two levels of voicing perception. Single-unit activity recordings in the cat auditory nerve (Sinex & McDonald, 1988; 1989), in the inferior colliculus (Chen et al., 1996; Chen & Sinex, 1999) and in the primary auditory cortex of monkeys (Steinschneider et al., 2003) and of humans (Liégeois-Chauvel et al., 1999) have provided some support for a hierarchical functional organization of the perceptual system.

Behavioural experiments have demonstrated that animals (Kuhl & Miller, 1975; 1978) and infants below six months of age raised in diverse linguistic backgrounds (Eimas et al., 1971; Lasky et al., 1975) share common voicing categories delineated by -30 ms and +30 ms VOT

boundaries (Abramson & Lisker, 1970). By analogy with the "language-universal" categories of Miller and Eimas (1996), defined as the initial categories according to which the initial acoustic space is partitioned, we will use "universal boundaries" throughout the chapter to refer to the initial boundaries (-30 and +30 ms VOT) that delineate the voicing continuum, i.e. the acoustic boundaries to which animals and infants before six months of age are sensitive.
Other behavioural and electrophysiological studies have found evidence of the maturation from a universal and acoustic mode to a phonological mode of perception (e.g. Werker & Tees, 1984; Hoonhorst et al., in press) according to rules of the mother tongue. Indeed, it has been shown for voicing that the two universal boundaries (-30 and +30 ms) are used in Thai (Donald, 1978), whereas in English (Lisker & Abramson, 1967), only the +30 ms VOT boundary remains relevant. In French, Spanish, Polish, Dutch, Hebrew and Arabic (Serniclaes, 1987; Williams, 1977; Flege & Eefting, 1986; Maassen et al. , 2001; Horev et al., 2007; Yeni-Komshian et al., 1977), none of the universal boundaries are relevant. Indeed, a new phonological boundary located at 0 ms VOT emerges, just midway between the two universal predispositions for perceiving voicing. The emergence of this new boundary centred on 0 ms VOT is the most commonly used mechanism in two-category languages. With its single boundary at +30 ms, English appears as an exception among the languages with two VOT categories. It is nevertheless on English that the vast majority of studies have been conducted.

CATEGORICAL PERCEPTION

Categorical perception (CP), which allows the definition of a finite percept on the basis of an infinite number of acoustic realizations, is the major characteristic of the hierarchical system for speech processing. The study of perceptual boundaries and their mechanisms is therefore at the heart of research on speech perception.

Categorical perception is frequently presented as a transform by which continuous physical variations are encompassed into discrete perceptual categories, i.e. there is an analog-to-digital transformation (Harnad, 1987). Identification and discrimination tasks have been extensively used to find evidence of CP. While identification requires labelling auditory stimuli, discrimination requires the subject to determine whether pairs of stimuli are identical or different. Categorical perception is often presented as the fact that we are only able to discriminate stimuli that do not have the same labelling. However, according to Liberman et al.. (1957), who first defined this phenomenon, CP does not imply an all-or-none capacity but

refers to the observation that the contrasts between sounds separated by the same acoustic distance and belonging to different phoneme categories are more easily discriminated than those inside categories (Liberman et al., 1957). Carney et al. (1977) reinforced this relative nature of the CP phenomenon by emphasizing *"the improved discrimination of stimuli near a category boundary" (p. 969).*

Although the hierarchical organization as well as the categorical mode of speech perception is firmly established, their physiological correlates remain to be better delineated. Determining the link between basic auditory mechanisms and the perception of voicing is the core interest of the research presented below.

NEURAL ENCODING OF VOICING

The comparison between humans and non-human animals suggests that VOT universal boundaries, located at -30 and +30 ms, have developed according to constraints imposed by the auditory system. Several studies have been devoted to the understanding of the physiological correlates underlying these boundaries.

In 1959, Hirsh distinguished two psychoacoustical abilities: one was the minimal delay that must separate two sounds in order for each of them to be perceived independently and the other concerned the minimal delay necessary to allow the determination of their temporal order (i.e. which one appears first). The conclusions of the various experiments conducted with different stimuli presented in Hirsh's study were that while only two ms are sufficient for a subject to distinguish the presence of two sounds, about 20 ms are needed to determine the temporal order of the same two sounds. This difference by a factor of 10 is, according to Hirsh, related to the involvement of two different mechanisms, i.e. *"more central structures for the anatomical and physiological correlates" (p. 767)* must be involved in the second task. This higher threshold of the perceptive system which determines the temporal order between two successive events is not specific to auditory stimuli, since Piéron (1964) reached similar conclusions for the visual modality. Considering that temporal order is the critical cue to perceive the sign of VOT or TOT[2] (Pisoni, 1977) and that at least 20 ms are needed to determine temporal order between two sounds, several studies have sought to find a correlate of the corresponding neural code.

[2] Tone Onset Time is the non-speech analogue of VOT and corresponds to the time elapsed between two tone onsets.

The neural code is defined by Eggermont (2001) as the link between behaviour (discrimination performances, for our purpose) and neural activity. Because by definition discrimination implies two percepts, it is expected that discrimination between two stimuli occurs as soon as their two neural representations are sufficiently different as to evoke different percepts.

Regarding the +30 ms VOT boundary, several studies have suggested that the neural code for VOT values greater than 30 ms is a double-peaked *("double-on")* response time-locked to closure release and voicing onset as opposed to a single-peak *("single-on")* response for shorter VOT values. In this respect, it is worth noting that the majority of the world's languages make the most of this neurophysiological signature by producing voiced and voiceless phonemes with VOT values strategically positioned with respect to the single- vs double-on response boundaries.

The following section presents some of the key results in the field of neural encoding of voicing.

Data on animals

Testing the discrimination abilities of Macaque monkeys on a /bae-dae-gae/ continuum, Kuhl and Padden (1983) showed that they were sensitive to the same boundaries as human beings, demonstrating in this way that neither categorical perception nor the sensitivity to acoustic contrasts are human-specific. Acoustic boundaries were therefore associated to natural psychophysical boundaries (Kuhl & Miller, 1975) and attributed to auditory constraints that could serve as natural markers to shape the perceptive map.

Further evidence of the contribution of universal boundaries to categorical perception was given by Sinex and McDonald in numerous studies (e.g. 1988; 1989) in which they recorded the discharge pattern of the auditory nerve fibre responses in chinchillas. Results of these studies highlighted the role of neural variability in perceiving acoustic categories. Indeed, Sinex and McDonald (ibid) showed that neural representations of voicing in the auditory nerve are non-linear since the variance in the latency of neural responses is far more important for within-category exemplars of a stimulus than for between-category exemplars. This high degree of uncertainty within categories was responsible for weaker discrimination performances.

This non-linear neural encoding of linear acoustic changes was also shown by Chen et al. (1996) in the discharge pattern of the inferior colliculus of chinchillas and by Steinschneider

et al. (1994; 1995) and Steinschneider et al. (2003) in the primary auditory cortex of monkeys. More specifically, Steinschneider et al. (1995) described two patterns of neuron discharge responses when animals were submitted to stimuli varying in VOT values from 0 to 60 ms with 20 ms VOT steps. Stimuli like /da/, with a VOT value shorter than 20 ms, elicited a *"single-on response"* time-locked to closure release whereas /ta/ stimuli elicited a *"double-on response"* time-locked to both closure release and voicing onset. Interestingly, this differential pattern of responses according to the VOT values is the same as that observed by Sinex and McDonald (1988) in the auditory nerve fibres but is different from the one recorded in thalamocortical fibres (Steinschneider et al., 1994). Responses recorded in the primary auditory cortex are indeed characterized by transient time-locked responses on both closure release and voicing onset whereas neural responses in thalamocortical fibres are characterized by a time-locked response on closure release followed by a phase-locked repetitive response during vocal fold vibrations. This led Steinschneider et al. (1994) to state that the transformation of the pattern of response between the thalamus and the auditory cortex, i.e. the accentuation of the transient onset components, is at the root of the voiced-voiceless distinction.

Whatever the relative roles of auditory nerve and cortical mechanisms in the build-up of the single- vs double-on neural code of VOT categories, the fact that it is expressed in the time-locked response pattern of many cortical neurons allows it to be studied by means of non-invasive scalp recordings.

Data on humans

Dealing this time with the perception of human beings, Steinschneider et al. (1999) presented /ba-da-ga/ and /pa-ta-ka/ stimuli to English-speaking epileptic subjects while recording intra-cortical activity (Multi Unit Activity, Current Source Density, Auditory Evoked-Potentials – AEP) from the Heschl's gyrus, the planum temporale and the superior temporal gyrus. Results showed a differential pattern of results according to the VOT value: stimuli characterized by a VOT of 0 or 20 ms elicited a single-on response whereas a VOT of 40, 60 or 80 ms generated a double-on response time-locked to closure release and voicing onset. Results led these authors to suggest that a 20 ms refractory period after closure release is at the root of this differential response pattern. Only stimuli with a VOT longer than at least 20 ms would be able to elicit the second transient response associated with voicing onset. These results stressed the link between the psychoacoustical boundary (+30 ms in English), typically

obtained with identification and discrimination tasks, and the auditory constraints that underlie this sensitivity. Simos et al. (1998a; 1998b) also highlighted this link by showing a non-linear decrease in the N100 magnetic response amplitude when subjects were presented with TOT stimuli with values increasing from +20 to +40 ms. Simos et al. (1998) showed the same positive correlation between the amplitude of evoked N100 magnetic field (N100m) and VOT values with /ga-ka/ stimuli.

Regarding the representation of voicing in the auditory cortex, Steinschneider et al. (1999; 2005) showed that categorical perception of voicing relies in particular on primary and secondary auditory cortical fields. They found evidence that the anterior portion of Heschl's gyrus, which corresponds to the primary auditory cortex, is more specialized than the posterior portion in the representation of the two transient events of VOT, i.e. closure release and voicing onset. More specifically, electrodes localized in the lateral sites of the anterior portion of Heschl's gyrus provided larger responses time-locked to voicing onset relative to responses time-locked to closure release than electrodes localized in central and medial sites. Using surface recorded cerebral evoked magnetic fields, Simos et al. (1998a; b) demonstrated a more medial localization of the N100m for stimuli with VOT values of 40 and 60 ms than for those with 0 and 20 ms VOT values. Although they used quite different methods, these two studies suggest that the detailed representation of voicing is different between different cortical regions.

As far as hemispheric lateralization is concerned, Liégeois-Chauvel et al. (1999) demonstrated the specialization of the left Heschl's gyrus for processing transient acoustic information contained in voiced and voiceless syllables in human beings while general temporal processing in animals is treated bilaterally in the primary auditory cortex. Trébuchon-Da Fonséca et al. (2005) reached the same conclusions with /ba-pa/ stimuli. This left hemispheric dominance for the processing of VOT was, however, not supported by Steinschneider et al. (2005), who obtained the same results when recording neural responses in the right Heschl's gyrus. Given the possible role of the preexisting epileptogenic lesions in the subjects participating in the latter two studies, one must remain cautious before drawing definite conclusions on lateralization of the VOT-related double-on response.

Concerning the correspondence between intra-cortical recordings and scalp recordings, Trébuchon-Da Fonséca et al. (2005) compared both methodologies and obtained comparable results. As different locations in the Heschl's gyrus respond differently to the transient events contained in voiced and voiceless stimuli, one must be prepared when performing scalp

recordings to obtain, as stated by Steinschneider et al. (2003), *"a composite wave that reflects activation of multiple auditory cortical fields, each with its own capacity to follow temporal features of complex sounds" (p. 318).*

Sharma and Dorman (1999) recorded scalp auditory evoked-potentials and found a single-on response with short VOT values (perceived as /da/) as opposed to a double-on response pattern (perceived as /ta/) for longer VOT values. The first peak, labelled N100', was evoked by the closure release, and the second, labelled N100, by voicing onset. Although this result led Sharma and Dorman to suggest that the morphology of the N100 component was a neurophysiological correlate of voicing perception, they later provided evidence for a different view (Sharma & Dorman, 2000). In the latter study, they presented syllables with different places of articulation. When performing identification and discrimination tasks, English-speaking subjects showed a VOT perceptual boundary centred on 27.5 ms for the /ba-pa/ continuum and on 46 ms for the /ga-ka/ continuum. This well-known shift of the VOT boundary according to the place of articulation (the consonants with more back articulation display longer VOTs: Lisker & Abramson, 1967; for the trade-off between F1 and VOT associated with this shift, see Summerfield & Haggard, 1977; Parker, 1988) was not associated with a corresponding shift in the neurophysiological boundary between the double- and single-on patterns. It was therefore concluded that the N100 morphology (single vs double-peak) was not a reliable indicator of voicing contrast. Although a correspondence between the psychoacoustic boundary and the physiological one was found by Steinschneider et al. (2005) with intra-cortical recordings when varying place of articulation, they acknowledged that the matching between these two types of boundaries was not perfect, showing in this way the complexity of the trade-offs between acoustic events involved in the perception of voicing.

As to the N100 morphology, Eggermont and Ponton (2002) concluded that it does not reflect cortical correlates of voicing perception but rather intrinsic properties of the auditory cortex that *"may have been exploited by speech production mechanisms and tailored to the various temporal contrasts that are found around the 30- to 50-ms range" (p. 95).* This could mean that the appearance of a double-peak indexes the threshold of VOT values that can be reliably evaluated in terms of relative temporal order (Hirsh, 1959). Whereas VOT processing by the auditory cortex accounts for the universal acoustic boundaries, the perceptual boundaries in a phonological framework are language-specific and depend on the relationship between different features (Carden et al., 1981; Serniclaes & Wajskop, 1992).

It is therefore not surprising that several examples of the discrepancy between the N100 morphology boundary and the phonological boundary have been reported in different languages. An example of such dissociation was given by Sharma and Dorman (2000), who showed that Hindi- and English-speaking adults, who do not have the same phonological boundary for voicing, do however share the same N100 pattern of results. This result was replicated for Hebrew, a language for which neither the -30 nor the +30 ms universal boundary is relevant (Horev et al., 2007). Horev et al. (ibid) showed that /ba-pa/ stimuli with VOT values of -15 and +15 ms, which belonged to different phonological categories in Hebrew, evoked similar N100 components.

Understanding of the neurophysiological correlates of voicing perception in languages such as Hebrew, in which the phonological boundary is located around 0 ms, is particularly of interest. The emergence of this new phonological boundary, just midway between the two universal ones, is reputed to be based on mechanisms derived from our basic ability to determine the temporal relationships between two different acoustic events. Indeed, the 0 ms boundary corresponds to the perception of the temporal order between events with simultaneity as a boundary, not as a category. The mere accumulation of the outputs of the negative and positive VOT detectors would give rise to poor categorical perception: the 0 ms boundary is positioned in the middle of a 60 ms region (i.e. between -30 and +30 ms) in which neither negative VOT nor positive VOT are detected. The perception of temporal order would be completely ambiguous for the stimuli located inside this region if the outputs of the negative and positive VOT detectors were simply cumulated. Some kind of integration between VOT detectors must therefore be necessary for the categorical perception of temporal order, and behavioural data collected with adult French listeners support this assumption (Serniclaes, 1987; 2000). Labelling data collected with stimuli generated by factorial variation of negative and positive VOT suggest that outputs of the two VOT detectors interact in such a way that their boundaries converge to 0 ms, a process which has been referred to as "coupling". The higher complexity of temporal order perception is congruent with the fact that the neural correlates of the 0 ms VOT boundary are not reducible to those found for the processing of the universal VOT boundaries. Understanding these correlates is therefore of special interest.

The present study

We are not aware of any study having explored the electrophysiological N100 morphology over a single VOT continuum encompassing both the negative and the positive universal boundaries. Our study capitalized on the fact that the phonological boundary is centred on 0 ms in French to explore whether evidence can be found of the two universal boundaries by recording AEP. Should these neurophysiological correlates be available, insight into the mechanism by which they contribute to the phonological boundary could be gained by manipulating them.

The aim of this study was to determine whether /də-tə/ syllables with VOT values varying from -75 to +75 ms presented to French-speaking subjects would show an N100 correlate of the two universal boundaries. As the N100 component reflects the encoding of the successive events that compose the stimuli, it was expected that for positive VOT values greater than +30 ms, the N100 component would be composed of two peaks, the first evoked by closure release and the second by voicing onset. For negative VOT values longer than -30 ms, the N100 component should also be double-peaked but with the first peak evoked by voicing onset and the second by closure release. Finally, for VOT values between -30 and +30 ms, the N100 should be single peaked.

This study involves two specific features that set it apart from other studies, one pertaining to the selection of French-speaking subjects and another pertaining to the stimuli used. French is indeed characterized, in contrast to English, by a mismatch between the acoustic (-30 and +30 ms VOT) and the phonological (0 ms VOT) boundaries. This discrepancy enabled us to unambiguously dissociate the effects of the acoustic and phonological boundaries on the morphology of the N100 component. As to the stimuli, the use of symmetrical positive and negative VOT values encompassing the -75 to +75 ms range allowed us to investigate both universal VOT boundaries.

Our working hypothesis was that if the N100 morphology is the neurophysiological correlate of basic acoustic mechanisms implied in voicing perception, we should observe a double-peak response along the acoustic universal boundaries, even though the phonological boundary to which French-speaking subjects are sensitive does not correspond to either of these boundaries. Indeed, even though the universal boundaries are no longer phonologically relevant in French after six months of age (Hoonhorst et al., in revision), the demonstration

that there still exist neurophysiological correlates of the ability to discriminate these initial voicing boundaries would give some support to the assertion that the phonological mode of perception corresponds to a higher level of perception which is based on basic auditory mechanisms of which we can still find evidence in adult subjects.

Method

PARTICIPANTS

Five native French speakers (four women and one man) aged 22–35 years (mean: 28 years and 5 months) volunteered to participate in the study. All were right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971) questionnaire and reported normal hearing.

STIMULI

Speech stimuli were synthesized syllables composed of the apical stop consonant /d/ or /t/ followed by the neutral vowel /ə/ so as to limit contextual effects (Serniclaes, 2005). Six syllables with VOTs varying from -75 ms to +75 ms with a 30 ms acoustic step were created.
Stimuli were generated by a parallel formant synthesizer (Klatt, 1980) provided by R. Carré (http://www.tsi.enst.fr/~carre/). The onsets of the initial frequency transitions of F1, F2 and F3 were 200, 2200 and 3100 Hz, respectively. Steady state formant frequencies were 500, 1500 and 2500 Hz, respectively. The F0 value was 120 Hz, and the transitions lasted 20 ms. The overall duration of all syllables was 200 ms (figure 1). Negative VOT was synthesized with periodic energy (60 dB), F1 bandwidth at 50 Hz, and F2 and F3 bandwidths both at 600 Hz. Positive VOT was synthesized with aperiodic energy (30 dB), with F1 bandwidth at 600 Hz, and F2 and F3 bandwidths at 70 and 100 Hz, respectively. The voiced vocalic segment was synthesized with periodic energy (60 dB) and with F1, F2 and F3 bandwidths at 50, 70 and 100 Hz, respectively.
Results of research conducted by Medina and Serniclaes (2005) with the same stimuli showed that French-speaking adults' identification boundary was centred on 0 ms VOT and that adults' AX discrimination scores peaked for the contrast straddling this boundary.

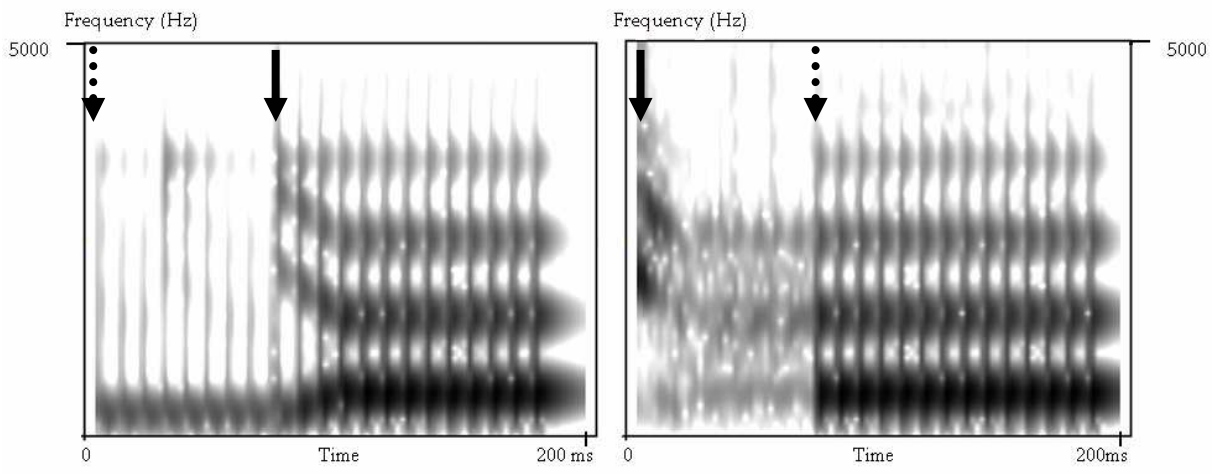

Figure 1 : /də/ and /tə/ stimuli with -75 ms VOT (left) and +75 ms VOT (right). The solid black arrow represents closure release and the dotted black arrow the onset of voicing.

PROCEDURE

Threshold measurement: Subjects were tested in a sound-attenuated booth using an adaptive three-alternative forced choice procedure, with a three-down, one-up decision rule using Tucker-Davis Technologies (TDT) hardware and software (System III ; PsychRP 1.05). The target stimulus used for measuring the auditory threshold was one of the six stimuli used in the experiment, a /də/ syllable characterized by a VOT value of -45 ms. The inter-stimulus-interval (ISI) lasted 500 ms.

Each trial, composed of three intervals marked by lights, one containing the target sound, was initiated by the subject. Intensity was first reduced in 5 dB steps until the fourth reversal occurred and then by 2 dB steps. The procedure stopped after 12 reversals and the first four reversals were excluded from the calculation of the average threshold. The threshold was measured two to three times. The final threshold corresponded to the mean threshold separated by less than 2 dB, i.e. the reference used for delivering the stimuli 50 dB above the individual threshold in the electrophysiological recordings.

EEG recording: Subjects were seated in a comfortable armchair. They passively listened to auditory stimuli while reading a book of their choice. They could take a break whenever they wished.

Stimuli were presented in 30 blocks of 360 stimuli, each lasting +/- seven minutes depending on the ISI, which was randomized and fluctuated between 500 and 1500 ms. Each block was composed of 60 exemplars of each stimulus delivered in a pseudo-randomized order at 50 dB

SL, binaurally through insert earphones (ER-1 TUBEPHONE, Etymotic Research) connected to eartips (12 mm) placed in the ear canal through a 280 mm long silicon tubing. The pseudo-randomized sequence was created on an Eevoke device (ANT Software, The Netherlands) and interfaced with TDT hardware (System III), the former to control for presentation order and the latter to provide digital-to-analog conversion and to control for intensity.

Brain electrical activity was recorded using the 32-channel ASA EEG/ERP system (ANT). Thirty-two electrodes were embedded in a waveguard cap (ANT) based on the 10-10 system (Chatrian et al., 1985). The tip of the nose was taken as the reference and an electrode placed on the forearm served as ground. Electro-ocular activity (EOG) was monitored from two bipolar pairs of Bluesky electrodes located at the outer lateral canthi and the infra- and supra-orbital areas of the left eye. Impedances were kept below 5 kΩ.

DATA ANALYSIS

The signal was amplified (x20 for EEG and EOG channels), band-pass filtered (0.1–30 Hz) and continuously digitized with a sampling rate of 512 Hz. Continuous EEG was segmented in 1000 ms epochs including a 200 ms pre-stimulus onset baseline on which baseline correction was carried out. Responses exceeding +/-100 µV were excluded from averaging. An averaged waveform for each of the six stimuli was computed for each subject and across subjects to obtain grand averaged waveforms. Data analysis consisted of a visual examination for the presence of a single- or a double-peak pattern. Two time windows were delineated, the first one lasting from 0 to 120 ms to search for the first peak and the second one from 120 to 200 ms to search for the second peak (Sharma et al., 2000). A double-peak pattern was acknowledged only if two peaks were segregated between the two pre-defined time windows covering their respective expected latencies. Owing to the small number of subjects (five) involved in this pilot study, no statistical analyses were performed.

Results

The N100 component was observed for each VOT value (e.g. for +75 ms VOT stimuli, figure 2). Its scalp distribution showed the classic polarity reversal across the Sylvian fissure.

Two distinct N100 morphologies can be dissociated relative to the VOT value: VOT values between -30 ms and +30 ms gave rise to one single N100 peak while VOT values outside the -30 ms to +30 ms range evoked a double-peaked N100, at least for temporo-parietal

localizations. For the sake of clarity, we will label single peaks as N100 and designate the two components of double-peaked waveforms as N100c or N100v, according to which acoustic event (closure release or voicing onset), on the basis of the peak latency, evoked the component.

For most electrodes on the scalp convexity (see figure 2), a double-on response, composed of N100c followed by N100v, was easily observed for +45 and +75 ms VOT stimuli whereas single-on responses were associated with -15 and +15 ms VOT stimuli. For long negative VOT values (-75 and -45 ms), double-on responses composed of N100v followed by N100c were observed at parietal (Pz) and temporal (T7, T8) electrodes only. For +75 ms VOT stimuli, a more complex pattern, composed of three negative peaks, was observed at the T7 and T8 electrodes.

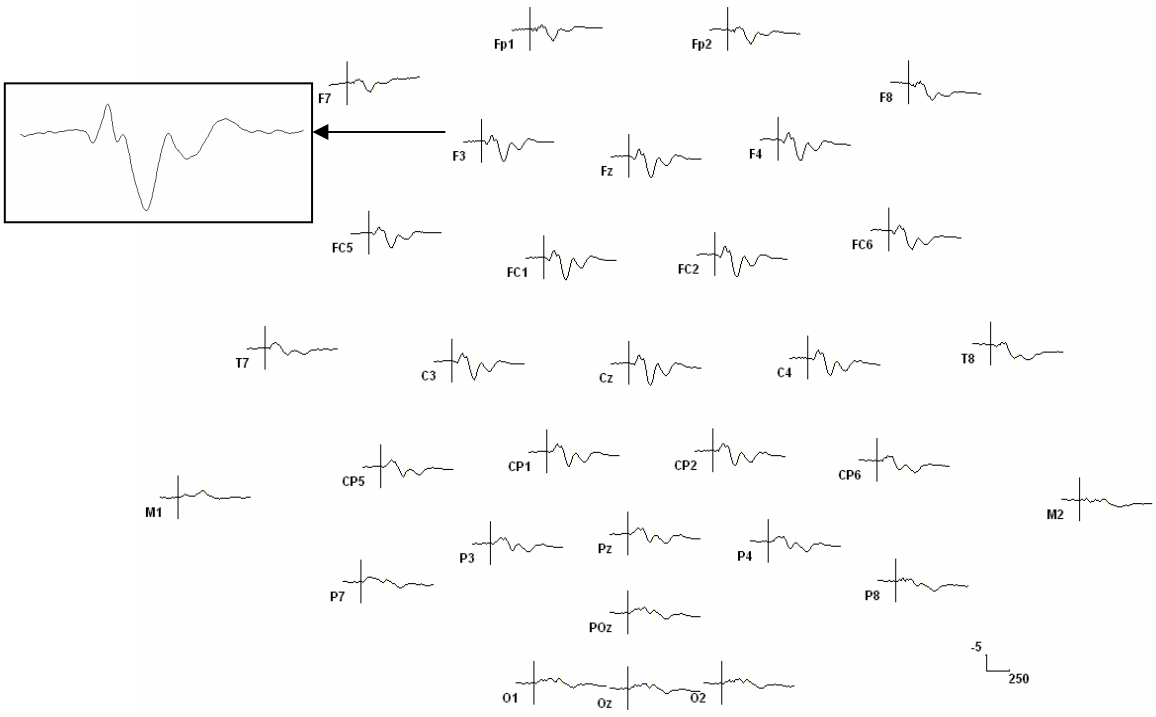

Figure 2: Grand averaged AEP evoked by +75 ms VOT stimulus. The vertical black line represents stimulus onset.

Since we did not notice any difference according to electrode lateralization and because the double-peak pattern was only observed for parietal and temporal electrodes in the negative VOT region, only waveforms recorded at the Cz, Fz , Pz, T7 and T8 electrodes (figures 3, 4, 5, 6 and 7) are shown.

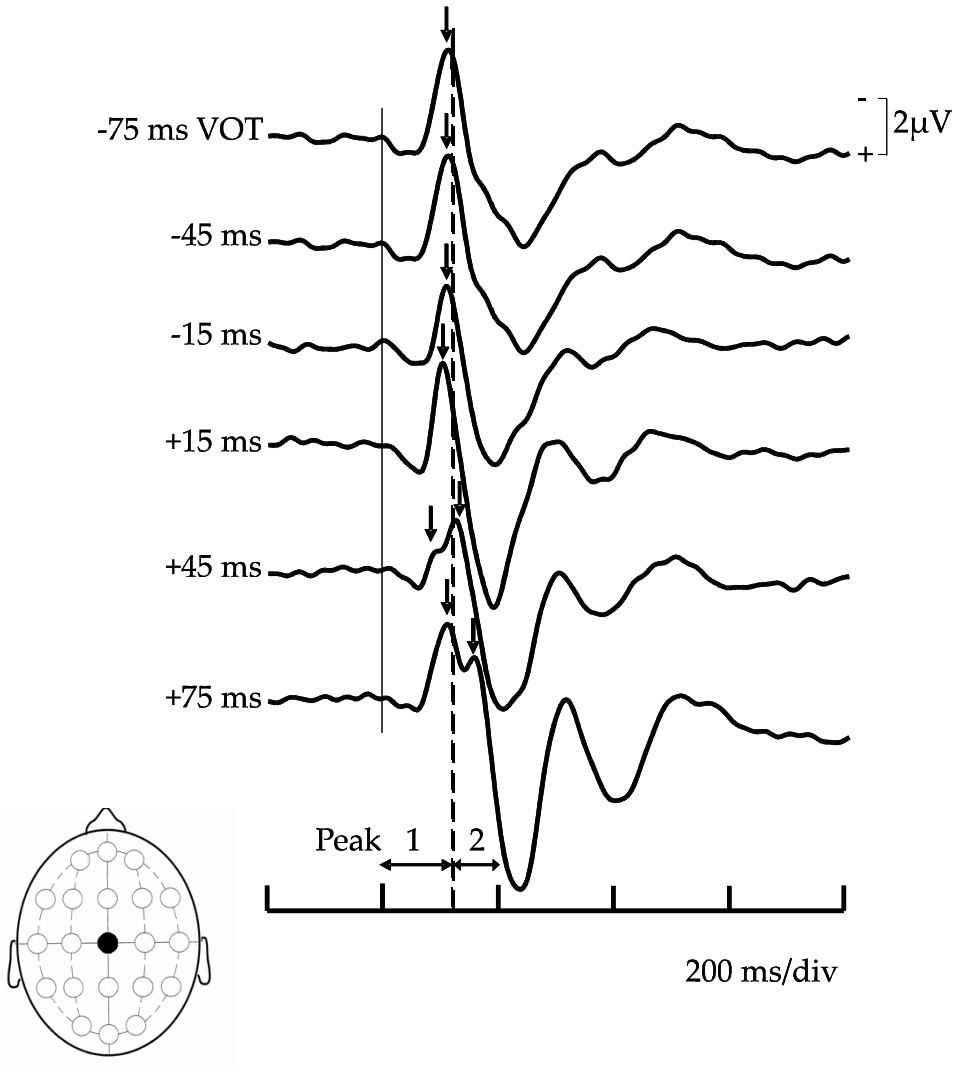

Figure 3: Grand averaged AEP recorded at the Cz electrode for -75, -45 and -15, +15, +45 and +75 ms VOT stimuli. The thin vertical black line represents stimulus onset. The dashed vertical black line delineates the limit between the two temporal windows (from 0 to 120 ms and from 120 to 200 ms, respectively) used to locate the first and the second peaks, marked by arrows.

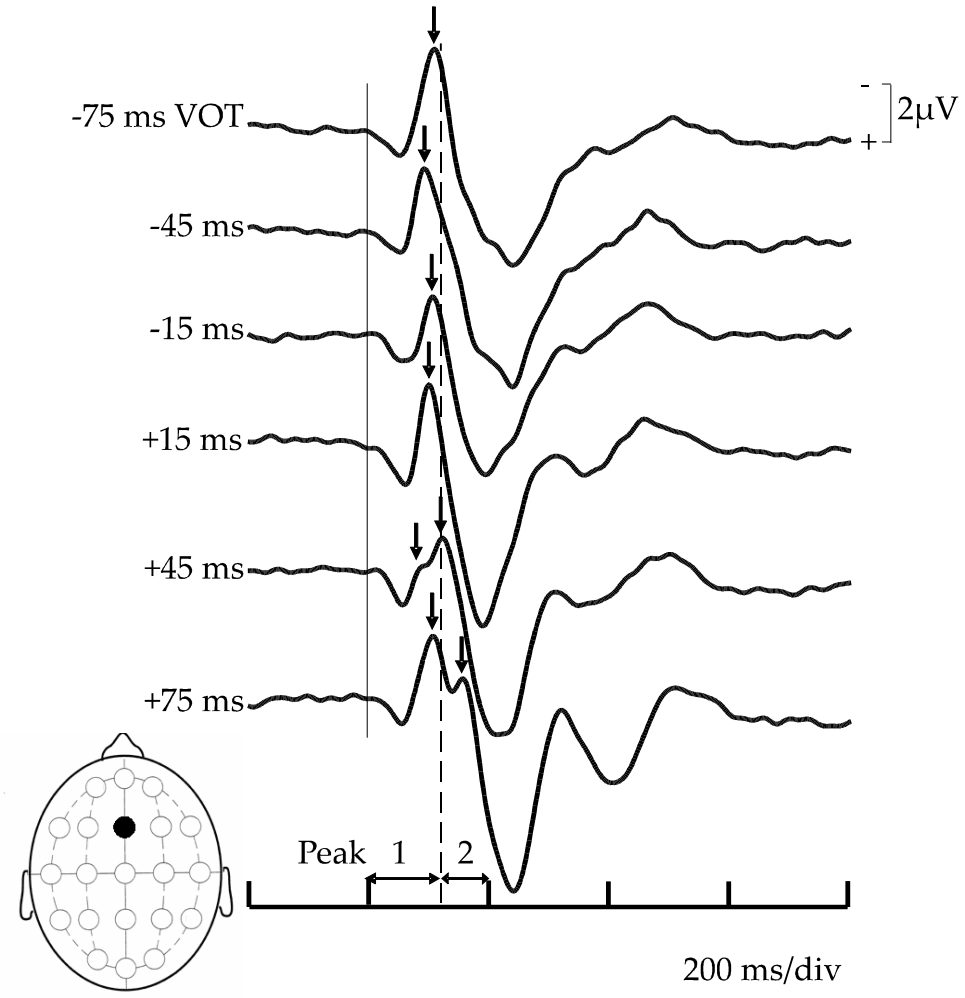

Figure 4: Grand averaged AEP recorded at the Fz electrode for -75, -45 and -15, +15, +45 and +75 ms VOT stimuli. The thin vertical black line represents stimulus onset. The dashed vertical black line delineates the limit between the two temporal windows (from 0 to 120 ms and from 120 to 200 ms, respectively) used to locate the first and the second peaks, marked by arrows.

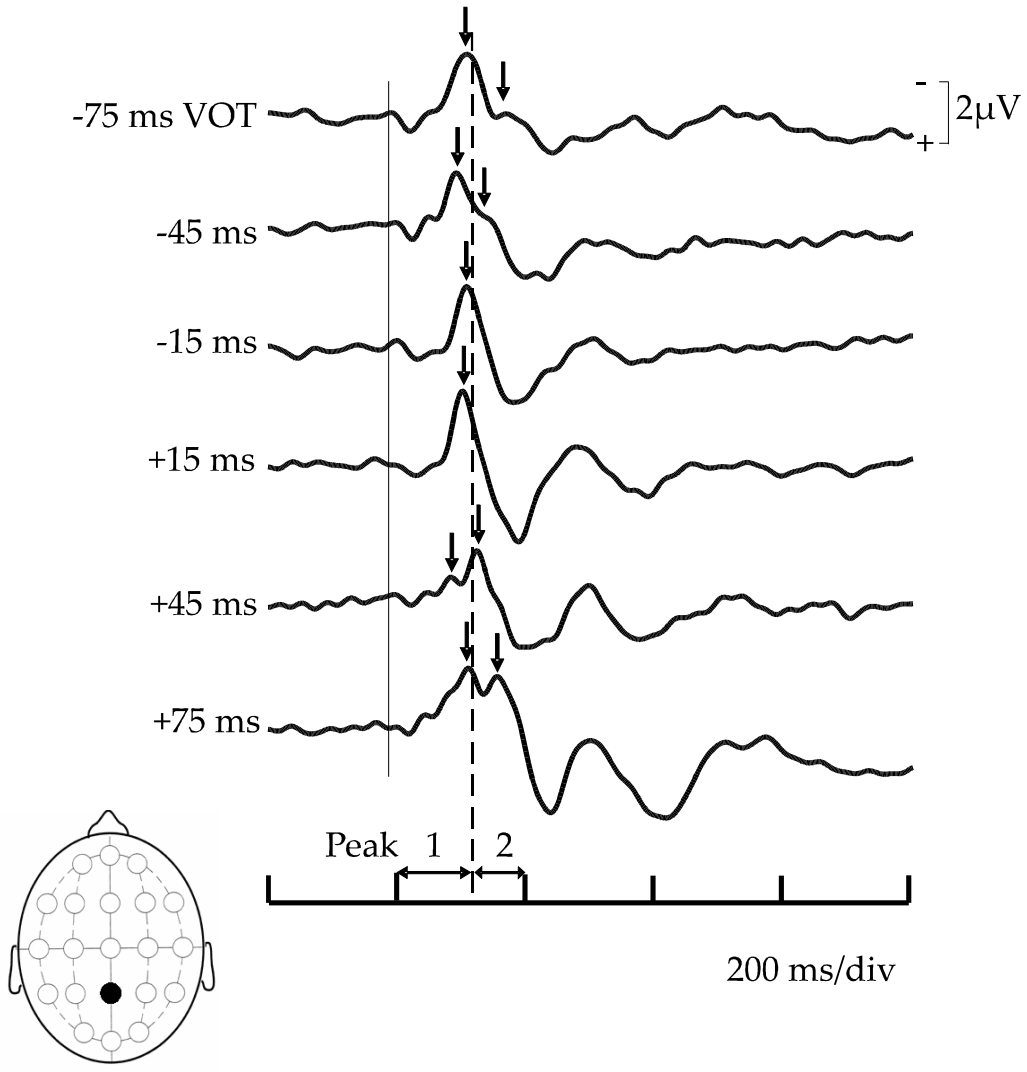

Figure 5: Grand averaged AEP recorded at the Pz electrode for -75, -45 and -15, +15, +45 and +75 ms VOT stimuli. The thin vertical black line represents stimulus onset. The dashed vertical black line delineates the limit between the two temporal windows (from 0 to 120 ms and from 120 to 200 ms, respectively) used to locate the first and the second peaks, marked by arrows.

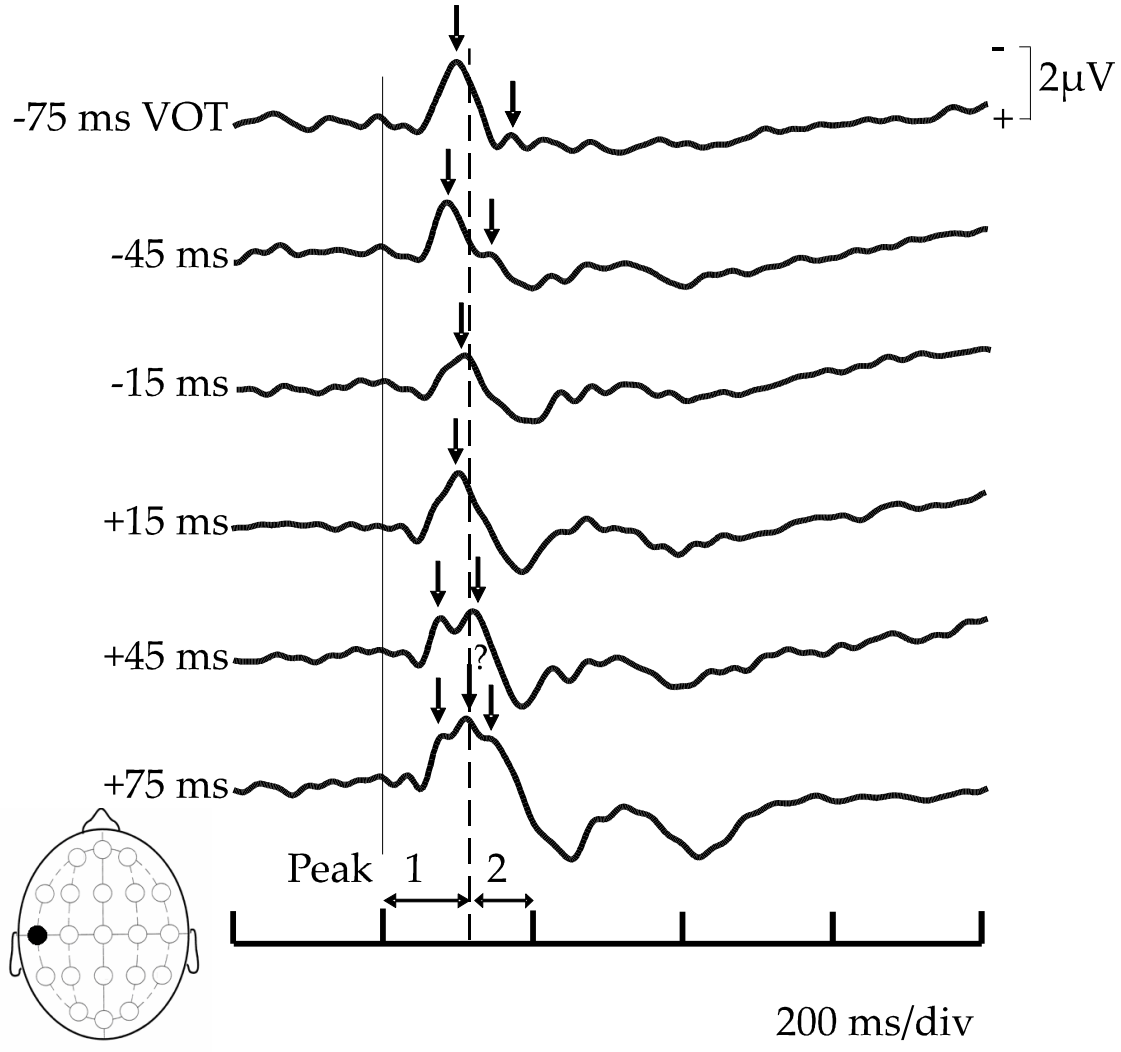

Figure 6: Grand averaged AEP recorded at the T7 electrode for -75, -45 and -15, +15, +45 and +75 ms VOT stimuli. The thin vertical black line represents stimulus onset. The dashed vertical black line delineates the limit between the two temporal windows (from 0 to 120 ms and from 120 to 200 ms, respectively) used to locate the first and the second peaks, marked by arrows. An unexpected middle peak (arrow with a question mark) is present for +75 ms VOT.

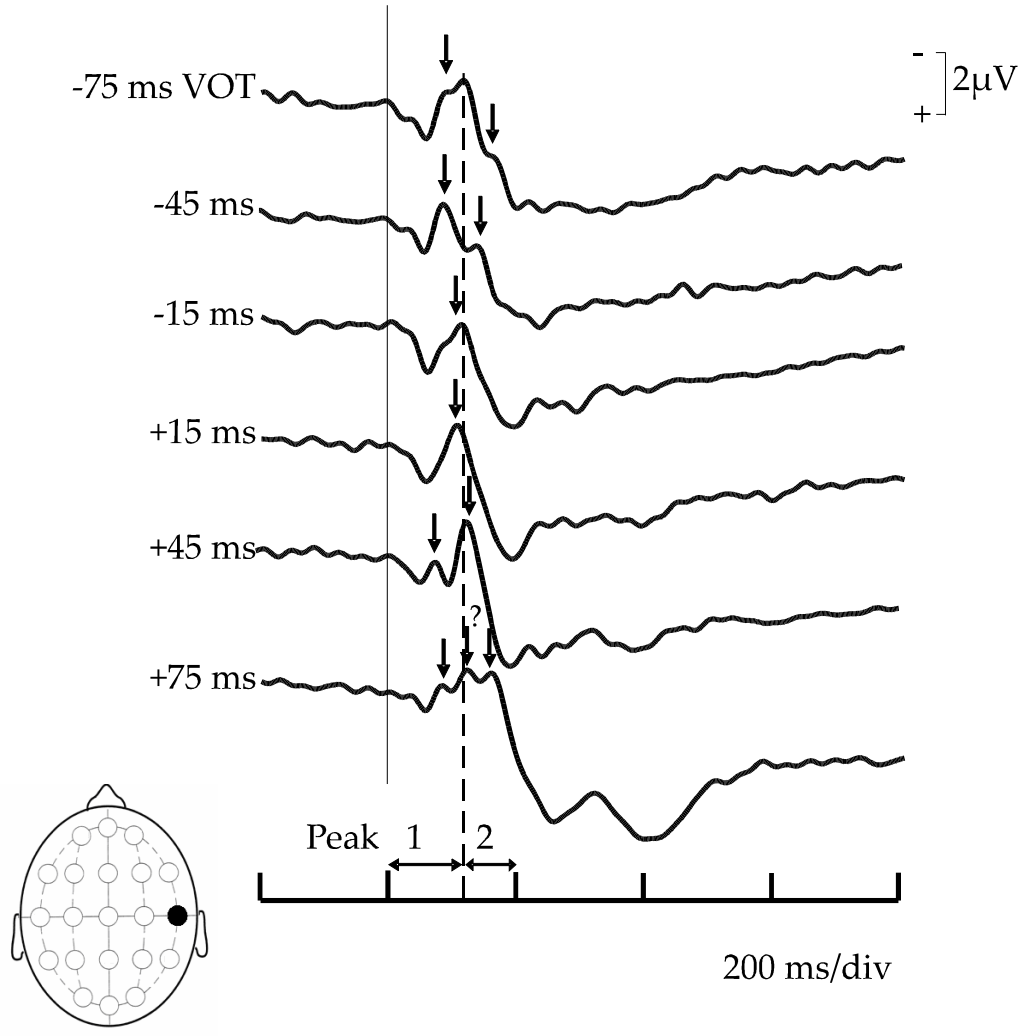

Figure 7: Grand averaged AEP recorded at the T8 electrode for -75, -45 and -15, +15, +45 and +75 ms VOT stimuli. The thin vertical black line represents stimulus onset. The dashed vertical black line delineates the limit between the two temporal windows (from 0 to 120 ms and from 120 to 200 ms, respectively) used to locate the first and the second peaks, marked by arrows. An unexpected middle peak (arrow with a question mark) is present for +75 ms VOT.

Discussion

The results presented above clearly show that, for French-speaking subjects, whose phonological boundary is distinct from the universal acoustic ones, there is still a clear evoked-potential correlate of the acoustic boundaries. This trace takes the common form of a double-on N100 wave appearing when VOT values exceed both the positive and negative universal boundaries. The expression of the double-on pattern on the scalp is, however, more widespread for positive than for negative VOT values. In the latter case, the double-on pattern is restricted to temporo-parietal electrodes.

The results obtained in the positive VOT region replicate, for French-speaking subjects, what has already been shown for English-speaking subjects. This adds weight to the demonstration that the double-on effect is independent of the phonological percepts (Sharma & Dorman, 2000). To the best of our knowledge, this is the first time that a scalp-recorded double-on boundary effect has been demonstrated at the level of the negative acoustic boundary. The results therefore confirmed our working hypothesis, according to which there remains a N100 trace of the basic neurophysiological mechanism from which phonological percepts are developed. Also, it is worth noting that whatever the region of the VOT continuum exploited by productions of a particular language, the voiced-voiceless opposition is associated with the single- vs double-on responses. Indeed, the opposition between long-lead (with VOT productions between -125 and -75 ms; Lisker & Abramson, 1964) and short-lag voicing distributions (with VOT productions between 0 and +25 ms; ibid) in French is echoed by differential double- vs single-on peaks of N100. English-speaking subjects benefit from the same correspondence. The short-lag distribution with VOT productions between 0 and +30 ms (associated with voiced phonemes; Kessinger & Blumstein, 1997) is coded by a single-peaked N100 waveform whereas the long-lag distribution with VOT productions between 50 and 170 ms (associated with voiceless phonemes; ibid) is coded by double-peaked N100.

An unanticipated finding was the differential scalp topography of the double-on response between the positive and the negative VOT regions. The underlying reason for this peculiar distribution remains to be investigated. A first hypothesis that can be put forward is that there is a difference in the relative contributions of the several cortical areas engaged in the perception of voicing (primary auditory cortex, anterior auditory field, secondary auditory cortex as showed by Eggermont, 2000) and of the different portions of the anterior part of the Heschl's gyrus (lateral, central and medial as showed by Steinschneider et al., 2005). This

difference in the cortical contributions is likely to be due to different interactions between the successive acoustic events according to their temporal order. For consonants characterized by a positive VOT value, the first formant transition plays an additional significant role (Burnham et al., 1991) while for consonants characterized by a negative VOT value the forward masking effect between the voicing bar and the closure release is larger. These acoustical interactions contribute to the greater psychoacoustic saliency described by Stevens and Klatt (1974) for phonemes with a positive VOT value.

As regards the forward masking effect, Eggermont (1995) showed that the characteristics (frequency, duration, amplitude) of the first feature that evokes the first N100 have a direct influence on the amplitude of the second response evoked by the second feature. Eggermont (1995) pointed out *"a strong interaction, reminiscent of forward masking, between the efficacy of the voiceless burst* [i.e. closure release] *in evoking neural activity and the size of the subsequent response to the onset of response" (p. 919).*

Whatever the genuine mechanism underlying the difference in the scalp topography for symmetrical positive and negative VOT values (-75 vs +75 ms and -45 vs +45 ms), the results suggest that it reflects the neural code for determining the VOT sign.

The differential contribution of the different cortical areas engaged in the perception of voicing (Eggermont, 2000) and of the different portions of the auditory cortex (i.e. lateral, central and medial; Steinschneider et al., 2005) in the perception of voicing could also explain the unanticipated pattern of three negative responses obtained at the T7 and T8 electrodes for the +75 ms VOT stimulus. One hypothesis for the appearance of the middle peak could be that this response is the result of the algebraic sum of different neuronal subpopulations with slightly different latencies.

Finally, the fact that we did not notice any lateralization effect reinforced the conclusions drawn by Shtyrov et al. (2000), who stated that *"P1, N100 and P2 responses, being non-specific responses to acoustic stimuli, are insensitive to the speech context of the sound and cannot be used for determining the brain's hemisphere dominant in speech perception" (p. 2896).*

These preliminary results open the perspective of assessing how the acoustic boundaries are integrated to give rise to the phonological boundary. The idea is to manipulate the acoustic boundaries either by studying pathologies such as acquired sensory hearing loss (Tremblay et al., 2003), simulated hearing loss (Martin et al., 1997), and auditory-based learning disorders

(Cunningham et al., 2000) in which they may be modified or by using auditory training techniques (Tremblay & Kraus, 2002).

Conclusion

In this chapter, we have summarized research aimed at cracking the neural code of voicing perception. The core idea underlying the present work was to seek neurophysiological correlates of the initial and universal basic auditory sensitivity that we share with animals and which underlies our ability to process the temporal order between two events. No matter whether our mother tongue has selected one of these initial acoustic boundaries for phonological use (e.g. English) or has built an entirely new phonological boundary (e.g. French), our data show that it remains possible to non-invasively record a neurophysiological signature of this primary ability from the scalp.

The single- vs double-on N100 recorded in our experiment with French-speaking subjects for either positive or negative VOT values seems a promising tool to investigate the encoding of voicing. A logical next step is to investigate how modifications of the basic acoustic boundaries influence the phonological percepts.

Etude 3.2

N1b and Na subcomponents of the N100 long latency auditory evoked-potential: neurophysiological correlates of voicing in French-speaking subjects[1]

Objective: to look for the presence of neurophysiological correlates of language-general voicing boundaries in French by analyzing the morphology of two N100 subcomponents (N1b and T-complex).

Methods: /də/ and /tə/ syllables with a VOT value varying evenly from -75 and +75 ms were presented to French-speaking adults as stimuli for scalp-recorded auditory evoked-potentials. Morphologies and peak latencies of N1b and T-complex subcomponents were assessed.

Results: the Na subcomponent of the T-complex was double-peaked for VOT values below -30 ms and above +30 ms. N1b subcomponent revealed a double-peaked response above +30 ms VOT and a single-peaked response for all other VOT values. Whenever the response was double-peaked, there was a correlation between the VOT value and the N1b or Na supplementary peak latency.

Conclusions: the combined morphologies of N1b and Na yield clear neurophysiological correlates of the language-general boundaries. For negative VOT values, the differential behavior of N1b and Na subcomponents suggests that only Na possesses physiological properties indexing the two language-general boundaries.

Significance: rather than being lost, the universal sensitivity of human newborns to language-general boundaries remains present even if in some languages such as French, they do not separate phonological categories.

[1] Hoonhorst I, Serniclaes, W, Collet G, Colin C, Markessis E, Radeau M, Deltenre P (in press). N1b and Na subcomponents of the N100 long latency auditory evoked-potential: neurophysiological correlates of voicing in French-speaking subjects. Clinical Neurophysiology.

Introduction

Speech perception requires the neural encoding of both spectral and temporal acoustic cues. Voice Onset Time (VOT), defined as the delay between voicing onset (the starting time of vocal cords vibration) and consonant closure release (Lisker & Abramson, 1967), is one of the most studied temporal acoustic cues. Even if the perception of voicing contrasts is affected by the variation of several acoustic changes, including spectral ones, (e.g. the frequency and transition length of F1), the perception of voicing relies largely on VOT (Lisker et al. 1978). This is especially true for phonemes made of an occlusive consonant followed by a vowel, so that by changing the VOT value only, the perceived phoneme can be switched from a /də/ to a /tə/ to take an example with an alveolar occlusive.

At the perceptual level, the monotonic variation of the VOT cue leads to a non-monotonic perception (Liberman et al., 1957), a phenomenon known as categorical perception (CP). It has been demonstrated that CP is already present in infants as young as one month of age (Eimas et al., 1971) and that the boundaries delimitating the categories undergo some maturational changes during the first year of life: whereas newborns perceive voicing according to two "language-general" boundaries located at -30 and +30 ms VOT, a few months later they become sensitive to the "language-specific"[2] boundary/ies of their linguistic environment. In most of the world languages, this general-to-phonological remapping leads to a single boundary centered on 0 ms VOT (e.g. French: Serniclaes, 1987; Spanish: Williams, 1977; Polish: Flege & Eefting, 1986; Dutch: Maassen et al., 2001; Hebrew: Horev et al., 2007 and Arabic: Yeni-Komshian et al., 1977) whereas in English the phonological boundary is located around some +30 ms VOT (Lisker et al., 1978). The occurrence of language-general voicing boundaries has been demonstrated to be neither speech-specific nor human-specific. Results obtained using Tone Onset Time (Simos et al., 1998a; 1998b), Formant Onset Time (Horev et al., 2007) and Noise Onset Time (Miller et al., 1976) in humans and in non-human animals (Kuhl & Miller, 1975; 1978) have shown a natural discontinuity in the discrimination

[2] Throughout the paper, the term "language-general" will be used to designate the universal ability shared by non-human animals and humans before six months of age to discriminate voicing according to -30 and +30 ms voicing boundaries and "language-specific" to refer to the phonological boundaries that differ according to the language spoken by the subject.

of stimuli in which the experimenter has manipulated the delay between acoustic events (Hirsh, 1959; Pisoni, 1977). It is now proposed in the literature that the perceptual system exploits this natural discontinuity to perceive voicing (Holt et al., 2004).

Both scalp and intra-cranially-recorded Auditory Evoked-Potentials (AEP) have been extensively used in search of neurophysiological correlates of voicing perception and in the investigation of auditory mechanisms at the root of voicing perception. The temporal unfolding of the late scalp-recorded exogenous responses, mainly the N100 component, has been particularly studied in relation to VOT perception. The N100 component "responds to a steep change in a level of physical energy that has remained constant for at least a short time" and is therefore regarded as reflecting several aspects of the sensory encoding of time-varying stimuli (Näätänen & Picton, 1987). As far as VOT is concerned, Martin and Boothroyd (2000) showed that the transition from noise to a periodic signal (and vice versa) elicited a N100 component. More specifically, both intra-cortical (e.g. Steinschneider et al., 1999) and scalp-recorded AEPs (e.g. Sharma & Dorman, 1999) have shown that short positive VOT values (from 0 to +30 ms VOT) evoked a "single-on" N100 response whereas long positive VOT values (+30 ms and above) evoked a "double-on" N100 response, i.e. two N100 peaks, respectively time-locked to closure release and voicing onset. Anatomically, the N100 is generated by multiple areas on the superior temporal plane with a major contribution from the lateral part of Heschl Gyrus and the Planum Temporale (Liégeois-Chauvel et al., 1991; Godey et al., 2001). Physiologically, the appearance of the N100 single- vs. double-on patterns appears to be determined by several non-linear interactions between the responses to two successive acoustic events. Recent studies and especially those using intra-cortical recordings (e.g. Steinschneider et al., 2005) suggested that the N100 pattern evoked by Consonant-Vowel (CV) phonemes is not a linear summation of two N100 evoked by successive acoustic events but the combined activity of distributed yet simultaneously active sources (Budd et al., 1998). These synchronized onset responses provide a specific temporal processing mechanism that overcomes all the potential sources of variability e.g. forward masking effects for positive VOT stimuli (the voicing neural response masking the closure release response, Horev et al., 2007) and vice versa or intrinsic cell properties (refractory period duration, characteristic frequency, Eggermont, 2000). As stated by Steinschneider et al. (2005, p.180), the synchronization of the neural responses to closure release and voicing onset is "an especially

powerful means" of allowing our perceptive system to determine the temporal order between these two acoustic events and so to discriminate voiced vs. voiceless phonemes.

Relying on previous studies showing that speech processing largely depends on basic auditory mechanisms specialized for the representation of temporal acoustic information (Eggermont, 2001; Eggermont & Ponton, 2002), we intended to explore the neurophysiological correlates of language-general voicing boundaries in adults whose mother tongue was French, a language in which language-general (-30 and +30 ms VOT) and phonological (0 ms VOT) boundaries do not match, so that they can be optimally distinguished. The present study encompasses two original features, one pertaining to the VOT continuum used, and the other to the data analysis methodology. There is paucity of data pertaining to the neural encoding of negative VOT values. Using a -15 ms VOT stimulus, Horev et al. (2007) evidenced a single N100 response but predicted in their conclusions that they should have recorded a double-on response, if longer negative values had been presented. This prediction was at variance with the results from Sharma and Dorman (2000) who, by studying scalp-recorded AEP evoked by stimuli with a VOT continuum extending from -90 to 0 ms, did not find a double-peaked N100 response for stimuli with a VOT value below -30 ms. However, in an intra-cortical study, Liégeois-Chauvel et al. (1999) did obtain a double-on response with a -120 ms VOT natural stimulus produced by a French-speaking subject. Faced with these contradictory results, we decided to revisit the matter by presenting stimuli with VOT values systematically varying from -75 ms to +75 ms with a 30 ms step, i.e. a continuum that encompasses symmetrical positive and negative values crossing the two general boundaries. As to analysis of the AEP components, it bore not only on the N1b subcomponent (the N100 subcomponent maximally recorded at Fz with a latency at around 100 ms ; Näätänen & Picton, 1987) but also on the T-complex (N100 subcomponents maximally recorded at temporal electrodes and composed of a negative subcomponent at 70-80 ms (Na), of a positive subcomponent (Ta) at 100 ms and of a negative peak (Tb, also labeled N1c) at 140-160 ms ; Wolpaw & Perry, 1975; Tonnquist-Uhlen et al., 2003).

This interest in the scalp-recorded T-complex which, to the best of our knowledge, has never been investigated in the context of both negative and positive VOT perception, stemmed from the results of a pilot study conducted on five French-speaking subjects (Hoonhorst et al., 2009). These preliminary results, recorded from a smaller number of subjects and computed with a nose reference, suggested that a double-on pattern emerged for negative VOT values below -30 ms at the temporal electrodes but was not seen when looking at the midline

recording sites. As midline responses correspond to the summation of electrical fields from both temporal lobes (Scherg & Van Cramon, 1985), the use of a nose reference located on the zero-equipotential line (Vaughan, 1974) is well suited to record the N1b subcomponent. Since this reference yielded illegible T-complexes, the latter were analyzed using an average reference. Moreover, resorting to an average reference is mandatory to evidence the inverted activity between N1b and Ta that allows positive identification of the T-complex waves (Picton et al., 2000). For all these reasons, we decided to extend the former study to a greater number of subjects and to use an average reference to be able to study the subcomponents recorded at temporal sites.

Our working hypothesis was that there would be neurophysiological correlates of both -30 and +30 ms language-general boundaries in French-speaking adults even if these boundaries are not phonological in French. According to this hypothesis, the N100 single vs. double-peaked morphology would index acoustic sensory encoding.

Methods

PARTICIPANTS

Ten right handed (Edinburgh Handedness Inventory ; Oldfield, 1971) normally hearing native French speakers (seven women and three men) aged 22–35 years (mean: 26 years and 10 months) volunteered to participate in the study. The experimental protocol was approved by the ethical committee of Brugmann Hospital (Brussels, Belgium) where the electrophysiological recordings took place.

STIMULI

The speech stimuli were synthesized syllables composed of the apical stop consonant /d/ or /t/ followed by the neutral vowel /ə/ so as to limit contextual effects (Serniclaes, 2005). Six syllables with VOTs varying from -75 ms to +75 ms with a 30 ms acoustic step were created. Stimuli were generated by a parallel formant synthesizer (Klatt, 1980) provided by R. Carré (CNRS, France). The onset of the initial frequency transitions of F1, F2 and F3 were 200, 2200 and 3100 Hz respectively and the formant transitions lasted 20 ms. Steady state formant frequencies were 500, 1500 and 2500 Hz respectively. F0 value was maintained constant at 120 Hz. All syllables had an overall duration of 200 ms (Fig. 1). Negative VOT was

synthesized with periodic energy (60 dB) with F1 bandwidth at 50 Hz, and F2 and F3 bandwidths both at 600 Hz. Positive VOT was synthesized with aperiodic energy (30 dB), with F1 bandwidth at 600 Hz, and F2 and F3 bandwidths at 70 and 100 Hz respectively. The voiced vocalic segment was synthesized with periodic energy (60 dB) and with F1, F2 and F3 bandwidths at 50, 70 and 100 Hz respectively. Results of research conducted by Medina and Serniclaes (2005) with the same stimuli showed that French-speaking adults perceived negative VOT stimuli as /də/ and positive VOT values as /tə/.

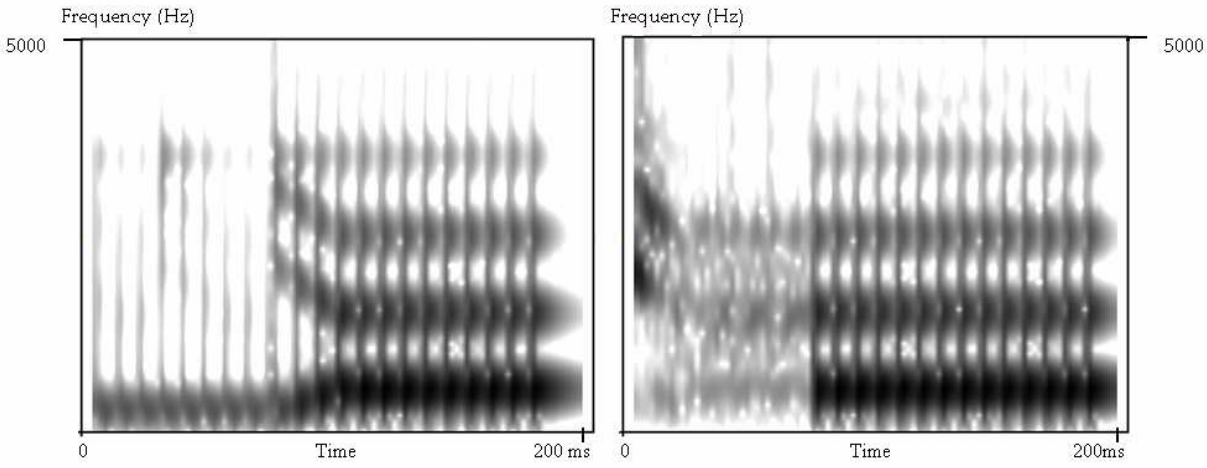

Figure 1. Spectrograms of /də/ and /tə/ stimuli with -75 ms VOT (left) and +75 ms VOT (right).

PROCEDURE

Threshold measurement: subjects were tested in a sound-attenuated booth using an adaptive three-alternative forced choice procedure, with a three-down, one-up decision rule using Tucker-Davis Technologies (TDT) hardware (System III) and software (PsychRP 1.05). The stimulus used for measuring the auditory threshold was one of the six stimuli of the experiment, a /də/ syllable characterized by a VOT value of -45 ms[3]. The inter-stimulus-interval (ISI) lasted 500 ms. Each trial was composed of three presentation intervals marked by lights and was initiated by the subject. For each trial, the stimulus was randomly assigned to one of the three intervals. The subject's task was to identify which of them contained the stimulus. Intensity was first reduced in five dB steps until the fourth reversal occurred and

[3] This stimulus was chosen for the threshold measurement because its RMS voltage was closest to the median RMS value computed across the whole range of stimuli used for electrophysiological recordings.

then by two dB steps. The procedure stopped after 12 reversals and the first four reversals were excluded from the calculation of the average threshold. The threshold was measured two to three times in order to obtain two replicated measurements separated by less than two dB. The final threshold corresponded to the average of these two values and was used as the reference for delivering the stimuli 50 dB above the individual threshold in the electrophysiological recordings.

EEG recording: AEPs were recorded in inattentive condition. Subjects were seated in a comfortable armchair and passively listened to auditory stimuli while reading a book. They could take a break whenever they wished. Stimuli were presented in 30 blocks of 360 stimuli, each block lasting approximately seven minutes depending on the average ISI (individual ISI values were randomized and fluctuated between 500 and 1500 ms). Each block was composed of 60 exemplars of each stimulus delivered binaurally in a pseudo-randomized order at 50 dB SL, through insert earphones (ER-1 TUBEPHONE, Etymotic Research) connected to eartips (12 mm) placed in the ear canal through a 280 mm long silicon tube. The pseudo-randomized sequence was created on an Eevoke device (ANT Software, The Netherlands) and interfaced with TDT hardware (System III), the former to control for presentation order and the latter to provide digital-to-analog conversion and to control for intensity. Brain electrical activity was recorded using 31 channels of an ASA EEG/ERP system (ANT). The tip of the nose was taken as the physical reference. Electro-ocular activity (EOG) was monitored from two bipolar pairs of Bluesky electrodes located at the outer lateral canthi and the infra- and supra-orbital areas of the left eye. Impedances were kept below 5 kΩ. The 31 active electrodes were embedded in a waveguard cap (ANT) based on the 10-10 system (Chatrian, Lettich & Nelson, 1985).

DATA PROCESSING

The signal was amplified (x20 for EEG and EOG channels), band-pass filtered (0.1–30 Hz) and continuously digitized with a sampling rate of 512 Hz. Continuous EEG was segmented in 1000 ms epochs including a 200 ms pre-stimulus onset baseline on which baseline correction was carried out. Responses exceeding +/-100 µV were excluded from averaging. Averaged waveforms were computed for each subject and then across subjects to obtain grand averaged waveforms for each VOT value. Regarding the T-complex, which subcomponents are expressed on the right and left temporal sites and do not exhibit the classical morphology

with the nose reference, we decided to re-reference our tracings to an average reference (average of the activity recorded at Fp1, Fp2, F7, F3, Fz, F4, F8, M1, T7, C3, Cz, C4, T8, M2, P7, P3, Pz, P4, P8) so as to avoid single-reference bias and to identify Na, Ta and Tb waves (Scherg & Von Cramon, 1985; Picton at al., 2000). As advised by Picton et al. (2000), we ensured that the sum of the reference voltages was not significantly different from 0 V for each subject ($t(8)= - .46$, $p> .05$).

DATA ANALYSIS

The N1b first peak was defined as the largest negativity occurring between 90 and 120 ms (Pratt et al., 2007). T-complexes were identified on the basis of the classically-described inversion around 100 ms between the positive Ta subcomponent recorded at temporal sites and the negative N1b recorded at central sites (Lück, 2005). Na, Ta and Tb first peaks were identified if they fell within the following latency ranges after stimulus onset respectively: 60-90 ms, 90-110 ms, 130-170 ms (on the basis of Wolpaw & Penry, 1975). Supplementary peaks were searched in latency ranges having the same width as those used to identify first peaks but shifted by a value equal to the absolute VOT value.

Results

Due to a poor signal to noise ratio, data from one subject had to be rejected from the final analysis.

N1B SUBCOMPONENT:

Clear N1b subcomponents with maximal amplitudes along the fronto-central midline were observed for each VOT value. They exhibited the classic polarity reversal across the Sylvian fissure. Each of the nine subjects presented the same individual pattern of results. Figure 2 illustrates the grand averages obtained at the 31 scalp locations for the stimulus with a VOT value of + 75 ms.

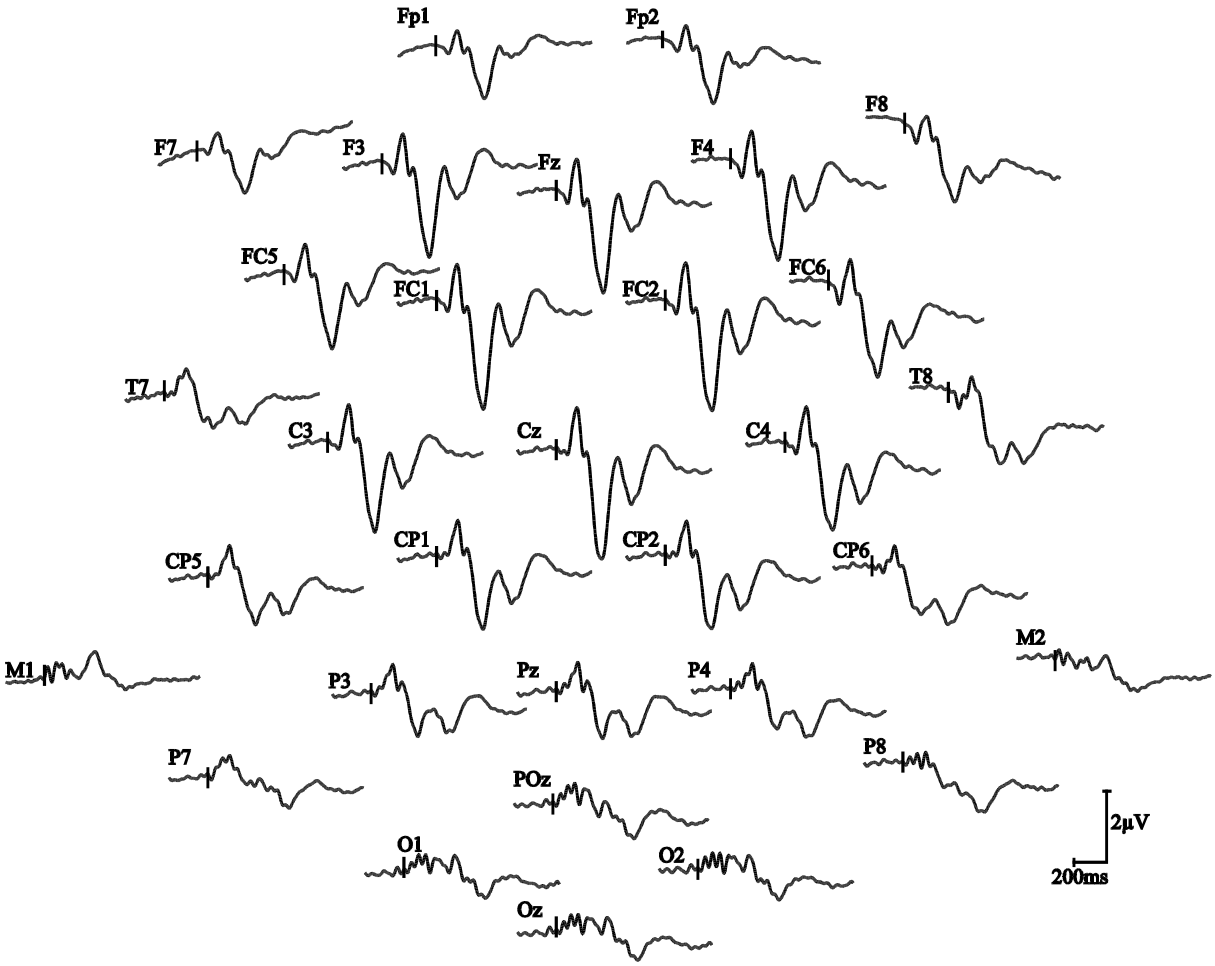

Figure 2. Grand averaged scalp distribution of AEPs in response to +75 ms VOT stimulus recorded with the tip of the nose as reference. The vertical black line represents stimulus onset. Negativity upwards.

Grand-averaged tracings obtained at the Fz electrode for each of the six VOT values are presented in figure 3.

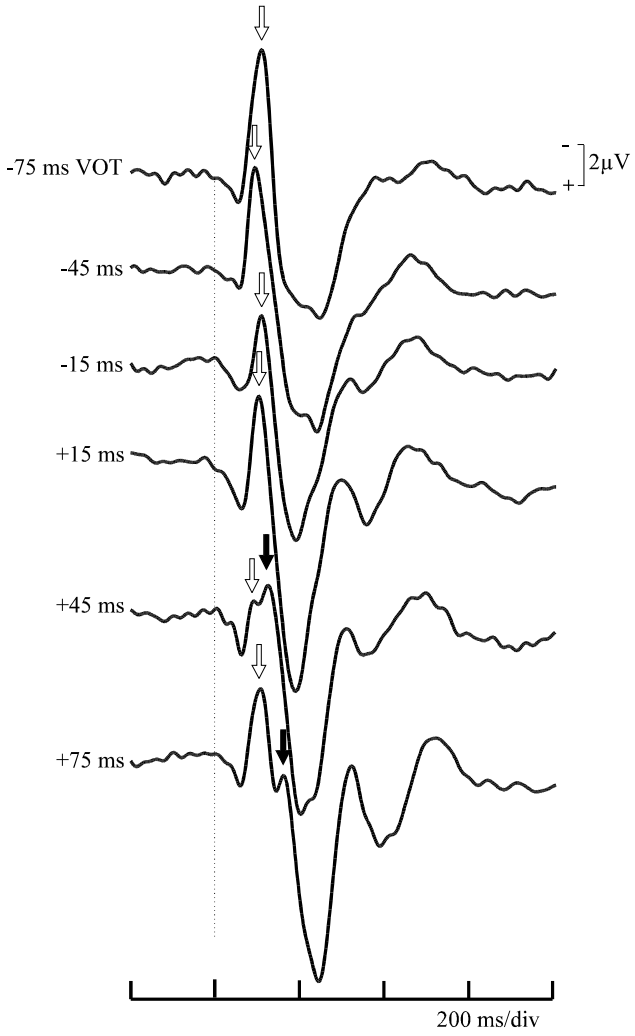

Figure 3. AEPs in response to -75, -45, -15, +15, +45 and +75 ms VOT stimuli recorded at Fz electrode with a nose reference. The dotted vertical black line represents stimulus onset. Empty arrows mark N1b first peak and black arrows mark N1b supplementary peaks.

A supplementary N1b peak appeared for positive VOT values when crossing the +30 ms VOT language-general boundary (i.e. when passing from +15 to +45 ms VOT) whereas for negative VOT values, no additional peak was observed below VOT values of -30 ms. A one-way repeated measures analysis of variance (ANOVA) revealed a significant effect of VOT on the N1b supplementary peak latency ($F(1,17)= 51.35$, $p< .001$). This VOT effect was further characterized by computing the correlation coefficients between N1b supplementary peak latencies and VOT values (+45 and +75 ms). This revealed a strong positive correlation ($r= .873$, $p< .001$). Moreover, the mean difference between latency obtained at +45 and +75

ms equated 36 ms, a result that did not significantly ($t(8)= 1.51$, $p> .05$) differ from the 30 ms step separating the VOT of the two stimuli.

T-COMPLEX:

T-complexes were identified in all subjects at right and left temporal sites with a clear polarity inversion between the positive Ta subcomponent recorded at temporal sites and the negative N1b recorded at central sites.

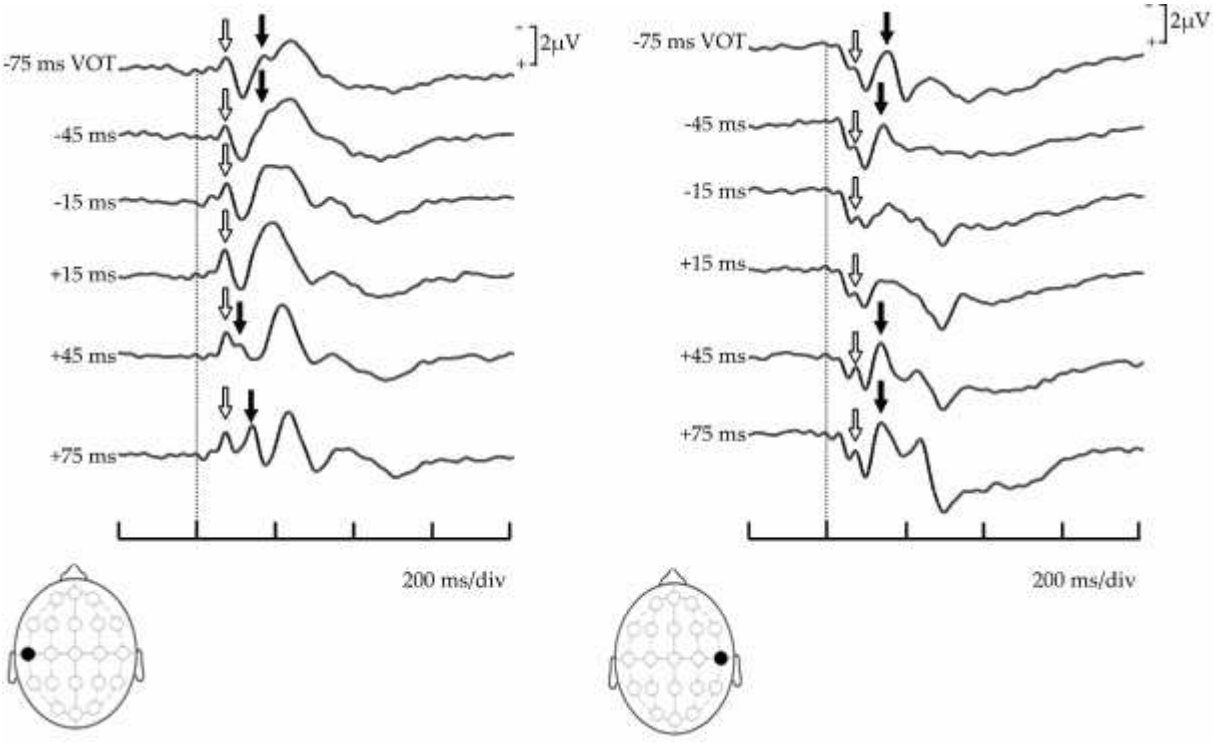

Figure 4. T-complexes in response to -75, -45, 15, +15, +45 and +75 ms VOT stimuli at T7 and T8 electrodes obtained after re-referencing to an average reference. The dotted vertical black line represents stimulus onset. Empty arrows mark Na first peak and black arrows mark Na supplementary peaks.

Figure 4 showed T-complexes for -15 and +15 ms VOT values composed of a first negativity identified as Na followed by a positive subcomponent corresponding to Ta itself followed by a negative peak referred to as Tb. Similar to what was observed for N1b, a supplementary peak was observed in the T-complex morphology when crossing the +30 ms VOT boundary. Moreover, this single vs. double-peaked morphology was also observed for negative VOT values below the -30 ms boundary (Fig.4). Since this supplementary peak appeared between

Na and Ta and because its relative latency with respect to Na was correlated to the VOT value, we identified it as a second Na peak. No supplementary peak was observed in the latency interval of Ta and Tb.

Repeated measures ANOVA was computed with VOT (four levels: -75, -45, +45, +75) and lateralization (two levels: right (T8) and left (T7)) as the independent variables and Na supplementary peak latency as the dependant variable. The results showed a main effect of VOT ($F(2.08,33.32)= 18.29$, $p< .001$ after Greenhouse-Geisser correction) on Na supplementary peak latency. This VOT effect was further characterized by computing the correlation coefficients between Na supplementary peak latencies and VOT values (-75, -45, +45 and +75 ms). Although both were rather loose, the negative correlation evidenced between supplementary Na latency and -75 and -45 ms VOT values ($r= -.42$, $p= .01$) and the positive correlation between supplementary Na latency and +75 and +45 ms VOT values ($r= .54$, $p= .001$) were significant. Moreover, the VOT x Lateralization interaction was significant ($F(2.08,33.32)= 6.77$, $p< .01$ after Greenhouse-Geisser correction). Although the Na supplementary peak latency was longer for -75 than for -45 ms and for +75 than for +45 ms for both T7 and T8 electrodes, the differences were significant for the T7 electrode only (respectively ($F(1,8)= 5.48$, $p< .05$ and $F(1,8)= 25.33$, $p= .001$), i.e. for the left temporal side. Finally the mean difference between Na supplementary peaks latencies obtained on the one hand at -75 and -45 ms and on the other hand at +45 and +75 at T7 electrode respectively equated 16 and 22 ms, results that did not differ significantly from the 30 ms step differentiating the two stimuli ($t(8)= -2.11$, $p> .05$ and $t(8)= -1.82$, $p> .05$ respectively). For T8, the mean difference between latency obtained at -75 and -45 ms and between +45 and +75 equated respectively 13 and 9 ms, results that differ significantly from the 30 ms step differentiating the two stimuli ($p< .05$). Mean latencies at T7 and T8 electrodes for Na supplementary peak are presented in figure 5.

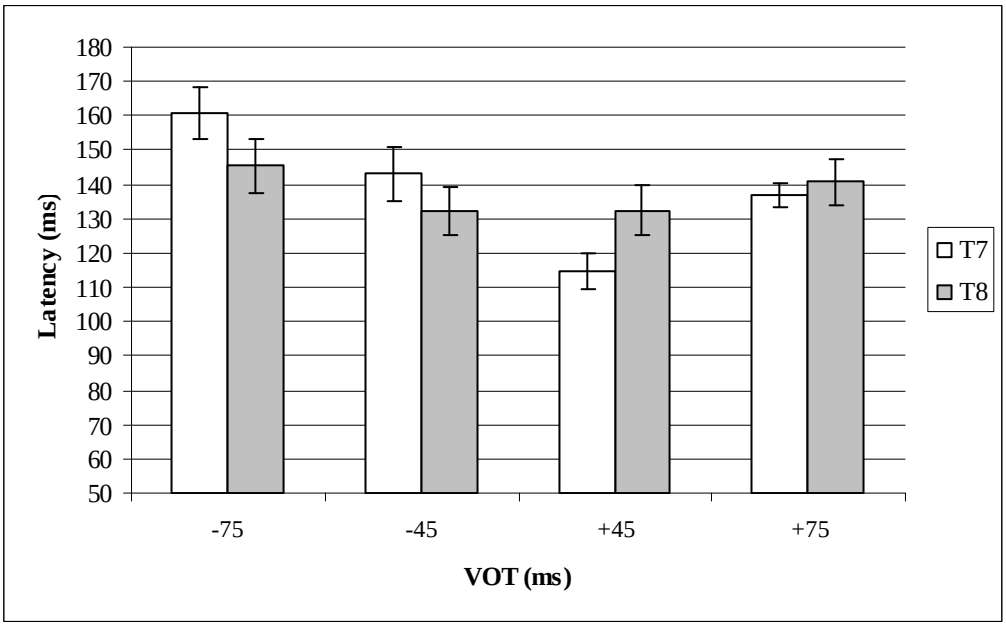

Figure 5. Mean latencies for Na supplementary peak at T7 and T8 electrodes for -75, -45, +45 and +75 ms VOT. Error bars represent standard deviations from the mean.

Discussion

The aim of the present study was to investigate whether, despite the development of a language-specific phonological boundary set at 0 ms VOT, adult native French speakers still possess electrophysiological correlates of the language-general voicing boundaries located at -30 and +30 ms.

The morphology of the N100 component yielded clear evoked-potential correlates of the language-general boundaries. The double-on pattern of N1b response observed for positive VOT values above +30 ms replicated for French-speaking subjects what had already been evidenced for English-speaking subjects (Sharma & Dorman, 1999). Contrary to this last study in which the authors limited their observations to the N1b subcomponent, we extended our analysis to the Na subcomponent. This supplementary analysis enabled us to evidence a similar discontinuity in the Na morphology for negative VOT values and a correlation between VOT value and N1b and Na supplementary peak latencies. The present results therefore confirmed the prediction made by Horev et al. (2007) according to which a double-on response should be recorded with negative VOT values (see Introduction) below -30 ms and reinforce the conclusion that auditory mechanisms for the language-general boundaries are still operational even though the boundaries they define have no phonological values in

the subject's language. Furthermore, the present data support the idea that the N100 component morphologies are a language-general correlate of voicing perception and not a phonological one, a conclusion already drawn by Sharma et al. (2000) and Sharma and Dorman (2000) for the N1b subcomponent. Indeed, their first study showed that the perceptual shift that occurred when modifying place of articulation was not correlated by a similar boundary shift between single and double-on N100 response. In their second study, Sharma and Dorman (2000) showed that the electrophysiological correlate of voicing was identical for Hindi- and English-speaking adults even though they did not share the same perceptual boundary.

The results also point out to a pair of asymmetries, one pertaining to the difference of N1b morphology between symmetrical negative and positive VOT values and the second relating to a difference between the effect of negative VOT values on Na and N1b morphologies. As explained below, these asymmetries are likely to be related to stimulus characteristics, i.e. the order of appearance between voicing onset and closure release as well as physiological differences between the T-complex and the N1b subcomponent.

Firstly, the finding that the N1b subcomponent is double-peaked when triggered by positive VOT values and single-peaked with negative VOT ones does not suit the hypothesis according to which the change from a single to a double-peaked N1b corresponds to a basic psychoacoustical natural threshold of discontinuity (Holt et al., 2004). Under this hypothesis, we should have found a double peak for both +30 and -30 ms boundaries. Moreover, this asymmetry between the language-general boundaries at -30 and +30 ms VOT was also found in several behavioral studies: some authors have pointed out the perceptive salience of the positive voicing boundary compared to the negative voicing boundary (e.g. Aslin et al., 1981). According to Stevens and Klatt (1974), the underlying reason is the presence of supplementary acoustic cues and, in particular, first formant transitions in positive VOT values. For Burnham et al. (1991), this salience is explained by a forward masking effect between the voicing bar and the closure release for negative voicing values. Reinforcing this idea, Eggermont (1995) pointed out *"a strong interaction, reminiscent of forward masking, between the efficacy of the voiceless burst* [i.e. closure release] *in evoking neural activity and the size of the subsequent response to the onset of response"* (p. 919). The comparison between behavioral and electrophysiological results, both showing an asymmetry between

negative and positive language-general VOT boundaries, suggests that the perceptive asymmetry is based upon general auditory mechanisms.

As to the second asymmetry, the Na subcomponent was, contrary to N1b, double-peaked for negative VOT values. This result supports those of Liégeois-Chauvel et al. (1999) who showed with intra-cortical recordings a double peak of neuronal discharge for a -120 ms natural voiced stimulus. The resulting question is: why is this double peak apparent when looking at the T-complex but not when looking at the N1b subcomponent? Answers to this question might come from physiological differences that have been evidenced between the T-complex and the N1b subcomponents: in young children between five and eight years, a T-complex has been obtained with short Inter-Stimulus-Interval (1.3/sec) whereas no N1b was recorded (Tonnquist-Uhlen et al., 2003). For Tonnquist-Uhlen et al. (2003), the *"robust presence of the T-complex without that of the N1b potential for children between five and eight years may indicate that the neural generators of the T-complex have a shorter recovery period and/or generators that mature earlier than those processes underlying other obligatory AEPs, specifically the N1b"* (p.697).

The pair of asymmetries found in our data might therefore be explained by a difference between the non-linear interactions (recovery period and forward masking) between the successive acoustic events according to which one of the two (voicing onset and closure release) occurs the first, potentially coupled with physiological differences between N1b and Na as suggested by Tonnquist-Uhlen et al. (2003).

Conclusion

The core idea underlying the present study was to seek neurophysiological correlates of the basic auditory sensitivity that humans share with animals and which are exploited by our linguistic perceptive system. Regardless of whether the mother tongue selected one of the initial language-general boundaries for phonological use (the positive VOT boundary in e.g. English) or built an entirely new phonological boundary (located at 0 ms VOT in e.g. French), our data show that it remains possible to non-invasively record a neurophysiological signature of this primary ability from the scalp. Both N1b and Na subcomponents yield information about the neural encoding of voicing but due to its better temporal resolution for less salient

features, the Na subcomponent of the T-complex appears as promising tool for investigating the neural encoding of voicing.

Discussion Générale

Synthèse

L'objectif de cette thèse était d'étudier le développement de la perception du trait de voisement. Notre travail s'inscrit dans la lignée des nombreuses études consacrées à ce sujet mais s'en détache à deux égards : la langue maternelle des sujets et la nature du continuum utilisé.

Nous avons vu dans l'introduction que le trait de voisement ne recouvre pas la même réalité perceptive dans toutes les langues. Les études menées indépendamment dans chaque langue concourent à la compréhension des règles de fonctionnement et de développement propres à chacune de ces langues mais seule la comparaison inter-linguistique permet de dissocier les mécanismes idiosyncrasiques des mécanismes généraux communs à toutes les langues. En étudiant la perception de sujets francophones, nos données s'ajoutent à celles déjà obtenues dans d'autres langues et notamment l'anglais, langue la plus étudiée dans la littérature. Par ailleurs, le processus de spécialisation phonologique en français se traduit par l'émergence d'une nouvelle frontière perceptive à 0 ms DEV qui ne correspond à aucune des frontières universelles de voisement (-30 et +30 ms DEV). Cette différence marquée entre frontières universelles et frontière phonologique devrait aider à identifier les différentes étapes du développement.

Eu égard au continuum utilisé, nous avons fait le choix de présenter des stimuli dont le DEV était positif mais aussi négatif. Ce choix s'impose dès lors que l'on étudie le processus par lequel les sujets francophones évoluent d'une perception universelle à une perception phonologique du voisement. D'autre part, il nous est apparu intéressant de tenter de comprendre pourquoi le DEV négatif est moins bien perçu que le DEV positif. Dans cette optique, la création de stimuli par synthèse vocale, nous a permis de mesurer les effets de la variation régulière du DEV négatif et positif. En ce sens, notre choix méthodologique basé sur la présentation de stimuli synthétiques est complémentaire de celui fait par Liégeois-Chauvel et al. (Liégeois-Chauvel, de Graaf, Laguitton & Chauvel, 1999 ; Laguitton, de Graaf, Chauvel & Liégeois-Chauvel, 2000 ; Trébuchon-Da Fonséca, Giraud, Badier, Chauvel & Liégeois-Chauvel, 2005) qui ont présenté à des sujets francophones des syllabes voisées et non voisées naturelles et dont nous avons fait mention à plusieurs reprises dans notre travail.

Avant d'entrer plus avant dans la discussion, reprenons les principales conclusions tirées des études présentées dans la présente thèse.

- De la première étude nous retiendrons que, entre 4 et 8 mois de vie, le processus de spécialisation phonologique pour le trait de voisement se traduit par un changement structurel de l'espace perceptif. En 4 mois, le nourrisson francophone passe d'un espace à trois catégories organisé autour de deux frontières universelles (-30 et +30 ms DEV) à un espace à deux catégories séparées par une frontière phonologique (0 ms DEV) localisée à mi-distance entre les frontières universelles. En confrontant nos données à celles de la littérature obtenues avec des nourrissons anglophones, nous avons mis en évidence que la bascule phonologique pour le voisement a lieu plus tôt en français qu'en anglais. Nous proposons que les distributions des productions des phonèmes voisés et non voisés sont à l'origine de cette différence développementale des habiletés de perception. Principalement, le fait qu'en anglais les distributions voisées et non voisées soient d'une part toutes deux caractérisées par un DEV positif et d'autre part adjacentes, requerraient que le système perceptif des sujets anglophones soit à même de dissocier des valeurs de DEV très proches. Cette acuité perceptive ne serait obtenue qu'au prix d'un allongement de la période de maturation.

- Dans la deuxième étude, la comparaison des résultats de PC (mesurée à l'aide de tâches d'identification et de discrimination) obtenus pour des syllabes voisées et non voisées, des couleurs et des expressions faciales, nous a permis d'étudier la maturation de la perception du voisement entre 5 et 8 ans et à l'âge adulte. De ces expériences, il ressort que la précision catégorielle (indexée par la raideur de la pente d'identification) se développe jusqu'à l'âge adulte tandis que la perception catégorielle relative (PCR ; mesurée en comparant les résultats de discrimination prédits sur base de l'identification et ceux réellement observés) arrive à maturité dès 6 ans. De manière plus spécifique, la divergence entre rythmes de développement des habiletés de perception des couleurs par rapport au voisement et aux expressions faciales, nous a amené à proposer que la durée du processus de maturation est liée à la complexité des percepts concernés. Par ailleurs, pour le continuum de voisement, aucune corrélation entre la PCR et le niveau de lecture n'a été établie, un résultat en défaveur de *l'hypothèse de lecture* défendue par Burnham (2003) selon laquelle l'apprentissage de la lecture a une influence sur la PC. Notons toutefois que nos résultats font état d'une

corrélation entre précision catégorielle et capacité à lire des mots irréguliers ou des non mots. L'interprétation de cette corrélation requiert de prendre en considération l'évolution développementale des habiletés de perception pour chacun des continua testés. La similarité des rythmes de développement constatée pour les continua de voisement et d'expressions faciales empêche de considérer l'apprentissage de la lecture comme cause commune de développement mais suggère à l'inverse que le développement des habiletés perceptives a des répercussions positives sur l'apprentissage de la lecture.

- L'étude du code neural sous-jacent à la perception du voisement chez l'adulte (études 3.1 et 3.2) nous a permis de mieux comprendre le mécanisme de spécialisation phonologique. Le fait de mettre en évidence chez l'adulte francophone des corrélats neurophysiologiques (simple vs double pic de N100) des frontières universelles de voisement qui, par ailleurs, sont aussi discriminées par des animaux et des nourrissons humains de moins de 6 mois, nous invite à considérer ces frontières comme les reliquats de mécanismes auditifs généraux sur lesquels sont élaborés des mécanismes perceptifs plus complexes.

La spécialisation phonologique : un processus contraint et dynamique

A la naissance, le cerveau de l'être humain et de l'animal en général n'est pas vierge mais pré-câblé dans le sens où diverses contraintes (génétique, physiologique…) influent sur le rapport du sujet au monde. Avant que la perception devienne une activité cognitive (telle que pensée par nombre de philosophes), elle est avant tout une interaction entre un environnement et un système biologique à la fois pré-câblé et doté de plasticité. Nous avons vu que des animaux partagent un certain nombre de ces contraintes : les chinchillas et l'être humain catégorisent un continuum de voisement selon des frontières identiques (-30 et +30 ms en moyenne) et qui se déplacent de manière semblable en fonction du contexte phonétique (+/- 15 ms). Sur base de cette observation, les partisans de la théorie auditive ont postulé que les mêmes contraintes étaient à l'œuvre chez l'ensemble de ces animaux : les frontières localisées à -30 et +30 ms sont universelles parce que contraintes par les mêmes *mécanismes auditifs généraux.* Bien que soumis à des contraintes initiales, le processus de spécialisation phonologique est aussi *dynamique :* par contact avec son environnement linguistique, l'être

humain se spécialise dans la perception des phonèmes de sa langue. En thaï, les frontières phonologiques correspondent aux frontières universelles. En anglais, seule la frontière centrée sur +30 ms DEV devient phonologique tandis qu'en français ainsi que dans la majorité des langues à deux catégories, les catégories voisée et non voisée sont situées de part et d'autre d'une frontière localisée à 0 ms DEV (étude 1).

Bien que l'adaptation à la phonologie de la langue débute vers 6 mois, la perception du voisement continue à évoluer pendant l'enfance. Dans l'étude 2, les résultats que nous avons obtenus avec des continua différents ont montré que la précision catégorielle s'affine avec l'âge tandis que la relation entre capacités d'identification et de discrimination est stable depuis au moins l'âge de 6 ans. L'augmentation de la précision catégorielle avec l'âge se traduit par une pente d'identification plus raide, une plus grande consistance des réponses et une moindre variabilité. En somme, le degré de fiabilité des réponses d'identification est corrélé avec l'âge du sujet. Ces données corroborent les résultats obtenus dans la littérature. Zlatin et Koenigsknecht (1975) ont abouti à cette conclusion en testant des enfants de 2 à 6 ans et des adultes anglophones avec des stimuli voisés et non voisés (de -150 à 150 ms DEV) issus de 3 continua de lieu différents (labial, apical et vélaire). Elliott, Busse, Partridge, Rupert, de Graff (1986) ont étendu cette conclusion à des enfants plus âgés en comparant les résultats d'identification d'un continuum /ba-pa/ obtenus par un groupe d'enfants âgés de 6 à 8 ans, un groupe d'enfants âgés de 6 à 11 ans et un groupe d'adultes, tous anglophones. Plus récemment Hazan et Barrett (2000) ont investigué les capacités d'identification d'enfants de 6 à 12 ans et d'adultes anglophones en présentant quatre continua, deux continua d'occlusives variant de par le DEV ou le lieu d'articulation et deux continua de fricatives variant de par ces mêmes traits acoustiques, et ont eux aussi conclu que la précision catégorielle s'améliorait avec l'âge. Bien que l'ensemble de ces données convergent, les raisons pour lesquelles la précision catégorielle continue à se développer restent floues.

Dans l'étude 2, nous avons pris comme point de départ de notre réflexion les hypothèses explicatives données par Burnham (2003) et Lalonde et Werker (1995). Pour Burnham (2003), la corrélation entre niveau de lecture et capacités de perception catégorielle (identification et discrimination) est interprétée comme facteur causal : l'acquisition de la lecture est la cause de l'amélioration des habiletés de perception catégorielle. Toutefois deux études dans la littérature suggèrent que la lecture affecte plutôt la *précision catégorielle*.

Mody, Studdert-Kennedy et Brady (1997) ont comparé les performances obtenues par des bons et des mauvais lecteurs âgés de 7 à 9 ans dont la tâche était d'identifier des syllabes issues de plusieurs continua (/ba-da ; /ba-sa/ ; /da-ʃa/ ; /seI-steI/), et leurs résultats montrent que la précision catégorielle est plus fine dans le groupe des bons lecteurs. Ces résultats sont appuyés par les données obtenues par Serniclaes, Ventura, Morais, Kolinsky (2005) qui, en présentant les syllabes /ba/ et /da/ (contraste de lieu d'articulation) à des adultes portugais lettrés et illettrés, ont montré que l'acquisition de la lecture a une incidence sur la précision mais pas sur la localisation de la frontière catégorielle. Nos résultats vont dans le même sens : nous n'avons pas mis en évidence une corrélation entre perception catégorielle relative et lecture mais bien entre précision catégorielle et lecture des mots irréguliers et des non mots.

Afin de tester l'hypothèse cognitive générale soutenue par Lalonde et Werker (1995), nous avons cherché à déterminer si les habiletés de perception catégorielle se développaient au même rythme quelle que soit la nature du continuum. Nos résultats de discrimination et d'identification obtenus avec un continuum de voisement, de couleurs et d'expressions faciales font état d'un rythme de développement semblable pour les contrastes de voisement et d'expressions faciales mais d'un développement plus rapide pour le continuum de couleurs. Ces résultats ne soutiennent donc pas l'idée que les habiletés de perception catégorielle se développent en synchronie mais soutiennent *l'hypothèse de spécificité* défendue par Gopnik et Meltzoff (1987). Pour ces auteurs qui ont étudié le lien entre développement des habiletés cognitives et développement sémantique chez le bébé de 15 à 20 mois, il existe bien un lien entre le développement de *certaines* habiletés cognitives et *certaines* habiletés sémantiques que l'on ne peut généraliser à l'ensemble des capacités cognitives.

Dans nos résultats, il existe une corrélation entre le développement de la capacité à percevoir les continua de voisement et d'expressions faciales. Cette dernière conclusion nous a amenés à proposer que la complexité des stimuli influence le rythme de développement des habiletés perceptives. La perception du trait de voisement fait intervenir différents niveaux d'intégration. A un niveau local d'intégration, chaque événement acoustique (relâchement de l'occlusion et vibration des cordes vocales) est analysé séparément et l'ensemble de ces indices est intégré à un niveau supérieur. A ce titre on peut dire que le corrélat acoustique du trait de voisement est *multiévénementiel.* Plutchik (1980) parle de stimuli multidimensionnels pour décrire les expressions faciales. Ces dernières résulteraient de combinaisons variables de quelques dimensions de base parmi lesquelles l'intensité, la similarité, les valences positive et

négative. Les couleurs peuvent elles aussi être considérées comme des stimuli multidimensionnels pour lesquels varient concomitamment la teinte, la luminosité et la saturation. La perception du voisement, des expressions faciales et des couleurs partageraient donc le fait que le percept intègre plusieurs constituants. Pour les couleurs, nous n'avons cependant fait varier que la teinte dans notre expérience, la luminosité et la saturation étant contrôlées. De ce fait, la réalisation des tâches d'identification et de discrimination était moins complexe pour le continuum de couleur que pour les continua de voisement et d'expressions faciales. Il est par ailleurs intéressant de noter que la similarité entre les continua de voisement et d'expressions faciales ne s'arrête pas là. En présentant des séries de points qui variaient au niveau de leur forme ou au niveau de leur configuration à des enfants de 4 à 10 mois, Deruelle et De Schonen (1995) ont en effet montré que l'hémisphère gauche traitait préférentiellement les informations locales (*traitement analytique*) contrairement à l'hémisphère droit, spécialisé dans une forme de traitement plus global (*traitement holistique*). Plus tard dans le développement, plusieurs études ont montré que l'enfant passait d'un mode analytique à un mode holistique de traitement des visages. Ce changement de 'stratégie' perceptive s'effectuerait vers 7 ans pour Schwarter (2000), vers 6 ans pour Mondloch, Le Grand et Maurer (2002) voire même plus tôt, vers 4 ans, pour Heering, Houthys et Rossion (2007).

La comparaison entre ces trois continua et la similarité de développement des continua de voisement et d'expressions faciales ouvrent des perspectives. Les termes 'analytique' et 'holistique' utilisés pour parler du traitement perceptif des visages soulignent la coexistence de plusieurs niveaux de traitement du percept. Pour les visages, Schwartzer (2000) a montré que lorsque un visage était présenté à l'envers, les adultes revenaient spontanément à un mode de traitement analytique. Pour le voisement, nous avons vu dans l'introduction qu'en fonction de la consigne donnée, les sujets adultes sont capables de faire des différences intracatégorielles fines, un niveau de traitement qui nécessite une analyse plus précise des composantes acoustiques du stimulus. Pour les couleurs, Bornstien et Korda (1984) parviennent aux mêmes conclusions avec un continuum bleu-vert. Lorsque aucune consigne particulière n'était donnée aux sujets adultes testés, leur perception était plus catégorielle que lorsque la consigne les invitait à une analyse plus fine des différences physiques entre stimuli. Différents niveaux de perception coexisteraient donc pour ces trois types de continua.

Aussi, au vu des ces observations, on peut penser que lorsque Pisoni (1973) pour le voisement, Raskin, Maital et Bornstein (1983) pour les couleurs et Mondloch et al. (2002) pour les expressions faciales évoquent le *développement des habiletés attentionnelles* pour expliquer la plus grande précision avec laquelle les sujets catégorisent chacun de ces types de continua, ils font référence à leur capacité de jongler avec différentes échelles de traitement, plus analytique ou plus intégré selon les exigences de la consigne (ou selon le contexte en situation plus écologique de perception).

La perception du voisement dans le cadre de la théorie du liage

Nous avons choisi d'étudier la perception du voisement dans une perspective développementale. Du nourrisson à l'adulte, les étapes de ce processus dynamique révèlent la complexité de la perception. La complexité inhérente au mécanisme de perception requiert de postuler l'existence de *plusieurs niveaux de traitement* qui, ensemble et de manière intégrée, permettent de penser la perception comme un phénomène unitaire.

TRAITEMENT AUDITIF GENERAL DES INFORMATIONS ACOUSTIQUES

L'enregistrement des potentiels évoqués auditifs nous a permis de comprendre comment les informations temporelles contenues dans le signal sont codées. Concernant l'indice acoustique de DEV, les informations acoustiques les plus saillantes sont le relâchement de l'occlusion et la vibration des cordes vocales. Au niveau cortical, ces informations sont codées par des pics de décharges neuronales synchronisées sur le début de chacun de ces deux événements. Dans notre travail, la morphologie de la composante N100 reflétait ces pics de décharges neuronales. Pour des valeurs courtes de DEV (entre -30 et +30 ms), un seul pic de N100 était observable tandis que pour des valeurs plus longues (au-delà de -30 et +30 ms DEV), deux pics étaient clairement dissociables (étude 3.2). Nous avons toutefois constaté que les tracés évoqués par des DEV positifs présentaient un double pic plus robuste que ceux évoqués par des DEV négatifs.

Actuellement, nous menons une étude pour déterminer si la moindre saillance acoustique du DEV négatif résulte d'un effet de *masquage proactif* i.e. le masquage de la composante de haute fréquence (le relâchement de l'occlusion de la consonne) par la composante de basse fréquence (la vibration des cordes vocales). Afin de démasquer le pic de N100 synchronisé sur le relâchement de l'occlusion, nous avons emprunté à la psychoacoustique le test de

Binaural masking level difference (BMLD) utilisé pour étudier les mécanismes de séparation spatiale des sources sonores. La figure 16 schématise une condition expérimentale dans laquelle on obtient un effet BMLD. Dans la situation A, le signal (qui correspond au son pur représenté par la sinusoïde) et le bruit masquant sont appliqués aux deux oreilles de manière identique alors que dans la condition B, la cible est déphasée de 180° (opposition de phase) dans l'une des deux oreilles. On parlera d'un *effet de démasquage binaural* lorsque le seuil de détection du signal mesuré dans la condition B est plus bas que dans la condition A. Ce démasquage s'explique par un effet de localisation de sources : dans la condition A les localisations spatiales du signal et du bruit sont confondues tandis que dans la condition B, elles sont dissociées.

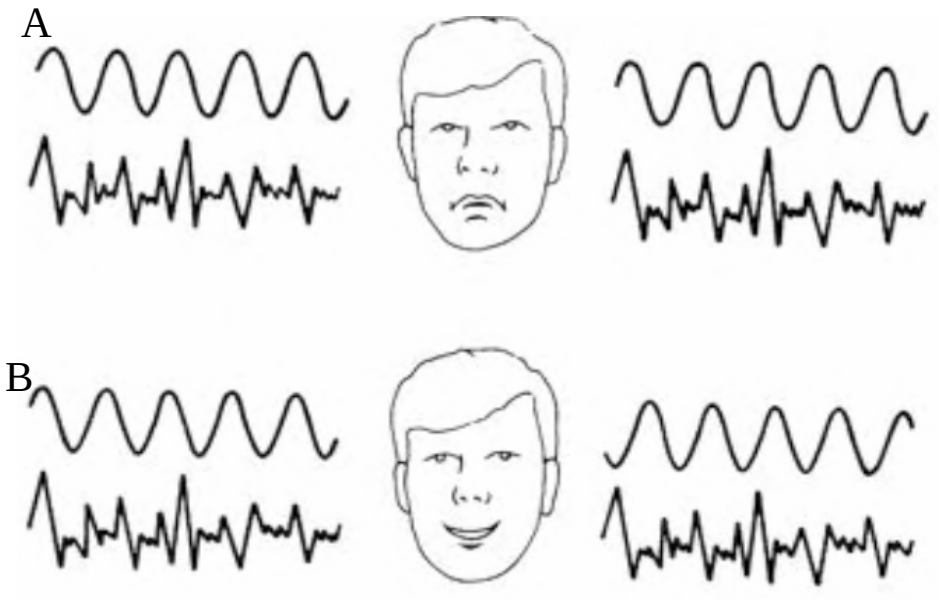

Figure 16 : Illustration de l'effet de démasquage binaural (BMLD). Lorsque le signal et le bruit sont appliqués aux deux oreilles de manière identique (A), les localisations spatiales de la source de chaque son sont confondues. Le déphasage du signal dans l'une des deux oreilles (B) permet au sujet de dissocier la localisation de chacune des deux sources ce qui induit un effet de démasquage du signal.

Nous avons adapté ce test à notre problématique en mettant les barres de voisement des syllabes /də/ (-75, -45, -15 ms DEV) en opposition de phase dans les oreilles du sujet. Ce déphasage a pour conséquence de percevoir le relâchement de l'occlusion et la barre de voisement comme venant de *deux sources spatiales différentes*. Bien que nous n'ayons pas encore réalisé d'analyse statistique sur les tracés obtenus en testant dix sujets francophones

adultes, nous pouvons déjà tirer une conclusion majeure de cette étude : le déphasage de la barre de voisement *ne provoque pas la désintégration du percept* (qui est toujours identifié comme un /də/). Cette observation nous invite à proposer que la perception unitaire de la syllabe /də/ repose sur la relation temporelle entre le relâchement de l'occlusion et la vibration des cordes vocales et non pas sur la localisation spatiale de ces deux événements. Cette proposition fait écho aux résultats observés par Mann et Liberman (1983) et Whalen et Liberman (1987) dans le cadre des travaux sur la *perception duplex* (cf section 2.3.2.2.1. de l'introduction). Bien que les deux premiers formants des syllabes /ba/ et /ga/ aient été présentés dans une oreille et le troisième formant dans l'autre oreille, les sujets testés percevaient non seulement les formants de manière isolée mais aussi les syllabes /ba/ et /ga/, perçues par intégration de l'ensemble des formants.

Nous proposons que la perception unifiée de la syllabe /də/ reflète un niveau d'intégration suprasegmental qui permet aux segments acoustiques constitutifs du DEV (relâchement de l'occlusion et vibration des cordes vocales) d'être liés quelle que soit la durée du DEV et quelles que soient les différences de localisation de source entre relâchement de l'occlusion et vibration des cordes vocales.

Cette observation et les résultats que nous avons obtenus à l'aide des potentiels évoqués nous amènent à aborder la question du passage du simple au double pic dans une *perspective cognitiviste*, celle proposée par la théorie du liage temporel.

TRAITEMENT COGNITIF DES INFORMATIONS ACOUSTIQUES

Treisman et Gelade (1980) proposent dans le cadre de leur théorie sur la perception des objets, que la perception unitaire d'un même objet implique (1) le codage neuronal des différentes propriétés de cet objet et (2) *le liage* (ou 'binding' en anglais) de ces propriétés. Cette dernière étape permettrait d'expliquer que notre vision du monde n'est pas fragmentaire mais unifiée. Le liage ne nécessite pas que les propriétés qui sont intégrées soient issues de la même modalité sensorielle (Damasio, 1989 ; 1990). Certains auteurs (Kaiser, Hertrich, Ackermann, Mathiak, Lutzengerger, 2005) ont d'ailleurs proposé que dans la parole audiovisuelle (voir Colin & Radeau 2003 pour une revue), l'intégration des informations issues des deux modalités sensorielles constituait une forme de liage. Physiologiquement, le liage résulterait de la synchronisation des potentiels d'action émis par les neurones recrutés

dans le codage des multiples attributs (taille, couleur, forme…) d'un même percept. On parle à ce propos *de liage temporel* (Von der Malsburg, 1981) et les neurones qui déchargent de manière synchrone forment quant à eux une *assemblée temporelle*. Comme chaque neurone peut appartenir à plusieurs assemblées temporelles, le processus de liage est dynamique : ce n'est pas le neurone qui code un attribut particulier mais c'est la synchronisation des décharges d'un ensemble de neurones qui crée l'intégration des attributs. La figure 17 illustre le mécanisme de liage temporel.

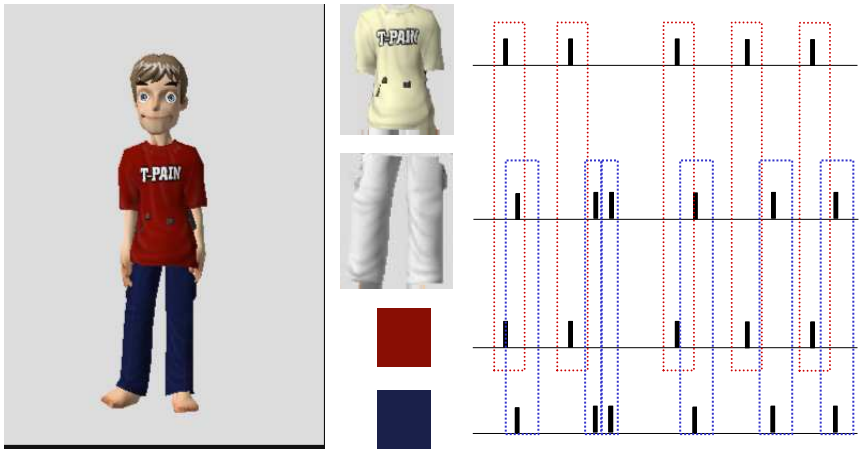

Figure 17 : Hypothèse du liage temporel (selon Meunier, 2007). Les attributs 'couleur' et 'forme' sont intégrés par synchronisation des rythmes de décharge des neurones codant chacun de ces attributs. Par exemple, la perception du tee-shirt rouge vient de la synchronisation des rythmes de décharges des neurones qui codent la perception de la forme du tee-shirt et la perception de la couleur rouge (cadre rouge).

Ce sont les *oscillations internes du cortex* qui sont à l'origine du liage temporel. Par oscillations internes, il faut entendre le rythme oscillatoire spontané du cerveau, i.e. non synchronisé par les caractéristiques fréquentielles ou spectrales des stimuli présentés. Ces oscillations spontanées sont définies par leur bande de fréquence : on distingue ainsi les bandes δ (<3.5 Hz), θ (5-10 Hz), α (10-20 Hz), β (20-30 Hz), γ (20-80 Hz). Par exemple, pendant le sommeil profond, ce sont les oscillations δ qui dominent tandis que pendant le sommeil paradoxal apparaissent des oscillations γ. On retrouverait ces mêmes oscillations γ lorsque le sujet est engagé dans une tâche perceptive. Par exemple, en présentant des paires de clics, Joliot, Ribary et Llinas (1994) ont enregistré une activité oscillatoire de fréquence 40 Hz (bande γ) lorsque les clics étaient séparés par moins de 15 ms. Au-delà de ce seuil, deux

activités oscillatoires à 40 Hz étaient enregistrées : entre 15 ms et 25 ms, elles se superposaient dans le temps puis se dissociaient clairement au-delà de 25 ms. Joliot et al. ont conclu de cette étude que le temps nécessaire pour identifier et ordonner deux événements auditifs distincts était lié à la *réinitialisation* des oscillations γ. La présentation d'un son provoquerait la mise en phase de l'activité oscillatoire spontanée la plus adaptée, i.e. la fréquence oscillatoire (θ, α, γ) la plus proche du rythme des informations temporelles contenues dans le son présenté (Shroeder & Lakatos, 2009). Par ailleurs, à l'intérieur d'un son, la présence d'événements acoustiques transitoires (i.e. brefs) réinitialiserait l'activité oscillatoire spontanée du cortex pour se synchroniser sur les événements acoustiques présents dans le stimulus (Engel, Fries, König, Brecht & Singer, 1999).

Le concept de liage a été appliqué à de nombreux domaines aussi divers que la reconnaissance, l'attention, la mémorisation et la récupération d'information en mémoire, l'intégration sensori-motrice, les inférences logiques et l'analyse de la parole (Engel & Singer, 2001). Poeppel (2003) se concentre sur cette dernière application. Partant d'évidences pyschophysiques et neuropsychologiques, Poeppel propose la théorie de l'échantillonnage asymétrique (« Assymetric sampling in time », AST) qui repose sur deux prémisses :

> (1) l'analyse du signal de parole s'effectue sur différentes échelles de temps qui correspondent à différents niveaux de traitement (i.e. on ne peut pas placer sur une même échelle temporelle l'analyse des indices acoustiques de la parole, le découpage de la parole en unité syllabique et l'analyse de la prosodie).
>
> (2) à un premier niveau d'analyse, le traitement du signal de parole fait intervenir les deux hémisphères cérébraux et ce n'est qu'à partir d'un niveau plus élaboré de représentation qu'on observe une latéralisation hémisphérique du traitement.

Dans la droite ligne de la théorie de Tallal qui a montré que l'hémisphère gauche était spécialisé dans le traitement des informations acoustiques rapides (e.g. Schwartz & Tallal, 1980), la théorie AST postule qu'au niveau du traitement des informations temporelles contenues dans le signal de parole, l'hémisphère gauche traite préférentiellement les informations contenues dans des fenêtres temporelles courtes (entre 20 et 40 ms) tandis que l'hémisphère droit est spécialisé dans le traitement d'informations contenues dans des fenêtres temporelles plus longues (150-250 ms). La taille de ces fenêtres temporelles est définie par le rythme des oscillations neuronales spontanées. Dans l'hémisphère gauche, qui traite préférentiellement les informations acoustiques rapides, on enregistre une activité γ tandis que

l'hémisphère droit, qui traite les informations acoustiques plus lentes, est caractérisé par une forte activité oscillatoire θ et α. En se synchronisant sur les différentes composantes acoustiques du signal de parole, ces activités oscillatoires spontanées définissent de par leur rythme des fenêtres temporelles plus ou moins longues dans lesquelles les paramètres acoustiques du langage sont liés. En 1996, Poeppel, Yellin, Phillips, Roberts, Rowley, Waxler et Marantz ont montré que l'écoute passive de syllabes de type Consonne-Voyelle activait les deux hémisphères tandis que la tâche d'identification des phonèmes initiaux (/d/ vs /t/ et /b/ vs /p/) activait préférentiellement l'hémisphère gauche : la latéralisation de l'activité cérébrale était donc contrainte par la nature de la tâche. Dans la conclusion de l'article de 2003, Poeppel revient sur ce résultat et propose que la tâche définit la fenêtre temporelle dans laquelle elle peut être réalisée. Par exemple, si la tâche impose de traiter les informations associées aux transitions de formants (une tâche d'identification sur un continuum de lieu d'articulation par exemple), la fenêtre temporelle d'analyse sera courte, i.e. entre 20 et 40 ms. L'hémisphère gauche sera donc sollicité.

Cette latéralisation du traitement des informations temporelles est corroborée par les données comportementales obtenues par Laguitton et al. (2000). Dans cette étude, des sujets adultes francophones réalisaient une tâche d'identification de syllabes voisées et non voisées issues de trois continua différents : bilabial (/ba-pa/), apical (/da-ta/) et vélaire (/ga-ka/). L'analyse des temps de réaction a montré un avantage de l'oreille droite (donc de l'hémisphère gauche) pour traiter les continua bilabial et apical, i.e. les continua caractérisés par une frontière phonologique courte en français (inférieures à 20 ms DEV). La perception des stimuli issus du continuum vélaire, pour lequel la frontière phonologique est localisée à une valeur de DEV supérieure à 20 ms, n'a donné lieu à aucune supériorité de l'oreille droite. De ces résultats, Laguitton et al. ont conclu que l'hémisphère gauche jouait un rôle majeur dans l'analyse des informations temporelles rapides.

Les données électrophysiologiques enregistrées en intracortical chez des sujets francophones par cette même équipe (Liégeois-Chauvel et al., 1999) vont dans le même sens. Les syllabes /ba-da-ga/ voisées et /pa-ta-ka/ non voisées caractérisées par un DEV moyen de -110 ms et 20 ms respectivement n'étaient pas codées de la même manière dans les deux hémisphères cérébraux. Les informations acoustiques contenues dans le DEV étaient traitées de manière séquentielle dans l'hémisphère gauche et de manière holistique dans l'hémisphère droit. Les données électrophysiologiques présentées dans cette thèse montrent aussi un effet de

latéralisation. Le complexe T enregistré au niveau temporal droit (T8) et temporal gauche (T7) avaient une morphologie différente. Dans notre étude 3.2, la différence de latence entre les deuxième pics de Na obtenus avec les stimuli séparés par +75 et +45 ms DEV ne différait pas significativement de 30 ms (i.e. la durée du pas acoustique) en T7 contrairement à cette même différence calculée en T8 (temporal droit).

L'ensemble de ces recherches nous amène à proposer que la valeur absolue des frontières de voisement (30 ms DEV) correspond à la durée de la fenêtre temporelle dans laquelle les informations acoustiques sont intégrées. En d'autres termes, les frontières universelles de voisement sont situées entre deux fenêtres temporelles dans lesquelles les informations sont intégrées.

Si les frontières universelles de voisement sont contraintes par des mécanismes auditifs généraux, eux-mêmes contraints par les activités oscillatoires du cerveau, qu'en est-il des frontières phonologiques ?

Ward (2003) propose une vision dynamique et unifiée des relations entre la synchronisation des oscillations neuronales spontanées et les fonctions cognitives. En faisant une revue de la littérature du domaine, elle propose d'associer les processus *d'encodage et de rappel mnésique* à la synchronisation des rythmes oscillatoires θ et γ tandis que *l'attention* serait corrélée aux activités oscillatoires α et γ. Tant pour la mémorisation que pour l'attention, l'activité oscillatoire rapide γ est modulée par une activité oscillatoire plus lente (θ ou α), i.e. à chaque début de cycle lent, l'activité oscillatoire rapide est réinitialisée (figure 18).

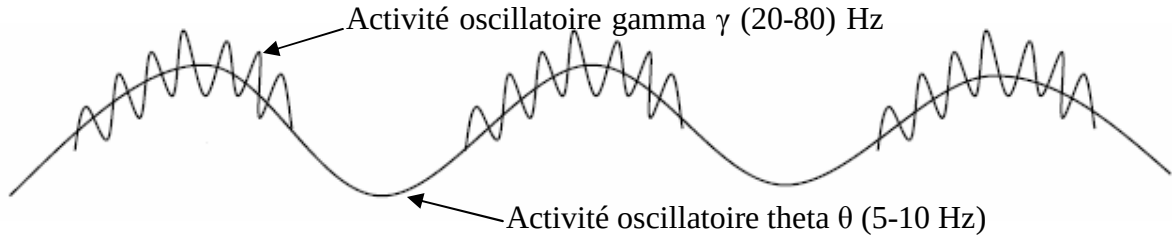

Figure 18 : modulation de l'activité γ en fonction de l'activité θ. Adapté de Ward (2003).

Cette dépendance entre rythmes oscillatoires spontanés serait à l'origine de différents niveaux hiérarchiques d'intégration (Engel & Singer, 2001). En effet, les fenêtres d'intégration

associées au rythme oscillatoire γ sont courtes (une période de cycle dure de 12 à 50 ms) tandis que celles associées au rythme oscillatoire θ (une période de cycle dure de 100 à 200 ms) et α (une période de cycle dure de 50 à 100 ms) sont plus longues. La présentation d'un stimulus provoquerait la synchronisation des oscillations lentes (θ et α) sur lesquelles se synchroniseraient elles-mêmes les oscillations plus rapides (γ). A un premier niveau de traitement, les différentes composantes d'un même stimulus seraient analysées de manière distribuée dans plusieurs fenêtres temporelles d'intégration courtes. Ces représentations distribuées seraient ensuite intégrées à un niveau plus élevé de traitement au sein de fenêtres d'intégration plus longues. Le premier niveau correspondrait à une intégration en parallèle des attributs isolés au sein de chaque fenêtre courte d'intégration tandis que le deuxième niveau correspondrait à une intégration sur une plus large échelle. La perception serait donc une fonction supérieure nécessitant l'intégration d'un grand nombre d'unités plus simples en interaction.

Nous proposons d'appliquer cette double hiérarchisation des niveaux de traitement cognitifs et des oscillations spontanées de différentes fréquences à la perception du voisement. Les figures 19 et 20 illustrent la spécialisation phonologique en anglais et en français.
En anglais, les productions sont toutes caractérisées par un DEV positif. Tous contextes confondus la moyenne de production du DEV des phonèmes /b-d-g/ est de 15 ms et celle des /p-t-k/ de 70 ms (Lisker & Abramson, 1967). Au niveau perceptif, la frontière phonologique de voisement est proche de 30 ms. Le fait que les phonèmes voisés et non voisés soient toujours produits avec un DEV positif implique qu'au niveau perceptif les locuteurs anglophones soient capables de déterminer de manière précise la valeur du DEV : le phonème est voisé pour des DEV positifs courts (< 30 ms) et non voisé pour des DEV positifs longs (> 30 ms). Cette précision temporelle impose au système perceptif d'analyser le flux de parole à une échelle où ces différences de délai de DEV sont encodées. En termes physiologiques, un DEV de 30 ms correspond à une fenêtre d'intégration courte, i.e. une période de cycle de fréquence γ. Lorsque le relâchement de l'occlusion et le début de la vibration des cordes vocales sont séparées par moins de 30 ms, les oscillations gamma ne sont pas réinitialisées sur le deuxième événement acoustique (qui correspond toujours au début de la vibration des cordes vocales en anglais), i.e. les deux événements sont intégrés dans une même fenêtre temporelle. Par contre, lorsque le relâchement de l'occlusion et le début de la vibration des cordes vocales sont séparés par plus de 30 ms, chacun des deux événements donne lieu à une

synchronisation des oscillations de fréquence γ. La frontière phonologique de l'anglais serait donc située entre deux frontières d'intégration courtes.

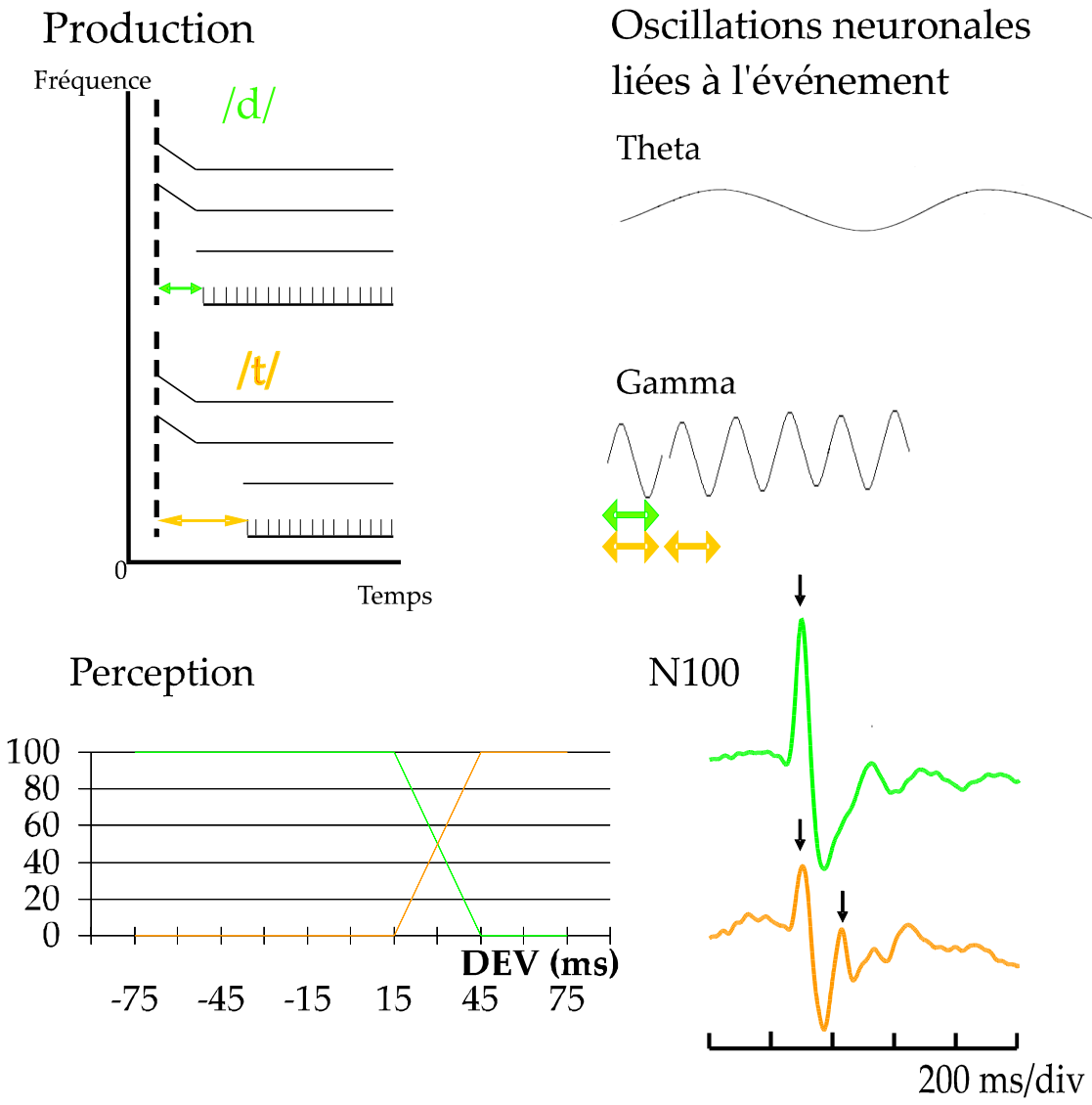

Figure 19 : Perception et production du voisement en anglais et corrélats neurophysiologiques associés (oscillations neuronales et potentiels évoqués).

En français (figure 20), les productions sont caractérisées par un DEV négatif et long (moyenne : -100 ms) pour les phonèmes voisés et positif et court (moyenne : 30 ms) pour les phonèmes non voisés (Serniclaes, 1987). Au niveau perceptif, la frontière phonologique de voisement est proche de 0 ms. Ce n'est donc plus la longueur du DEV qui constitue l'information nécessaire à la catégorisation mais *l'ordre d'apparition* du relâchement de

l'occlusion et du début de la vibration des cordes vocales : lorsque la vibration des cordes vocales précède le relâchement de l'occlusion, le phonème est voisé, lorsque l'ordre est inversé, le phonème est perçu comme non voisé. Les phonèmes non voisés peuvent être produits en français aussi bien avec un DEV de 20 ms, qu'avec un DEV de 40 ms (donc de part et d'autre de la frontière universelle à +30 ms) sans que cela n'ait de répercussions sur la nature du phonème perçu (/tə/ dans les deux cas). Pour ne pas prendre en compte la variabilité temporelle des informations contenues dans les fenêtres d'intégration courtes (disponibles grâce aux oscillations γ), les sujets francophones analyseraient le flux de parole sur des périodes d'intégration plus longues disponibles grâce aux oscillations de fréquence plus lentes (θ). Ce mode de traitement leur permettrait d'éviter de discriminer des variations allophoniques des phonèmes, i.e. des différences de durées de DEV qui ne sont pas phonologiques en français.

Ces réflexions nous amènent à proposer que la dispersion des distributions de DEV propres à chaque langue impose une échelle d'analyse préférentielle. Les anglophones, qui sont sensibles à la longueur du DEV, découperaient le signal de parole en fenêtres temporelles courtes tandis que les francophones, qui sont sensibles à l'ordre d'apparition des événements acoustiques du DEV, traiteraient ce même signal dans des fenêtres d'analyse plus longues, ce qui leur permettraient de ne pas considérer comme phonologiques de simples variations allophoniques.

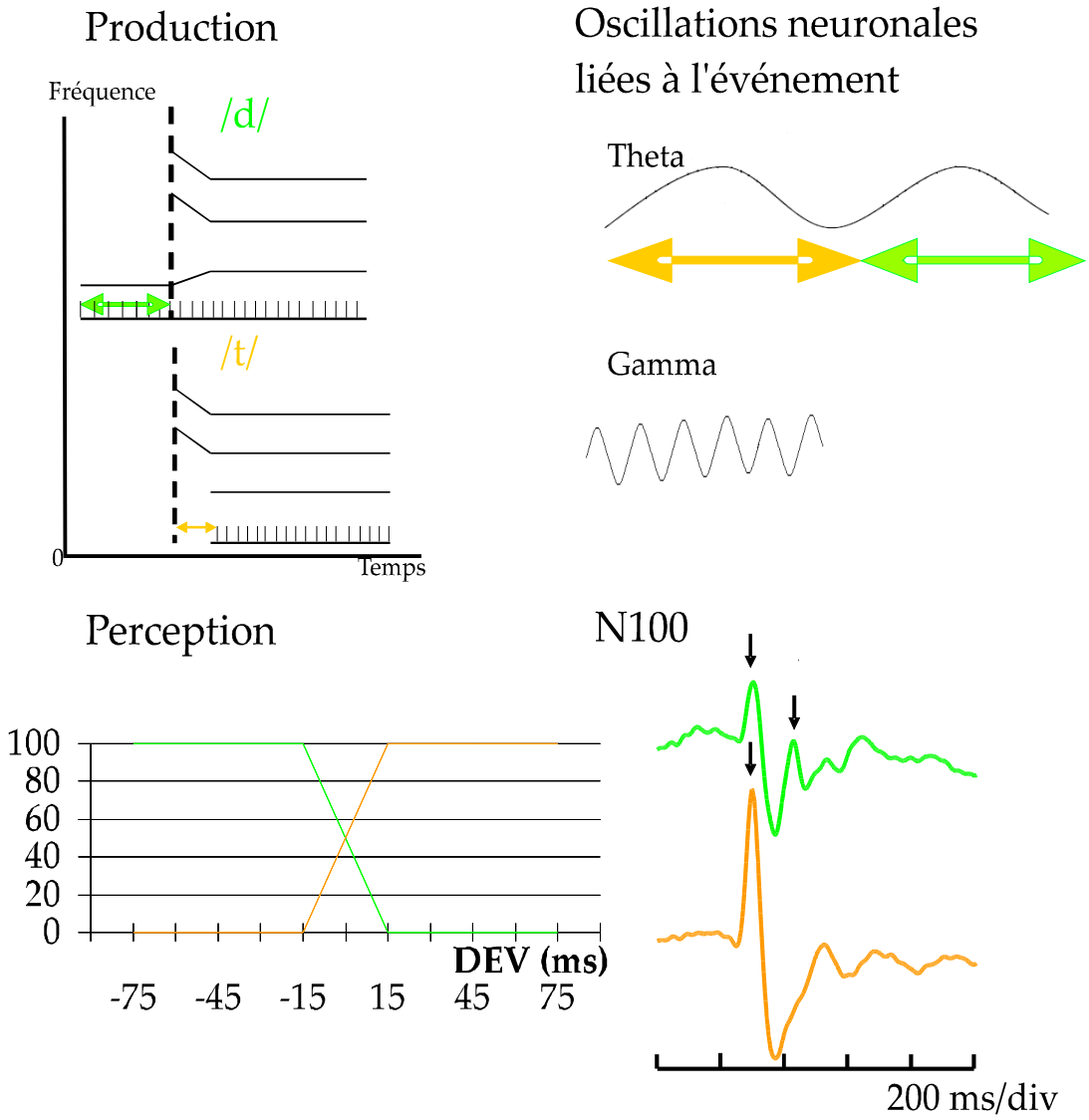

Figure 20 : Perception et production du voisement en français et corrélats neurophysiologiques associés (oscillations neuronales et potentiels évoqués).

Bien que le processus de spécialisation phonologique amène les sujets à privilégier une fenêtre d'analyse plus ou moins longue, ces différentes échelles de traitement pourraient coexister. Comme mentionné plus haut, les oscillations plus courtes de fréquence γ sont en effet modulées par les oscillations de fréquence θ (Engel & Singer, 2001). L'intégration des oscillations gamma rapides (de l'ordre de fréquence de 30 Hz), au moyen d'une fréquence « porteuse » theta plus lente (de l'ordre de 4 Hz, soit une fenêtre d'intégration de 250 ms) pourrait constituer une forme de couplage comme celui décrit dans l'introduction (Serniclaes, 1987, 2000).

La réinitialisation synchronisée des oscillations de différentes fréquences permettrait d'analyser la parole continue sur plusieurs échelles de temps : une échelle courte pour analyser des changements acoustiques précis, une échelle plus longue pour analyser les modulations lentes de l'enveloppe du son de parole (Poeppel, 2003). Un argument à l'appui de la coexistence de plusieurs échelles d'analyse vient de nos données comportementales : bien que les frontières universelles ne soient plus pertinentes pour discriminer le voisement en français, on observe chez la plupart des sujets francophones un pic de discrimination pour une valeur de DEV de 30 ms.

Dans les figures 19 et 20, nous avons représenté les potentiels évoqués par les syllabes /də/ et /tə/ chez des sujets anglophones et francophones. Chez les anglophones, les valeurs de DEV positives courtes (perçues comme /d/) évoque un seul pic de N100 tandis que les DEV positifs longs (perçus comme /t/) évoquent deux pics de N100 (Sharma & Dorman, 1999). En français, deux pics de N100 sont évoqués pour des DEV négatifs longs (perçues comme /d/) et un seul pic pour des DEV positifs courts (perçus comme /t/) (étude 3.2.).
A notre connaissance aucune théorie n'explique le lien entre la morphologie des potentiels évoqués et les mécanismes de synchronisation que nous venons d'évoquer. De ce fait, et bien que la théorie du liage temporel procure un moyen de sélectionner certains délais par rapport à d'autres, il n'y a pas de raison suffisante pour évacuer l'idée que d'autres mécanismes physiologiques soient responsables de la morphologie des potentiels évoqués. Les effets de réfraction neuronale et de masquage temporel dont nous avons fait mention dans les études 3.1 et 3.2 constituent deux exemples de ces mécanismes physiologiques. Il serait donc intéressant de parvenir à dissocier ce qui est du ressort des mécanismes d'intégration de ce qui a trait à des non-linéarités d'origine physiologique.

Mais, avant de déterminer le lien entre oscillations cérébrales spontanées et potentiels évoqués, une première question est de déterminer s'il existe un corrélat neurophysiologique de la frontière phonologique du voisement en français (0 ms DEV) ?
Dans plusieurs études, il a été montré que la négativité de discordance (MMN : Mismatch Negativity ; Näätänen, Gaillard & Mäntysalo, 1978) est un potentiel évoqué qui indexe la perception des contrastes phonologiques. La MMN est une composante négative dont la latence d'apparition varie entre 100 et 200 ms. Cette composante « indexe la détection automatique, pré-attentive, d'une disparité physique entre un stimulus déviant présenté dans

une séquence homogène de stimuli standards » (Colin, 2001, p.77). Bien que toutes deux soient pré-attentives, la N100 est un potentiel *exogène*, i.e. évoqué par une cause externe (le stimulus présenté) tandis que la MMN est un potentiel *endogène*, i.e. évoqué par une cause interne (le hiatus provenant de la comparaison entre la trace mnésique d'un stimulus standard et le stimulus déviant).

Sharma et Dorman (2000) ont comparé la morphologie des composantes N100 et MMN enregistrées chez des sujets anglophones et chez des sujets bilingues hindi/anglais. Les syllabes /ba-pa/ présentées aux sujets variaient sur un continuum de DEV de -90 à 0 ms. Alors que les sujets anglophones monolingues percevaient tous les stimuli de ce continuum comme un /ba/, les sujets bilingues hindi/anglais percevaient deux percepts : un /ba/ lorsque le DEV variait entre -90 et -20 ms et un /pa/ entre -20 et 0 ms. Les résultats de cette étude ont montré que la lange parlée par le sujet n'avait aucune incidence sur la morphologie de la composante N100. Par contre, l'amplitude de la MMN était plus importante pour le groupe des bilingues que pour le groupe des monolingues lorsque le contraste était situé de part et d'autre de la frontière phonologique de l'hindi (proche de -20 ms). Sharma et Dorman (2000) proposent dans leurs conclusions que la N100 et la MMN reflètent deux modes de traitement différents : *acoustique* pour la première composante et *phonologique* pour la seconde. Les données développementales de Cheour, Shestakova, Alku, Ceponiene, Näätänen (2002) vont dans le même sens : alors qu'aucune MMN n'est enregistrée chez des enfants de langue finnoise âgés de 4 à 6 ans à qui on a présenté un contraste de voyelles françaises, une MMN est évoquée chez ces mêmes enfants peu après leur arrivée dans une école maternelle francophone.

Avec les mêmes stimuli que nous avons utilisés pour enregistrer la N100 (/də/ et /tə/, DEV de -75 à +75 ms avec un pas acoustique de 30 ms), Jacques (2007), pour son mémoire de fin d'étude, a enregistré dans notre laboratoire les potentiels évoqués par ces stimuli chez trois groupes de sujets adultes francophones : un groupe de 6 sujets musiciens, un groupe de 4 ingénieurs du son ainsi qu'un groupe de 6 sujets de contrôle afin d'analyser les influences de l'expertise musicale sur les habiletés de perception catégorielle du voisement. Les potentiels évoqués par ces syllabes ont été enregistrés dans deux types de paradigmes, un paradigme « *simple* » qui consiste à présenter chaque stimulus isolément (au total 15500 fois) et un paradigme dit « *oddball* ». Dans ce dernier paradigme, les stimuli sont présentés par paires (-75 ms vs -45 ; -45 vs -15 ; -15 vs +15 ; +15 vs +45 ; +45 vs +75 et inversement) dont le premier stimulus est le standard et le deuxième le déviant. Seule la paire -15/+15 ms testait la

sensibilité à la frontière phonologique du français. Dans une séquence de 700 stimuli, le stimulus standard était présenté 600 fois (85%) et le déviant 100 fois (15%). La MMN était obtenue pour chaque stimulus en faisant la différence entre les potentiels évoqués par un stimulus dans le paradigme oddball et les potentiels évoqués par ce même stimulus dans le paradigme simple. Malgré le faible nombre de sujets, une analyse de variance a été réalisée avec comme variable dépendante l'amplitude de la MMN et comme variables indépendantes le groupe et la paire de stimuli. Les résultats n'ont pas mis en évidence un effet du groupe. Par contre l'effet de la paire de stimuli était proche de la signification (avec une amplitude plus grande pour le contraste phonologique (centré sur 0 ms) que pour tous les autres contrastes. **Bien que cette étude doive être poursuivie sur un plus grand nombre de sujets, ces données préliminaires, obtenues avec les mêmes stimuli que ceux utilisés dans cette thèse, suggèrent que la MMN constitue le corrélat neurophysiologique de la frontière phonologique du voisement en français.**

L'une des objections qui pourraient être faite au regard de cette proposition théorique est que la MMN n'est pas seulement évoquée par les contrastes phonologiques mais par tous les contrastes acoustiques. Notons cependant que l'amplitude de la MMN phonologique est plus importante que celle de la MMN acoustique. Näätänen et Alho (1997) suggèrent que le mécanisme mnésique sous jacent à ces deux types de MMN est différent. Une preuve empirique de cette différence est que la MMN évoquée par des contrastes acoustiques non phonologiques n'est pas latéralisée tandis que la MMN évoquée par un contraste phonologique est latéralisée dans l'hémisphère gauche. La MMN acoustique serait évoquée par comparaison entre le stimulus et la trace mnésique sensorielle laissée par le stimulus précédent (*mémoire échoïque*). La MMN phonologique, quant à elle, requerrait la comparaison entre le stimulus et la trace mnésique de ce stimulus stockée en *mémoire à long terme* (Näätänen, Paavilainen, Rinne & Alho, 2007). Ce dernier point nous permet de faire le lien avec les théories développementales et notamment avec le NLM-e (Kuhl et al., 2008 ; cf partie 2.4.2.3 de l'introduction). Dans ce cadre théorique, l'encodage d'un stimulus en mémoire à long terme viendrait de l'exposition du sujet aux *prototypes* de sa langue. Cette exposition aux prototypes aurait pour conséquence de renforcer le lien établi entre pattern perçu et pattern produit et donc de créer une trace mnésique en mémoire à long terme. Ce lien entre perception et production a notamment été souligné par Serniclaes, d'Alimonte et Alegria (1984) qui ont montré que les locuteurs sourds francophones testés dans cette étude avaient

des difficultés à produire des contrastes de voisement caractérisés par une valeur de DEV négative ; un résultat qui s'expliquerait par l'incapacité de ces sujets à réajuster leurs productions en fonction du retour acoustique qui en découle.

Conclusion et perspectives

« Ne comprends-tu donc pas que le moindre oiseau qui fend l'air est un immense monde de délices fermé par tes cinq sens » (p.18). Cette constatation faite par William Blake (1790) est complétée quelques pages après par cet aphorisme : « Si les fenêtres de la perception étaient nettoyées, chaque chose apparaîtrait à l'homme, -ainsi qu'elle l'est-, infinie » (p. 36). Nous avons commencé cette dissertation en posant une question : Que nous est-il donné de percevoir ? Blake y répond en insistant sur le caractère *contraint* de la perception.

En se concentrant sur la perception d'une réalité tout à fait particulière - le voisement -, nous avons présenté la perception comme une activité contrainte. Dès la naissance, le nourrisson discrétise le continuum de voisement autour de deux frontières universelles dont la localisation (-30 et +30 ms) ne semble pas aléatoire. Nos données électrophysiologiques montrent que la morphologie (simple vs double pic) de la composante N100 constitue le corrélat physiologique de ces frontières universelles. Dans la discussion, nous avons jugé opportun d'analyser ces résultats à l'aune de la théorie du liage temporel et proposé que les activités oscillatoires spontanées du cerveau jouaient un rôle dans la définition des différentes échelles d'analyse du flux de parole. Spécifiquement, nous avons proposé que les frontières universelles de voisement étaient localisées entre deux fenêtres temporelles d'intégration courtes dont la période était définie par le rythme des oscillations γ.

Mais, acceptons avec Elman, Bates, Johnson, Karmiloff-Smith, Parisi et Plunkett (1996) que le processus de spécialisation phonologique est aussi *dynamique* : « […] developmental timing becomes more crucial as the hierarchical complexity of an ontogenetic system increases. And genes are algorithms which operate sequentially in time » (p. 16). En étudiant le développement de la perception du voisement chez des francophones, nous avons observé qu'autour de six mois, l'espace perceptif est remodelé en fonction des productions disponibles dans l'environnement de l'enfant. En français, cette redéfinition de l'espace perceptif passe par l'émergence d'une nouvelle frontière, située à mi-distance des frontières universelles (0 ms). En anglais, le processus de spécialisation phonologique est tout autre puisqu'il conduit à la désactivation de l'une des frontières universelles (-30 ms) tandis que l'autre reste active (+30 ms). Par ailleurs, nous avons constaté que jusqu'à l'âge adulte, les transactions entre le substrat cérébral plastique de l'enfant et son environnement sont constantes et donnent lieu à un affinement des habiletés perceptives. Les données électrophysiologiques présentées dans la

littérature et les résultats de Jacques (2007) obtenus avec les mêmes stimuli que ceux que nous avons utilisés dans cette thèse nous ont amené à proposer que la MMN constitue le corrélat physiologique de la frontière phonologique du voisement en français. Dans le futur, cette hypothèse devra faire l'objet de nouvelles études pour être vérifiée.

Sans amoindrir cette vision dynamique du processus de spécialisation phonologique, il faut cependant constater que, malgré la diversité des environnements dans lesquels ce processus peut avoir lieu, la frontière de voisement n'est pas localisée n'importe où sur le continuum de voisement. Dans l'échantillon des langues qui ont fait l'objet d'études, l'espace perceptif du voisement est divisé en deux catégories et dans la plupart de ces langues la frontière phonologique est située à 0 ms DEV. On peut dès lors se poser la question du mécanisme neurophysiologique sous jacent à cette frontière. Dans la discussion de cette dissertation, nous avons évoqué les travaux de Poeppel (2003) selon lequel le flux de parole peut être analysé sur différentes échelles. Selon le type d'informations acoustiques nécessaires pour catégoriser les sons de parole, le système perceptif adapterait sa fenêtre d'analyse. Pour catégoriser les phonèmes voisés vs non voisés, les sujets francophones ont besoin de déterminer lequel des deux événements acoustiques (relâchement de l'occlusion et début de la vibration des cordes vocales) a été produit en premier tandis que cette même catégorisation requiert de la part des sujets anglophones de mesurer le délai de séparation de ces deux événements acoustiques. Nous avons proposé que la distribution des productions auxquelles le nouveau-né est exposé l'amenait à privilégier un mode de perception analytique ou intégré. Les informations acoustiques seraient intégrées dans des fenêtres temporelles longues (grâce aux oscillations spontanées de rythme θ) chez le sujet francophone et dans des fenêtres temporelles plus courtes (grâce aux oscillations spontanées de rythme γ) chez le sujet anglophone. Le processus de spécialisation phonologique consisterait dès lors à privilégier l'échelle d'analyse la plus adéquate au regard de la statistique des productions de la langue à laquelle l'enfant est exposé. Malgré ce mode de traitement privilégié, plusieurs échelles de traitement peuvent coexister grâce au fait que les différents rythmes oscillatoires du cerveau sont assujettis les uns aux autres (e.g. modulation de l'activité γ en fonction de l'activité θ).

Au terme de ce travail, nous mesurons combien les différentes propositions faites dans cette conclusion sont spéculatives, et nécessitent d'être testées. Il nous apparaissait toutefois intéressant, étant donné le caractère intrinsèquement multidisciplinaire des études sur la

perception, de tenter de proposer une théorie unifiée du processus de spécialisation phonologique. Les choix méthodologiques que nous avons faits nous ont amené à considérer le processus de spécialisation phonologique sous divers angles. Les données comportementales révèlent en effet ce qu'il nous est donné de percevoir tandis que les données physiologiques reflètent les mécanismes intimes de cette perception et les contraintes qui la limitent.

Pour poursuivre ce travail, il serait intéressant de continuer à utiliser de concert des méthodologies comportementales et neurophysiologiques. Les questions qu'ont suscitées les résultats obtenus dans ce travail de thèse ont mis en évidence l'intérêt d'intégrer le rôle des rythmes cérébraux dans la méthodologie et le raisonnement liés à la compréhension du processus de spécialisation phonologique. De par leur haute résolution temporelle, les potentiels évoqués constituent une méthodologie de choix pour comprendre comment les informations temporelles du DEV sont codées au niveau neuronal. Il serait intéressant à présent de mieux comprendre la relation entre les potentiels évoqués et les mécanismes de synchronisation des oscillations neuronales. Mécanismes indépendants ou complémentaires ?

Références bibliographiques

Abramson, A.S. (1977). Laryngeal timing in consonant distinctions. *Phonetica, 34,* 295-303.

Abramson, A.S. & Lisker, L. (1968). Voice timing: Cross-language experiments in identification and discrimination. *Journal of the Acoustical Society of America, 44,* 377.

Abramson, A.S. & Lisker, L. (1970). Discriminability along the voice onset time continuum: Cross-language tests. In B. Hala, M. Romportl and P. Janota (eds.) *Proceedings of the 6th International Congress of Phonetic Sciences, Prague 1967,* Prague: Academia, 569-573.

Abramson, A.S. & Lisker, L. (1973). Voice timing perception in Spanish word initial stops. *Journal of Phonetics, 1,* 1-8.

Anderson, J.L., Morgan, J.L. & White, K.S. (2003). A statistical basis for speech sound discrimination. *Language and Speech, 46,* 155-182.

Aslin, R.N. & Pisoni, D.B (1980). Some developmental processes in Speech perception. In G.H. Yeni-Komshian, J.F. Kavanagh and C.A. Ferguson (eds.), *Child Phonology-Volume 2-Perception*, New-York: Academic Press, 67-96.

Aslin, R.N. & Smith, L.B (1988). Perceptual development. *Annual Review of Psychology, 39,* 435-473.

Aslin, R.N., Pisoni, D.B., Hennessy, B.L. & Perey, A.J. (1981). Discrimination of Voice Onset Time by human infants: new findings and implications for the effects of early experience. *Child Development,* 52, 1135-1145.

Aslin, R.N., Werker, J.F. & Morgan, J.L. (2002). Innate phonetic boundaries revisited. *Journal of the Acoustical Society of America, 112,* 1257-1260.

Bartoschuk, A.K. (1962). Human neonatal cardiac acceleration to sound: habituation and dishabituation. *Perceptual and Motor Skills, 15,* 15-27.

Best, C.T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J.C. Goodman and H.C. Nusbaum (eds). *The development of speech perception: The transition from speech sounds to spoken words,* Cambridge, Massachussets : MIT Press, 167-224.

Best, C.T. & McRoberts, G.W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech, 46,* 183-216.

Beaupré, M. & Hess, U. (2005). Montreal Set of Facial Displays of Emotions.

Blake, W. (1790). *Mariage du ciel et de l'enfer*. Traduction française de André Gide. Collection Romantique N°2.

Bornstein, M.H., Kessen, W. & Weiskopf, S. (1976). The categories of hues in infancy. *Science, 191*, 201-202.

Bornstein, M.H. & Korda, N.O. (1984). Discrimination and matching within and between hues measured by reaction times: some implications for categorical perception and levels of information processing. *Psychological Research, 46*, 207-222.

Browman, C.P. & Goldstein, L. (1987). Tiers in articulatory phonology, with some implications for casual speech. *Status report on speech research, 92*, 1-31.

Bruyer, R., Granato, P. & Van Gansberghe, J.P. (2007). Un score individuel de reconnaissance d'une série de stimuli intermédiaires entre deux sources: la perception catégorielle des expressions faciales. *Revue européenne de psychologie appliquée, 57*, 37-49.

Budd, T.W., Barry, R.J., Gordon, E., Rennie, C. & Michie, P.T. (1998). Decrement of the N1 auditory event-related potential with stimulus repetition: habituation vs. refractoriness. *International Journal of Psychophysiology, 31*, 51-68.

Burnham, D.K. (1986). Developmental loss of speech perception: Exposure to and experience with a first language. *Applied Psycholinguistics,* 7, 207-240.

Burnham, D.K. (2003). Language specific speech perception and the onset of reading. *Reading and writing: an interdisciplinary journal, 16*, 573-609.

Burnham, D.K., Earnshaw, L.J. & Clark, J.E. (1991). Development of categorical identification of native and non-native bilabial stops: Infants, children and adults. *Journal of Child Language, 18,* 231-260.

Burns, T.C. & Ward, W.D. (1978). Categorical perception-phenomenon or epiphenomenon : evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America, 63*, 456-468.

Burns, T.C., Yoshida, K.A., Hill, K. & Werker, J.F. (2007). The development of phonetic representation in bilingual and monolingual infants. *Applied Psycholinguistics, 28,* 455-474.

Buzsaki, G. & Draghun, A. (2004). Neuronal oscillations in cortical networks. *Science, 304*, 1926-1929.

Caramazza, A. & Yeni-Komshian, G.H. (1974). Voice Onset Time in two French dialects. *Journal of Phonetics, 2*, 239-245.

Caramazza, A., Yeni-Komshian, G.H., Zurif, E. & Carbone, E. (1973). The acquisition of a new phonological contrast: the case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America, 54,* 421-428.

Carden, G., Levitt, A., Jusczyk, P.W. & Walley, A. (1981). Evidence for phonetic processing of cues to place of articulation: Perceived manner affects perceived place. *Perception and Psychophysics, 29,* 26-36.

Carney, A.E., Widin, G.P. & Viemeister, N.F. (1977). Non categorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America, 62,* 961-970.

Changeux, J.P., Courrège, P. & Danchin, A. (1973). A theory of the epigenesis of neural networks by selective stabilization of synapses. *Proceedings of the Natural Academy of Sciences, 70,* 2974-2978.

Chatrian, G.E., Lettich, E. & Nelson, P.L. (1985). Ten percent electrode system for topographic studies of spontaneous and evoked EEG activity. *American Journal of EEG Technology, 25,* 83-92.

Chen, G.D., Nuding, S.C., Narayan, S.S. & Sinex, D.G. (1996). Responses of single neurons in the chinchilla inferior colliculus to consonant-vowel syllables differing in voice-onset-time. *Auditory Neuroscience, 3,* 179-198.

Chen, G.D. & Sinex, D.G. (1999). Effects of interaural time differences on the responses of chinchilla inferior colliculus neurons to consonant-vowel syllables. *Hearing Research, 138,* 29-44.

Cheour, M., Shestakova, A., Alku, P., Ceponiene, R. & Näätänen, R. (2002). Mismatch negativity shows that 3-6 years-old children can learn to discriminate non-native speech sounds within two months. *Neuroscience Letters, 325,* 187-190.

Clayards, M., Tanenhaus, M.K., Aslin, R.N. & Jacobs, R.A. (2007). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition, 108,* 804-809.

Colin, C. (2001). *Etude comportementale et électrophysiologique des processus impliqués dans l'effet McGurk et dans l'effet de ventriloquie.* Thèse de doctorat : Université Libre de Bruxelles.

Colin, C. & Radeau, M. (2003). Les illusions McGurk dans la parole : 25 ans de recherches. *L'année psychologique, 104,* 497-542.

Cunningham, J., Nicol, T., Zecker, S. & Kraus, N. (2000). Speech-evoked neurophysiologic responses in children with learning problems: Development and behavioral correlates of perception. *Ear & Hearing, 21,* 554-568.

Cutting, J.E., & Rosner, B.S. (1974). Categories and boundaries in speech and music. *Perception & Psychophysics, 16*, 564-570.

Damasio, A.R. (1989). Time-locked multiregional retroactivation: a system-level proposal for the neural substrates of recall and recognition. *Cognition, 33*, 25-62.

Damasio, A.R. (1990). The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation, 1*, 123-132.

Damper, R.I. & Harnad, S.R. (2000). Neural network models of categorical perception. *Perception & Psychophysics, 62*, 843-867.

DeCasper, A.J., Lecanuet, J.P., Busnel, M.C., Granier-Deferre, C. & Maugeais, R. (1994). Fetal reactions to recurrent maternal speech. *Infant Behavior and Development, 17*, 159-164.

De Cordemoy, G. (1668). *Discours physique de la parole*. Le Petit.

Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Mérieux, S., Roche, A., Sigman, M. & Dehaene, S. (2006). Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proceedings of the National Academy of Sciences of the Unites States of America, 103*, 14240-14245.

Delattre, P. (1958). Les indices acoustiques de la parole. *Phonetica, 2*, 108-118.

Delattre, P. (1968). From acoustic cues to distinctive features. *Phonetica, 18*, 198-230.

Deruelle, C. & De Schonen, S. (1995). Pattern processing in infancy: Hemispheric differences in the processing of shape and location of visual components. *Infant Behavior and Development, 18*, 123-132.

Donald, S.L. (1978). *The perception of voicing contrasts in Thai and English.* PhD Thesis : University of Connecticut.

Dunn, L.M., & Dunn, L.M. (1981). *Peabody Picture Vocabulary Test - Revised*. Circle Pine, MN, American guidance Service.

Dunn, L.M., Theriault-Whalen, C.M. & Dunn, L.M. (1993). *Echelle de vocabulaire en images Peabody*. Adaptation française du Peabody Picture Vocabulary test-revised. Toronto: Psycan.

Eggermont, J.J. (1995). Representation of a voice onset time continuum in primary auditory cortex of the cat. *Journal of the Acoustical Society of America, 98*, 911-920.

Eggermont, J.J. (2000). Sound-induced synchronization of neural activity between and within three auditory cortical areas. *Journal of Neurophysiology, 83*, 2708-2722.

Eggermont, J.J. (2001). Between sound and perception: Reviewing the search for a neural code. *Hearing Research, 157,* 1-42.

Eggermont, J.J. & Ponton, C.W. (2002). The neurophysiology of auditory perception: From single-units to evoked-potentials. *Audiology & Neuro-Otology, 7,* 71-99.

Eilers, R., Gavin, W. & Wilson, W. (1979). Linguistic exposure and phonetic perception in infancy: a cross-linguistic study. *Child Development, 50,* 14-18.

Eimas, P.D. & Corbit, J.D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology, 4,* 99-109.

Eimas, P.D., Siqueland, E.R., Jusczyk, P. & Vigorito, J. (1971). Speech perception in infants. *Science, 171,* 303-306.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotions, 6,* 169-200.

Elliott, L.L., Busse, L.A., Partridge, R., Rupert, J. & de Graaf, R. (1986). Adult and child discrimination of CV syllables differing in Voice Onset time. *Child Development, 57,* 628-635.

Elman, J., Bates, E., Johnson, M., Karmiloff-Simth, A., Parisi, A. & Plunkett, K. (1996). *Rethinking innateness : a connectionist perspective on development,* Cambridge, MA. : MIT Press.

Engel, A.K., Fries, P., König, P., Brecht, M. & Singer, W. (1999). Temporal binding, binocular rivalry, and consciousness. *Consciousness and Cognition, 8,* 128-151.

Engel, K. & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Sciences, 5,* 16-25.

Epstein, W. (1982). Percept-percept coupling. *Perception, 11,* 75-83.

Finney, D. J. (1971). *Probit Analysis,* Cambridge: Cambridge University Press.

Flege, J. (1988). The production and perception of speech sounds in a foreign language. In H. Winitz (ed.). *Human communication and its disorders: A review,* Norwood, NJ: Ablex, 224-401.

Flege, J. (1992). The intelligibility of English vowels spoken by British and Dutch talkers. In R. Kent (ed.*). Intelligibility and Speech Disorders: Theory, Measurement and Management.* Amsterdam: Benjamins, 157-232.

Flege, J. & Eefting, W. (1986). Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica, 43,* 155-171.

Fodor, J.A. (1983). *The modularity of Mind,* Cambridge, Massachussets : The MIT Press.

Fowler, C.A. (1986). An event approach to the study of speech perception. *Journal of Phonetics, 14,* 2-28.

Fowler, C.A. (1996). Listeners do ear sounds, not tongues. *Journal of the Acoustical Society of America, 99,* 1730-1741.

Fowler, C.A. & Rosenblum, L.D. (1990). Duplex perception: a comparison of monosyllables and slamming doors. *Journal of Experimental Psychology. Human perception and performance, 16,* 742-754.

Franklin, A. & Davies, I.R.L. (2004). New evidence for infant colour categories. *British Journal of Developmental Psychology, 22,* 349-377.

Fromkin, V.A. (1979). Persistent questions concerning distinctive features. In B. Lindblom and S. Ohman (eds.), *Frontiers of speech communication research*. London: Academic Press, 323-334.

Fujisaki, H. & Kawashima, T. (1969). *On the modes and mechanisms of speech perception. Annual report of the engineering research institute,* Faculty of Engineering, University of Tokyo, 28, 67-73.

Fujisaki, H. & Kawashima, T. (1970). *Some experiments on speech perception and a model for the perceptual mechanism,* Faculty of Engineering, University of Tokyo, 29, 207-214.

Galantucci, B., Fowler, C.A. & Turvey, M.T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review, 13,* 361-377.

Gibson, J.J. (1966). *The senses considered as perceptual systems,* Houghton Mifflin: Boston.

Godey, B., Schwartz, D., de Graaf, J.B., Chauvel, P. & Liégeois-Chauvel, C. (2001). Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: a comparison of data in the same patients. *Clinical Neurophysiology, 112,* 1850-1859.

Gopnik, A. & Meltzoff, A.N. (1987). The development of categorization in the second year and its relation to other cognitive and linguistic developments. *Child Development, 58,* 1523-1531.

Grieser, D. & Kuhl, P.K. (1989). Categorization of speech by infants: Support for speech prototypes. *Developmental Psychology, 25,* 577-588.

Guenther, F.H., Husain, F.T., Cohen, M.A. & Shinn-Cunningham, B.G. (1999). Effects of categorization and discrimination training on auditory perceptual space. *Journal of the Acoustical Society of America, 106,* 2900-2912.

Guenther, F.H., Nieto-Castanon, A., Ghosh, S.S. & Tourville, J.A. (2004). Representation of sound categories in auditory cortical maps. *Journal of Speech, Language and Hearing Research, 47*, 46-57.

Harnad, S. (1987). *Categorical perception: The groundwork of cognition*, New York: Cambridge University Press.

Hay, J.S.F. (2005). *How auditory discontinuities and linguistic experience affect the perception of speech and non-speech in English- and Spanish-listeners*. PhD Thesis: University of Texas.

Hazan, V. & Barrett, S. (2000). The development of phonemic categorization in children aged 6-12. *Journal of Phonetics, 28*, 377-396.

Heering, A. Houthys, S. & Rossion, B. (2007). Holistic face processing is mature at 4 years of age: Evidence from the composite face effect. *Journal of Experimental Child Psychology, 96*, 57-70.

Hillenbrand, J. (1984). Perception of sine-wave analogs of voice onset time stimuli. *Journal of the Acoustical Society of America* 75, 231-340.

Hirsh, I.J. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America, 31*, 759-767.

Holt, L.L., Lotto, A.J. & Diehl, R.L. (2004). Auditory discontinuities interact with categorization: implications for speech perception. *Journal of the Acoustical Society of America*, 116, 1763-1773.

Horev, N., Most, T. & Pratt, H. (2007). Categorical perception of speech (VOT) and analogous non speech (FOT) signals: Behavioral and electrophysiological correlates. *Ear & Hearing, 28*, 111-128.

Immelmann, K. (1969). Song development in the zebra finch and other estrildid finches. In R. Hinde (ed.). *Bird vocalizations,* London, UK : Cambridge University Press, 61-74.

Jacques, H. (2007). *L'expertise musicale influence-t-elle la perception catégorielle de la parole ? Etude comportementale et électrophysiologique*. Mémoire de fin d'étude en psychologie : Université Libre de Bruxelles.

Jakobson, R., Fant, C.M. & Halle, M. (1952*). Preliminaries to speech analysis. The distinctive features and their correlates*. Cambridge Massachussets : MIT Press.

Jakobson, R. (1973). Observations sur le classement phonologique des consonnes. In R. Jakobson (ed.). *Essais de Linguistique Générale*, Paris: Editions de Minuit, 123-130.

Joliot, M., Ribary, U. & Llinas, R. (1994). Human oscillatory brain activity near 40 Hz coexists with cognitive temporal binding. *Proceedings of the National Academy of Sciences of the Unites States of America, 91*, 11748-11751.

Jusczyk, P.W., Pisoni, D.B., Walley, A. & Murray, J. (1980). Discrimination of relative onset time of two-component tones by infants. *Journal of the Acoustical Society of America, 67*, 262-270.

Kaiser, J. Hertrich, I., Ackermann, H., Mathiak, K. & Lutzenberger, W. (2005). Hearing lips: gamma-band activity during audiovisual speech perception. *Cerebral Cortex, 15*, 646-653.

Karmiloff-Smith, A., (1991). Innate constraints and developmental change. In S. Carey and R. Gelman (eds.). *Epigenesis of the Mind: Essays in Biology and Knowledge*, New Jersey: Erlbaum, 171-197.

Kessinger, R. & Blumstein, S. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics, 25*, 143-168.

Kessinger, R. & Blumstein, S. (1998). Effects of speaking rate on voice-onset time and vowel production: some implications for perception studies. *Journal of Phonetics, 26*, 117-128.

Klatt, D.H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America, 67*, 971-995.

Knudsen, E.I. (2004). Sensitive periods in the development of the brain and behavior. *Journal of Cognitive Neuroscience, 16*, 1412-1425.

Kluender, K.R., Diehl, R.L. & Killeen, P.R. (1987). Japanese quail can learn phonetic categories. *Science, 237*, 1195-1197.

Kuhl, P.K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50*, 93-107.

Kuhl, P.K. (1993). Innate predispositions and the effect of experience in speech perception: The native language magnet theory. In B. de Boysson-Bardies, S. Schonen, P. Jusczyk, P. MacNeilage and J. Morton (eds). *Developmental neurocognition: Speech and face processing in the first year of life,* The Hague, Netherlands: Kluwer, 259-274.

Kuhl, P.K., Conboy, B.T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M. & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B, 363*, 979-1000.

Kuhl, P.K., Conboy, B.T., Padden, D., Nelson, T. & Pruitt, J. (2005). Early speech perception and later language development: implications for the “critical period”. *Language Learning and Development, 1,* 237-264.

Kuhl, P.K. & Miller, J.D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science, 190,* 69-72.

Kuhl, P.K. & Miller, J.D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America, 63,* 905-917.

Kuhl, P.K. & Padden, D.M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *Journal of the Acoustical Society of America* 73, 1003-1010.

Kuhl, P.K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S. & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9,* F13-F21.

Kuhl, P.K., Tsao, F.M. & Liu, H.M. (2003). Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences of the Unites States of America, 100,* 9096-9101.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N. & Limdblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255,* 606-608.

Lacey, J.I. & Lacey, B.C. (1958). The relationship of resting autonomic activity to motor impulsivity. In Wiliams and Wilkins (eds.), *The brain and human behaviour*. Baltimore, 144-209.

Laguitton, V., de Graaf, J.B., Chauvel, P. & Liégeois-Chauvel, C. (2000). Identification reaction times of voiced/voiceless continua : a right-ear advantage for VOT values near the phonetic boundary. *Brain and Language, 75,* 153-162.

Lalonde, C.E. & Werker, J.F. (1995). Cognitive influences on cross-language speech perception in infancy. *Infant Behavior and Development, 18,* 459-475.

Lane, H.L. (1965). Motor theory of speech perception: A critical review. *Psychological Review, 72,* 275-309.

Lasky, R.E., Syrdal-Lasky, A. & Klein, R.E. (1975). VOT discrimination by four to six and a half month old infants from Spanish environment. *Journal of Experimental Child Psychology, 20,* 215-225.

Laufer, A. (1998). Voicing in contemporary Hebrew in comparison with other languages. *Hebrew studies, 39*, 143-179.

Leavitt, L.A., Brown, J.W., Morse, P.A. & Graham, F.K. (1976). Cardiac orienting and auditory discrimination in 6-week-old infants. *Developmental Psychology, 12*, 514-523.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P. & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431-461.

Liberman, A.M., Harris, K.S., Hoffman, K.S. & Griffith, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*, 358-368.

Liberman, A.M. & Mattingly, I.G. (1985). The motor theory of speech perception revised. Cognition, 21, 1-36.

Liégeois-Chauvel, C., de Graaf, J.B., Laguitton, V. & Chauvel, P. (1999). Specialization of left auditory cortex for speech perception in Man depends on temporal coding. *Cerebral Cortex, 9*, 484-496.

Liégeois-Chauvel, C., Musolino, A. & Chauvel, P. (1991). Localization of the primary auditory area in man. *Brain, 114*, 139-151.

Lisker, L. (1978). Rapid vs Rabid: A catalogue of acoustic features that may cue the distinction. *Status report on speech research, 54*, 127-132.

Lisker, L. (1985). The pursuit of invariance in speech signals. *Journal of the Acoustical Society of America, 77*, 1189-1202.

Lisker, L. & Abramson, A.S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*, 384-422.

Lisker, L. & Abramson, A.S. (1967). Some effects of context on voice onset time in English stops. *Language and Speech, 10*, 1-28.

Lisker, L. & Abramson, A.S. (1970). The voicing dimension: some experiments in comparative phonetics. In *Processing of the 6th International Congress of Phonetic Science, Prague 1967*. Prague: Academia, 563-567.

Lisker, L., Liberman, A.M., Erickson, D.M. & Dechovitz, D. (1978). On pushing the voice onset time boundary about. *Language and Speech, 20*, 209-216.

Liu, H., Kuhl, P. & Tsao, F. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science, 6*, F1-F10.

Liu, H., Ng, M.L., Wan, M., Wang, S. & Zhang, Y. (2008). The effect of tonal changes on voice onset time in Mandarin esophageal speech. *Journal of voice, 22*, 210-218.

Lorenz, K. (1970). The enmity between generations and its probable ethological causes. *Studium Generale, 23*, 963-997.

Lück SJ. (2005). *An introduction to the Event-Related Potential Technique*. Cambridge MA: MIT Press.

Maassen, B., Groenen, P., Crul, T., Assman-Hulsmans, C. & Gabreëls, F. (2001). Identification and discrimination of voicing and place of articulation in developmental dyslexia. *Clinical Linguistics and Phonetics, 15*, 319-339.

Macmillan, N.A. (1987). Beyond the categorical/continuous distinction: A psychophysical approach to processing modes. In S. Harnad (ed.). *Categorical perception: The groundwork of cognition*, New York: Cambridge University Press, 53-88.

Macmillan, N.A., Kaplan, H.L. & Creelman, C.D. (1977). The psychophysics of categorical perception. *Psychological Review, 84*, 452-471.

Mann, V.A. & Liberman, A.M. (1983). Some differences between phonetic and auditory modes of perception. *Cognition, 14*, 211-235.

Martin, B.A. & Boothroyd, A. (2000). Cortical auditory evoked potentials in response to changes of spectrum and amplitude. *Journal of the Acoustical Society of America, 107*, 2155-2161.

Martin, B.A., Sigal, A., Kurtzberg, D. & Stapells, D.R. (1997). The effects of decreased audibility produced by high-pass noise masking on cortical event-related potentials to speech sounds /ba/ and /da/. *Journal of the Acoustical Society of America, 101*, 1585-99.

Massaro, D.W. (1987). *Speech perception by ear and by eye: A paradigm for psychological inquiry*, Hillsdale, NJ: Lawrence Erlbaum Associates.

Massaro, D.W. & Oden, G.C. (1980). Evaluation and integration of acoustic features in speech perception. *Journal of the Acoustical Society of America, 67*, 996-1013.

Mattys, S.L., Jusczyk, P.W., Luce, P.A. & Morgan, J.L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38*, 465-494.

Maye, J., Weiss, D.J. & Aslin, R.N. (2008). Statistical phonetic learning in infants: facilitation and feature generalization. *Developmental Science, 11*, 122-134.

Maye, J. Werker, J.F. & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*, B101-B111.

Mc Cullagh, P. & Nelder, J.A. (1983). *Generalized Linear Models*, Chapman and Hall: London.

Medina, V., Hoonhorst, I., Bogliotti, C., Sprenger-Charolles, L. & Serniclaes, W. (submitted). Development of voicing perception in French: Comparisons between adults, children and adolescents.

Medina, V. & Serniclaes, W. (2005). Late development of the categorical perception of speech sounds in pre-adolescent children. *ZAS Papers in Linguistics, 42*, 13-32.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J. & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29*, 143-178.

Meunier, D. (2007). *Une modélisation évolutionniste du liage temporal*. Thèse de doctorat : Univerité Lumière Lyon2.

Miller, J.L. & Eimas, P.D. (1983). Studies on the categorization of speech by infants. *Cognition, 13*, 135-165.

Miller, J.L. & Eimas, P.D. (1996). Internal structure of voicing categories in early infancy. *Perception & Psychophysics*, 58, 1157-1167.

Miller J.D., Wier, C.C., Pastore, R.E., Kelly, W.J. & Dooling, R.J. (1976). Discrimination and labeling of noise-buzz sequences with varying noise-lead times: an example of categorical perception. *Journal of the Acoustical Society of America, 60*, 410-417.

Mody, M., Studdert-Kennedy, M. & Brady, S. (1997). Speech perception deficits in poor readers: auditory processing or phonological reading? *Journal of Experimental Child Psychology, 64*, 199-231.

Moffit, A.R. (1971). Consonant cue perception by twenty-to-twenty-four-week-old infants. *Child Development, 41*, 1159-1171.

Mondloch, C.J., Le Grand, R. & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception, 31*, 553-566.

Morais, J. & Kolinsky, R. (1994). Perception and awareness in phonological processing: the case of the phoneme. *Cognition, 50*, 287-297.

Moushegian, G. & Rupert, A.L. (1970). Neuronal response correlates of cochlear nucleus: evidence for restrictive and multiple parameter information transfer. *Experimental Neurology, 29*, 349-365.

Näätänen, R. & Alho, K. (1997). Mismatch negativity: the measure for central sound representation accuracy. *Audiology and Neuro-Otology, 2*, 341-353.

Näätänen, R., Paavilainen, P., Rinne, T. & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical Neurophysiology, 118*, 2544-2590.

Näätänen, R., Gaillard, A.W. & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica, 42*, 313-329.

Näätänen, R. & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a Review and an analysis of the component structure. *Psychophysiology, 24*, 375-425.

Nazzi, T., Bertoncini, J. & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology. Human Perception and Performance, 24*, 756-766.

Nearey, T.M. (1990). The segment as a unit of speech perception. *Journal of Phonetics, 18*, 347-373.

Oldfield, R.C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97-113.

Parker, E. (1988). Auditory constraints on the perception of voice-onset time: The influence of lower tone frequency on judgments of tone-onset simultaneity. *Journal of the Acoustical Society of America, 83*, 1597-1607.

Pastore, R.E. (1987). Categorical perception: Some psychophysical models. In S. Harnad (ed.). *Categorical perception: The groundwork of cognition*, New York: Cambridge University Press, 29-52.

Pastore, R.E., Ahroon, X., Baffuto, K.J., Friedman, C., Puelo, J.S. & Fink, E.A. (1977) Common-factor model of categorical perception. *Journal of Experimental Psychology : Human Perception and Performance, 3*, 686-696.

Pastore, R.E., Schmuckler, M.A., Rosenblum, L. & Szczesiul, R. (1983). Duplex perception with musical stimuli. *Perception & Psychophysics, 33*, 469-474.

Picton, T.W., Bentin, S., Berg, P., Donchin, E., Hillyard, S.A., Johnson, R. JR, Miller, G.A., Ritter, W., Ruchkin, D.S., Rugg, M.D. & Taylor, M.J. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology, 37*, 127-152.

Piéron, H. (1964). *La sensation. Que sais-je?*: Presses Universitaires de France.

Pisoni, D.B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics, 13*, 253-260.

Pisoni, D.B. (1977). Identification and discrimination of the relative onset time of two-component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America, 61*, 1352-1361.

Plutchik, R. (1980). *Emotion : a psychoevolutionary synthesis*, Harper and Row, New-York.

Poeppel, D., Yellin, E., Phillips, C., Roberts, T.P.L., Rowley, H.A., Waxler, K. & Marantz, A. (1996). Task-induced asymmetry of the auditory evoked M100 neuromagnetic field elicited by speech sounds. *Cognitive Brain Research, 4*, 231-242.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication, 41*, 245-255.

Pollack, I. & Pisoni, D. (1971). On the comparison between identification and discrimination tests in speech perception. *Psychonomic Science, 24*, 299-300.

Pratt, H., Starr, A., Michalewski, H.J., Bleich, N. & Mittelman, N. (2007). The N1 complex to gaps in noise: effects of preceding noise duration and intensity. *Clinical Neurophysiology, 118*, 1078-1087.

Raskin, L.A., Maital, S. & Bornstein, M.H. (1983). Perceptual categorization of color: A life-span study. *Psychological Research, 45*, 135-145.

Repp, B.H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin, 92*, 81-110.

Rivera-Gaxiola, M., Silva-Peyrera, J. & Kuhl, P.K. (2005). Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Developmental Science, 8*, 162-172.

Rothenberg, M. (1981). Acoustic interaction between the glottal source and the vocal tract. In K.N. Stevens and M. Hirano (eds.) *Vocal Fold physiology*, University of Tokyo Press, 305-328.

Saerens, M., Serniclaes, W. & Beeckmans, R. (1989). Acoustic versus contextual factors in stop voicing perception in spontaneous speech. *Language and Speech, 32*, 291-314.

Saffran, J.R., Aslin, R.N. & Newport, E.L. (1996). Statistical learning by 8-month old infants. *Science, 274*, 1926-1928.

Scherg, M. & Von Cramon, D. (1985). Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model. *Electroencephalography and Clinical Neurophysiology, 62*, 32-44.

Schouten, M.E.H. (1980). The case against a speech mode of perception. *Acta Psychologica, 44*, 71-98.

Schouten, B., Gerrits, E. & van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication, 41*, 71–80.

Schroeder, C.E. & Lakatos, P. (2009). The Gamma oscillation: Master or slave? *Brain Topography, Feb. 10.*

Schwartz, J. & Tallal, P. (1980). Rate of acoustic change may underlie hemispheric specialization for speech perception. *Science, 207,* 1380-1381.

Schwarzer, G. (2000). Development of face processing : The effect of face inversion. *Child Development, 71,* 391-401.

Serniclaes W. (1987). *Etude expérimentale de la perception du trait de voisement des occlusives du français*. Université Libre de Bruxelles: Belgium http://www.vjf.cnrs.fr/umr8606/DocHtml/PAGEPERSO/WSerniclaes.htm

Serniclaes, W. (2000). La perception de la parole. In P. Escudier, G. Feng, P. Perrier et J-L. Schwartz (eds). *La parole, des modèles cognitifs aux machines communicantes*, Paris: Hermès, 159-190.

Serniclaes, W. (2005). On the invariance of speech percepts. *ZAS Papers in Linguistics, 40,* 177-194.

Serniclaes, W., D'Alimonte, G. & Alegria, J. (1984). Production and perception of French stops by moderately deaf subjects. *Speech Communication, 3*, 185-198.

Serniclaes, W. & Geng, C. (in press). Cross-linguistic trends in the perception of place of articulation in stop consonants: A comparison between Hungarian and French. In I. Chitoran, C. Coupé, E. Marsico and F. Pellegrino (eds), *Approaches to phonological complexity*. The Hague: Mouton.

Serniclaes, W., Ventura, P., Morais, J. & Kolinsky, R. (2005). Categorical perception of speech sounds in illiterate adults. *Cognition, 98,* B35-B44.

Serniclaes, W. & Wajskop, M. (1992). Phonetic versus acoustic account of feature interaction in speech perception. In J. Alegria, D. Holender, J. Junça de Morais and M. Radeau (eds.) *Analytic Approaches to Human Cognition, Proceedings of the Conference in Honour of Paul Bertelson, Brussels, June 1991,* Amsterdam: North-Holland, 77-91.

Sharma, A. & Dorman, M.F. (1999). Cortical auditory evoked-potential correlates of categorical perception of voice onset time. *Journal of the Acoustical Society of America, 106,* 1078-1083.

Sharma, A. & Dorman, M.F. (2000). Neurophysiologic correlates of cross-language phonetic perception. *Journal of the Acoustical Society of America, 107,* 2697-2703.

Sharma, A., Marsh, C. & Dorman, M.F. (2000). Relationship between N1 evoked-potential morphology and the perception of voicing. *Journal of the Acoustical Society of America, 108,* 3030-3035.

Shtyrov, Y., Kujala, T., Lyytinen, H., Ilmoniemi, R.J. & Näätänen R. (2000). Auditory cortex evoked magnetic fields and lateralization of speech processing. *NeuroReport, 11,* 2893-2896.

Simon, C. & Fourcin, A.J. (1978). Cross-language study of speech-pattern learning. *Journal of the Acoustical Society of America,* 63, 925-935.

Simos, P.G., Diehl, R.L., Breier, J.I., Molis, M.R., Zouridakis, G. & Papanicolaou, A.C. (1998). MEG correlates of categorical perception of a voice onset time continuum in humans. *Cognitive Brain Research,* 7, 215-219.

Simos, P.G., Breier, J.I, Zouridakis, G. & Papanicolaou, A.C. (1998a). Magnetic fields elicited by a tone onset time continuum in humans. *Cognitive Brain Research, 6,* 285-294.

Simos, P.G., Breier, J.I., Zouridakis, G. & Papanicolaou, A.C. (1998b). MEG correlates of categorical-like temporal cue perception in humans. *NeuroReport,* 9, 2475-2479.

Sinex, D.G. & McDonald, L.P. (1988). Average discharge rate representation of voice onset time in the chinchilla auditory nerve. *Journal of the Acoustical Society of America, 83,* 1817-1827.

Sinex, D.G. & McDonald, L.P. (1989). Synchronized discharge rate representation of voice onset time in the chinchilla auditory nerve. *Journal of the Acoustical Society of America, 85,* 1995-2004.

Sinnott, J.M. & Adams, F.S. (1987). Differences in human and monkey sensitivity to acoustic cues underlying voicing contrasts. *Journal of the Acoustical Society of America, 82,* 1539-1547.

Skinner, B.F. (1957). *Verbal behavior*. New-York, NY: Appleton-Century-Crofts.

Snowdon, C.T. (1987). A naturalistic view of categorical perception. In S. Harnad (ed.). *Categorical perception: The groundwork of cognition,* New York: Cambridge University Press, 332-354.

Steinschneider, M., Fishman, Y.I. & Arezzo, J.C. (2003). Representation of the voice onset time (VOT) speech parameter in population responses within primary auditory cortex of the awake monkey. *Journal of the Acoustical Society of America, 114,* 307-321.

Steinschneider, M., Schroeder, C.E., Arezzo, J.C. & Vaughan, H.G. (1994). Speech-evoked activity in primary auditory cortex: Effects of voice onset time. *Electroencephalography and Clinical Neurophysiology, 92*, 30-43.

Steinschneider, M., Schroeder, C.E., Arezzo, J.C. & Vaughan, H.G. (1995). Physiologic correlates of the voice onset time boundary in primary cortex (A1) of the awake monkey: Temporal response patterns. *Brain & Language, 48*, 326-340.

Steinschneider, M., Volkov, I.O., Fishman, Y.I., Oya, H., Arezzo, J.C. & Howard, M.A. (2005). Intracortical responses in human and monkey primary auditory cortex support a temporal processing mechanism for encoding of the Voice Onset Time phonetic parameter. *Cerebral Cortex, 15*, 170-186.

Steinschneider, M., Volkov, I.O., Noh, M.D., Garell, P.C. & Howard, M.A. (1999). Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *Journal of Neurophysiology, 82*, 2346-2357.

Stevens, K.N. (1989). On the quantal nature of speech. *Journal of Phonetics, 17*, 3-45.

Stevens, K.N. & Klatt, D.H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustical Society of America, 55*, 653-659.

Streeter,L.A. (1976). Language perception of two-month-old infants shows effects of both innate mechanisms and experience. *Nature, 259*, 39-41.

Studdert-Kennedy, M., Liberman, A.M., Harris, K.S. & Cooper, F.S. (1970). Theoretical notes. Motor theory of speech perception: a reply to Lane's critical review. *Psychological Review, 77*, 234-249.

Summerfield, Q. & Haggard, M. (1974). Perceptual processing of multiple cues and contexts: Effects of following vowel on stop consonant voicing. *Journal of Phonetics, 2*, 279-295.

Summerfield, Q. & Haggard, M. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America, 62*, 435-448.

Tanner, W.P. Jr & Swets, J.A. (1954). A decision-making theory of visual detection. *Psychological Review, 61*, 401-409.

Tonnquist-Uhlen, I., Ponton, C.W., Eggermont, J.J., Kwong, B. & Don, M. (2003). Maturation of human central auditory system activity: the T-complex. *Clinical Neurophysiology, 114*, 685-701.

Trébuchon-Da Fonseca, A., Giraud, K., Badier, J.M., Chauvel, P. & Liégeois-Chauvel C. (2005). Hemispheric lateralization of voice onset time (VOT) comparison between depth and scalp EEG recordings. *NeuroImage, 27*, 1-14.

Tresiman, A.M. & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97-136.

Treisman, M, Faulkner, A., Naish, P.L. & Rosner, B.S. (1995). Voice onset time and tone onset time: the role of criterion-setting mechanisms in categorical perception. *The Quarterly Journal of Experimental Psychology. A. Human Experimental Psychology, 48*, 334-366.

Tremblay, K.L. & Kraus, N. (2002). Auditory training induces asymmetrical changes in cortical neural activity. *Journal of Speech Language and Hearing Research, 45*, 564-572.

Tremblay, K.L., Piskosz, M. & Souza, P. (2003). Effects of age and age-related hearing loss on the neural representation of speech cues. *Clinical Neurophysiology, 114*, 1332-1343.

Vaughan, Jr H.G. (1975). The analysis of scalp-recorded potentials. In: R.F. Thompson and M.M. Patterson (eds). *Bioelectric recording techniques, part B.* New-York: Academic Press, 158-207.

Vihman, M.V. (1996). *Phonological development: The origins of language in the child*. Cambridge, MA: Blackwell.

Von der Malsburg, C. (1981). *The correlation theory of brain function*. Gottingen, Germany: Technical report, Max Planck Institute for Biophysical Chemistry.

Wajskop, M., & Sweerts, J. (1973) Voicing cues in oral stop consonants. *Journal of Phonetics, 1*, 121-130.

Ward, L.M. (2003). Synchronous neural oscillations and cognitive processes. *Trends in Cognitive Sciences, 7*, 553-559.

Werker, J.F. & Curtin, S. (2005). PRIMIR: a developmental framework of infant speech processing. *Language Learning and Development, 1*, 197-234.

Werker, J.F., Fennell, C.T., Corcoran, K.M. & Stager, C.L. (2002). Infants' ability to learn phonetically similar words: effects on age and vocabulary size. *Infancy, 3*, 1-30.

Werker, J.F. & Logan, J.S. (1985). Cross-language evidence for three factors in speech perception, *Perception & Psychophysics, 37*, 35-44.

Werker, J.F. & Tees, R.C. (1984a). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49-63.

Werker, J.F. & Tees, R.C. (1984b). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America, 75,* 1866-1878.

Werker, J.F. & Tees, R.C. (1999). Influences of infant speech processing: Towards a new synthesis. *Annual Review of Psychology, 50,* 509-535.

Whalen, D.H. & Liberman, A.M. (1987). Speech perception takes precedence over nonspeech perception. *Science, 237,* 169-171.

Whiteside, S.P. & Marshall, J. (2001). Developmental trends in voice onset time: some evidence for sex differences. *Phonetica, 58,* 196-210.

Williams, L. (1977). The voicing contrast in Spanish. *Journal of Phonetics, 5,* 169-184.

Wolpaw, J.R. & Penry, J.K. (1975). Hemispheric differences in the auditory evoked response. *Electroencephalography and Clinical Neurophysiology, 43,* 99-102.

Wood, C.C. (1976). Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *Journal of the Acoustical Society of America, 60,* 1381-1389.

Yeni-Komshian, G.H., Caramazza, A. & Preston, M.S. (1977). A study of voicing in Lebanese Arabic. *Journal of Phonetics, 5,* 35-48.

Zlatin, M.A. & Koeningsknecht, R.A. (1975). Development of the voicing contrast: Perception of stop consonants. *Journal of Speech and Hearing Research, 18,* 541-553.