École des Hautes Études en Sciences Sociales

Rôle de la Syllabe dans la Perception de la Parole : Études Attentionnelles

Thèse de Doctorat de Sciences Cognitives de l'EHESS nouveau régime soutenue le 12 décembre 1994.

par
Christophe Pallier

sous la direction de
Jacques Mehler

Jury :
M. Jeannerod
W. Marslen-Wilson
J. Segui

Le cerveau est un outil admirable : il se met en marche le jour de la naissance et ne cesse de fonctionner jusqu'au jour où l'on doit prendre la parole en public.

Remerciements

J'ai eu la chance d'être contredit par Juan Segui, Nuria Sebastian-Gallés, Jacques Mehler, Emmanuel Dupoux, Stanislas Dehaene, Anne Christophe, Luca Bonnati (dans l'ordre alphabétique inverse). Tous, je les remercie, car, à des titres divers, ils ont fortement contribué à façonner ma pensée.

J'ai des dettes intellectuelles envers de nombreuses autres personnes. Plus particulièrement, je voudrais signaler au lecteur que si l'occassion s'en présente à lui, il ne manque pas d'écouter Luigi Rizzi, Zenon Pylyshyn, Alan Prince, Michael Kenstowicz, Bela Julesz, François Dell et Paul Bertelson.

Je remercie Juan Segui, William Marlsen-Wilson et Marc Jeannerod d'avoir accepté d'être membres du Jury, ainsi que Nuria Sebastian-Gallés, José Morais et Uli Frauenfelder d'en être les prérapporteurs.

Enfin, ce travail n'aurait pas pu être possible sans le soutien financier de la D.R.E.T. (Contrat 91–8152) et de l'École polytechnique.

Comme il y a une vie avant la thèse, je remercie mes parents de la liberté et de la confiance qu'ils m'ont toujours accordées. En écrivant, je pensais aussi à Alexandre, Olivier, Serge... j'aimerais que vous lisiez les dix premières pages sans vous endormir. Finalement, je ne peux exprimer avec des mots tout ce que je dois à Géraldine à qui je dédie ce travail. Quant à Gabriel, eh bien il va enfin pouvoir jouer avec les poissons de l'économiseur d'écran, je laisse le PC libre...

Bien entendu, les fautes de frappe, de présentation, de grammaire et de logique sont imputables à un virus informatique.

Chapitre I

Introduction

La conception que les lettres révèlent les sons qui constituent la parole semble évidente pour celui qui emploie journellement l'alphabet. Bien sûr, tout le monde est conscient que l'orthographe ne respecte pas parfaitement la prononciation, particulièrement en France où la dictée est un passe-temps national. On admet aussi que les relations entre sons et lettres varient selon les langues, qui peuvent même utiliser différents alphabets, par exemple, le cyrillique ou le grec. Toutefois, avec suffisamment de soin, il doit être possible de représenter un énoncé d'une langue quelconque avec un alphabet adéquat qui, à chaque son, associerait une lettre.

Cette idée est le fondement de l'Alphabet Phonétique International. Les transcriptions phonétiques figurent dans les meilleurs dictionnaires, et il est courant de lire dans les introductions élémentaires à la linguistique que la parole « est une suite de sons, appelés des *phones* ». Pour celui qui, comme nous, s'intéresse à la façon dont le cerveau perçoit la parole, cette formulation conduit irrésistiblement à l'idée suivante : la perception auditive serait essentiellement similaire à une lecture séquentielle, où les formes visuelles des lettres sont remplacées par des « formes sonores » correspondant chacune à un phone. Par exemple, entendre le mot « thèse », ce serait percevoir successivement les sons /t/, /e/, et /z/.

Une première remarque vient tempérer ce modèle : la perception de la parole ne peut être le simple reflet d'objets acoustiques présents dans le signal puisque les catégories que perçoit un locuteur, dépendent de sa langue. Ainsi, dès le début du siècle, Sapir remarquait qu'il était « impossible d'apprendre à un Indien à établir des distinctions phonétiques qui ne correspondent à rien dans le système de sa langue, même si ces distinctions frappaient nettement notre oreille objective » (Sapir, 1921, p.56 de la traduction française). Par exemple, là où un Français percevra deux « sons » différents : /r/ ou /l/, un Japonais entendra deux exemplaires du même son.

Cette observation est à l'origine de la phonologie et de la notion de *phonème* : on propose généralement que les mots ne sont pas mémorisés comme des suites de sons, mais plutôt comme des suites d'unités abstraites, appelées des phonèmes (Halle, 1990). Quand un mot est prononcé, les phonèmes qui le constituent peuvent se réaliser comme différents phones en fonction du contexte. Par exemple, en français, le phonème final /b/ du mot « robe » se réalise comme le son [b], dans « une robe rouge », mais comme le son [p] dans « une robe fuchsia ». Les langues diffèrent par l'inventaire des phonèmes et par les règles qui régissent les relations entre les représentations phonémiques et les représentations phonétiques (voir, p.ex., Dell, 1973).[1]

On explique l'influence de la langue sur la perception en supposant que la conscience perceptive (phénoménologique) se fonde sur les phonèmes, plutôt que sur les phones. Cependant, ces derniers peuvent être décelés par l'oreille « objective » du phonéticien.[2] En conséquence, la réalité physique des phones, et leur disposition séquentielle dans le signal de parole, faisait peu de doute tant que les phonéticiens se fondaient sur leurs intuitions. Ainsi, Bloomfield annonçait en 1934 : « *On peut s'attendre à ce que la définition physique (acoustique) de chaque phonème de n'importe quel dialecte nous vienne du laboratoire dans les prochaines décennies* » (cité par Jakobson et Waugh, 1980, p. 23). Le modèle de traitement de l'information pour la reconnaissance des mots que suggèrent les conceptions linguistiques que nous venons d'évoquer, et qui était largement admis au milieu de ce siècle (cf, p.ex. Miller, 1951) est le suivant :

1° Il y a, dans le signal acoustique, des segments qui correspondent, un à un, avec les symboles de l'alphabet phonétique. Ces segments contiennent chacun de l'information acoustique caractéristique de la lettre phonétique

1. En fait, les théories phonologiques (structuralistes, fonctionnalistes, génératives...) diffèrent sur les méthodes d'identification des phonèmes, et sur la complexité des relations qu'elles autorisent entre la représentation phonémique et la représentation phonétique. En ce qui concerne la détermination des phonèmes, Chomsky (1955) a fait remarquer que les contrastes sémantiques habituellement employés ne sont pas satisfaisants parce qu'ils ne signalent pas *où* est la distinction phonémique, et également parce qu'il existe, d'une part des énoncés homonymes (p.ex. verre et vert), et d'autre part des énoncées synonymes phonétiquement distincts (p.ex. car et autobus). Il n'est pas question de faire ici une présentation des différentes conceptions linguistiques du phonème (voir Duchet, 1992 et la conclusion de Dell, 1973). Cependant, signalons qu'avec l'avènement des nouvelles représentations phonologiques multilinéaires (p.ex. cf Dell et Vergnaud, 1984 ; Goldsmith, 1990 pour une présentation), certains linguistes n'hésitent pas à affirmer qu'il n'y a plus de phonème en phonologie (p.ex. Kaye, 1990). Mais cela ne nous semble pas tout à fait vrai : au centre des descriptions, il demeure une « tire segmentale » qui est nécessaire pour synchroniser les traits.

2. Notons qu'un locuteur « naïf » peut parfois être capable de distinguer deux variantes du même phonème : un Français peut distinguer perceptivement un /r/ roulé et un /R/ uvulaire. Toutefois, discriminer des allophones (i.e. des variantes phonétiques du même phonème sous-jacent) est souvent plus difficile que des discriminer des phones correspondants à des phonèmes distincts (voir Best, McRoberts, & Nomathemba, 1988).

qui lui correspond.

2º notre appareil perceptif extrait une représentation phonétique du signal. Cette représentation est ensuite transformée en une représentation phonémique, c'est à dire une suite de phonèmes, spécifiques à la langue.[3]

3º la représentation phonémique perceptive est comparée aux mots qui sont également mémorisés comme des suites de phonèmes.

Disons le immédiatement : tout le monde (ou presque) pense désormais que la première affirmation est fausse. L'avènement des techniques d'enregistrement et de visualisation du signal de parole (Potter, Kopp, & Green, 1947), a révélé qu'il n'y avait pas de segments acoustiques correspondant un à un aux lettres de la représentation phonétique. Hockett (1955), Fant et Lindblom (1961), et Liberman, Cooper, Shankweiler, et Studdert-Kennedy (1967) ont popularisé l'idée que cela était dû à la *coarticulation* : la prononciation d'un phone serait influencée par ceux qui l'entourent ; à cause de l'inertie de nos articulateurs qui ne peuvent prendre que quatre ou cinq positions différentes par seconde (alors que le débit en phones atteint facilement 12 phones/sec), les phones seraient transmis en parallèle, ce qui explique qu'un même segment acoustique puisse contenir l'information de plusieurs phones (cf figure 1.1, page suivante).

Cette image d'une transmission en parallèle des phones est toutefois trompeuse. Elle laisse supposer que les informations relatives aux différents phones sont simplement superposées.[4] Pour Liberman et al. (1967), la réalisation acoustique d'un phone est tellement dépendante de ceux qui l'environnent qu'il n'existerait pas de caractéristique acoustique invariante associée à chaque phone. À l'appui de cette hypothèse, on montre grâce à des synthétiseurs de sons, que des objets acoustiques très variés peuvent évoquer perceptivement le même phone. Il pourrait du moins sembler possible de définir un phone comme une *classe* d'objets acoustiques. Mais une telle solution n'est pas satisfaisante, car un même segment acoustique peut être perçu différemment selon le contexte où il se trouve. Par exemple, Mann et Repp (1980) ont « collé » le même segment fricatif devant une voyelle /a/ et devant une voyelle /u/ ; dans le premier cas, les sujets entendaient /sa/, dans l'autre ils entendaient /ʃu/.[5] Selon Liberman et al. (1967), les phones ne seraient donc pas des objets *acoustiques*, mais plutôt, des objets *articulatoires*.

3. Le lien entre les deux représentations peut être simple (comme le supposent les structuralistes, pour qui c'est une « simple » surjection (« many-to-one mapping ») ou bien être réalisé par des règles transformationnelles (phonologie générative).

4. Ceci est, essentiellement, la vision qui a conduit au modèle TRACE de (Elman & McClelland, 1984, 1986).

5. Cf Repp (1982), Repp et Liberman (1987), et Pisoni et Luce (1986) pour des revues sur les effets de contextes et les relations de compensations entre indices acoustiques.

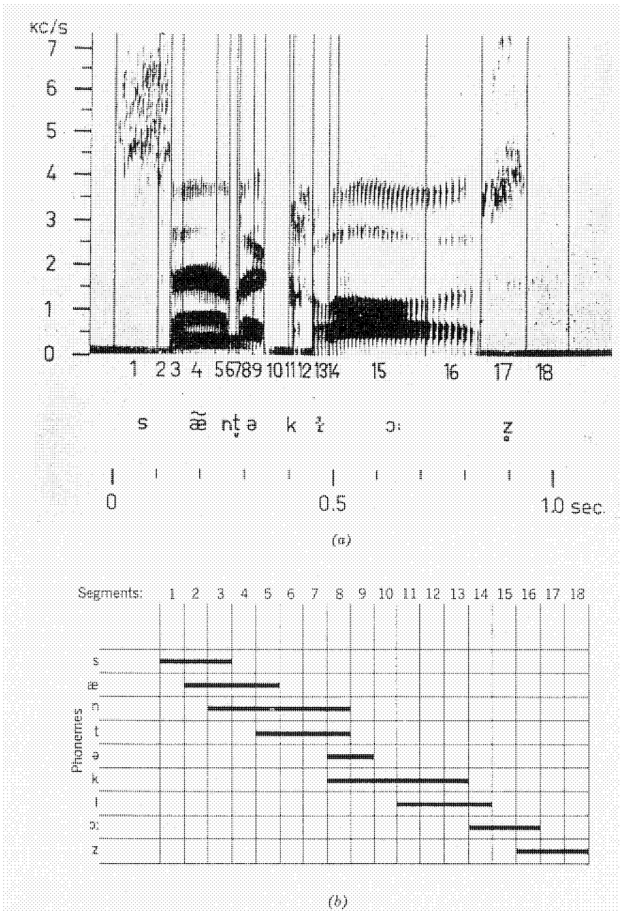


Fig.1.1: Schéma de Fant et Lindbloom (1961). (extrait de Lenneberg, 1967). En haut, figure un spectrogramme du mot « Santa-Klaus », segmenté selon les événements acoustiques remarquables. En bas, les auteurs ont fait figurer l'étendue de l'influence acoustique de chaque phone. On peut constater que le recouvrement entre phones est très important.

Avec cette théorie, connue sous le nom de Théorie Motrice, Liberman et ses collègues (Liberman, Cooper, Harris, & MacNeilage, 1963; Studdert-Kennedy, Liberman, Harris, & Cooper, 1970) font appel à la distinction classique entre stimulus distal et stimulus proximal : les phones correspondraient à des configurations des articulateurs, *transmises* par l'intermédiaire de l'onde acoustique. Liberman et al. proposent que le *décodage* de l'onde acoustique est effectué par un système perceptif spécialisé dans la reconnaissance des phones, et qui serait propre à l'espèce humaine. Ce système utiliserait des connaissances sur les contraintes de l'appareil articulatoire de l'humain pour effectuer des calculs permettant de déterminer quels sont les phones présents dans le signal.[6]

On peut affirmer qu'une large proportion des recherches menées dans le domaine de la perception de la parole durant les années 70-80, ont été conduites pour confirmer ou réfuter certaines hypothèses de la Théorie Motrice. Deux de ses prédictions ont été plus particulièrement étudiées : (a) la perception des sons de parole est-elle fondamentalement différente de celle des sons de « non-parole » (Pisoni, 1977) ? et (b) la perception phonétique est-elle propre aux humains ? Les recherches provoquées par ces questions ont permis, entre autres, de découvrir que les bébés catégorisaient la parole comme les adultes (Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Werker & Tees, 1984). Le cas des animaux est plus ambigu (Kuhl & Miller, 1978; Kuhl, 1991). Nous n'entrerons pas dans ces questions en détail (le lecteur pourra se référer à J. L. Miller, 1990, pour un début de revue).[7] La caractéristique de la Théorie Motrice, qu'il faut garder présente à l'esprit, est qu'elle conserve le plus intact possible la vision « classique », selon laquelle la perception des mots commence essentiellement par la récupération d'une représentation phonétique du signal. Cependant, contrairement au premier point du modèle « classique », cette théorie fournit l'image d'un signal acoustique de parole complexe, voire ambigu (puisqu'il nécessite des connaissances articulatoires pour être décodé). C'est ainsi qu'est né le « problème de la perception de la parole ». Parmi les propositions avancées pour le résoudre, nous allons en discuter trois :

1° la supposition qu'une utilisation intensive du contexte et des connaissances de haut-niveau permet de faciliter le traitement du signal.

2° l'affirmation que des invariants acoustiques (pour les phones ou les traits phonétiques) existent, bien qu'on ne les ait pas encore trouvés.

---

6. Il faut signaler qu'il existe un courant parmi les psychologues, pour lequel la perception des objets distaux n'implique pas des processus de transformation et de décodage de l'information, car celle-ci serait lisible « directement » dans le flux énergétique parvenant aux récepteurs (cf Gibson, 1966, 1979, et la critique de Fodor & Pylyshyn, 1981). C. A. Fowler est la représentante de ce courant la plus productive dans le domaine de la perception de la parole (Fowler, Rubin, Remez, & Turvey, 1980; Fowler, 1986, 1987; Fowler & Rosenblum, 1991). Les promoteurs de la perception directe ont cherché des arguments dans la théorie des systèmes dynamiques (Fowler et al., 1980; Browman & Goldstein, 1990) : pour eux les invariants seraient des paramètres de contrôle des dynamiques articulatoires ou acoustiques (voir également Petitot, 1985).

7. Notons que la théorie Motrice a évolué, car la recherche des invariants en production (qu'elle avait engendré) a révélé que l'invariance était autant problématique au niveau des l'articulation qu'au niveau de l'onde acoustique (p.ex. Abbs, 1986). Dans sa dernière mouture, l'objet de la perception sont les « intentions motrices » du locuteur (Liberman & Mattingly, 1985). Toutefois, pour beaucoup, cette caractérisation rend la théorie quasiment infalsifiable (voir les discussions dans Mattingly & Studdert-Kennedy, 1991, ainsi que la critique de Klatt, 1989).

3° l'abandon de l'idée que le système perceptif doive récupérer à tout prix une représentation phonétique et la proposition qu'il utilise une unité de reconnaissance plus large qui « capture » la variabilité du signal mieux que le phonème.

Le rôle des connaissances : L'idée que l'ambiguïté du signal est fondamentale, et qu'elle ne peut-être résolue que par le recours à des connaissances de haut-niveau, est commune dans le domaine de l'Intelligence Artificielle. Pour prendre un exemple (qui peut sembler extrême mais qui a été effectivement proposé) l'identification d'un savon serait grandement facilitée par le fait de savoir qu'on se trouve dans une salle de bain. Il est courant dans la littérature sur la reconnaissance automatique de la parole de proposer que des hypothèses lexicales, syntaxiques, sémantiques voire pragmatiques doivent être utilisées pour suppléer à l'ambiguïté du signal acoustique (p.ex. Young, 1990). Un exemple typique est le système HEARSAY (Reddy, Erman, Fennel, & Neely, 1973) dans lequel le message phonétique s'élabore progressivement sur un « tableau noir » où chaque source de connaissance peut proposer des hypothèses, le signal n'étant qu'une source d'information parmi les autres (lexicales, sémantique...). Ceci est le prototype même d'un système *interactif*, qui doit son nom aux processus d'interaction entre les connaissances.

La question de l'influence des sources de connaissances « supérieures » sur la perception de la parole est sans doute le problème qui occupé le plus d'espace dans les journaux de psychologie durant les années quatre-vingts. Ces recherches étaient stimulées par la thèse de la modularité (Fodor, 1983; Garfield, 1987) d'une part, et, d'autre part, par l'avènement du connexionnisme (McClelland & Rumelhart, 1986; Reilly & Sharkey, 1992), dont certains modèles faisaient une place centrale à l'interactivité.[8] Dans le domaine de la perception de la parole, les tenants de l'hypothèse interactive trouvent des justifications dans les recherches qui montrent l'existence d'influences dites de « haut en bas » (top-down) dans la perception de la parole par l'être humain. L'un des phénomènes les plus souvent cités est l'effet de restauration phonémique, découvert par Warren (1970) : quand on supprime un phonème (et qu'on le remplace par un bruit), les sujets affirment entendre le mot *intact*, avec le bruit *superposé*. Mieux encore, le contexte séman-

8. À l'heure actuelle, la plupart des modèles interactifs sont des modèles connexionnistes. Cependant ce serait une erreur de laisser croire que connexionnisme et interactivité sont indissolublement liés. On peut argumenter que HEARSAY, modèle symbolique par excellence, poussait l'interactivité beaucoup plus loin que le modèle connexionniste TRACE de Elman et McClelland (1986). Dans un modèle connexionniste commun, le perceptron, l'information va essentiellement de bas en haut, il n'y a pas de rétroaction des couches supérieures vers les couches inférieures (bien que, à cause de l'apprentissage, le perceptron puisse simuler des effets caractéristiques d'un système interactif (voir Norris, 1992).

tique peut influencer l'identité du phonème perçu : si '*' désigne le bruit, dans le contexte « it was found that the *eel was on the axle », le sujet entend « wheel » (roue), alors que si « axle » (essieu) est remplacé par « table », les sujets perçoivent « meal » (repas) (Warren & Warren, 1970).[9] Une autre démonstration de l'influence lexicale sur la perception des phonèmes est illustrée par l'expérience de Ganong (1980) : celui-ci a montré qu'un segment ambigu entre /t/ et /d/ était perçu plus souvent comme /t/ que comme /d/ dans le contexte /.ask/ (formant le mot /task/ par préférence au « non-mot » /dask/), mais qu'à l'opposé, dans le contexte /.ash/, il était plus souvent perçu comme comme /d/ que comme /t/ (/dash/ est un mot, /tash/ non).[10]

Ces faits sont, il est vrai, assez saisissants. Mais, après examen, ils ne démontrent pas que les premières étapes de traitement du signal sont influencées par des connaissances de haut niveau. En effet, notre perception *consciente* ne nous donne certainement pas un accès direct à la sortie des « transducers » (cf Pylyshyn 1984, p.174). Ce que nous percevons consciemment est le résultat d'une construction cognitive, influencée en partie non négligeable par nos connaissances et nos croyances ; l'hypothèse essentielle de la thèse de la modularité est qu'il y a une priorité aux informations sensorielles *pour ce qui concerne les premières étapes de traitement*, et que celles-ci sont peu affectées par nos connaissances ou croyances. On peut rendre compte des résultats de Warren et de Ganong (cf supra) avec un modèle de traitement non interactif en supposant que le sujet fonde sa réponse, non seulement sur une représentation phonémique extraite du signal, mais aussi sur une représentation post-lexicale après que le mot porteur ait été reconnu (Foss & Blank, 1980; Cutler & Norris, 1979; Cutler, Mehler, Norris, & Segui, 1987). Finalement, la distinction entre interactif et modulaire porte sur la question de savoir si un niveau supérieur (p.ex. lexical) peut influencer le traitement à un niveau inférieur (p.ex phonétique), question qu'on peut poser expérimentalement en essayant de déterminer précisément *à quel moment* les connaissances peuvent influencer la perception. En fait, les effets de contexte sont le plus évidents avec des stimuli ambigus ou dégradés et de plus, ils n'apparaissent qu'à des temps de réaction relativement lents (cf, p.ex. Fox, 1984; Pitt & Samuel, 1993) ; cela suggère que les premières étapes de traitement ne sont pas affectées par les connaissances, et que les effets de contexte peuvent être parfois purement décisionnels.[11]

9. Samuel (1981a, 1981b, 1987) a exploré cet « effet de restauration phonémique » avec la méthodologie de la détection du signal, dans l'intention de séparer les effets de biais décisionnels des effets perceptifs. Pour une revue de ces travaux, voir Samuel (1990), ainsi que la critique de Tyler dans le même volume.

10. Voir Fox (1984), Connine (1987), Connine et Clifton (1987), McQueen (1991), Pitt et Samuel (1993) pour des expériences qui continuent et précisent celles de Ganong (1980).

11. Certains refusent la distinction entre décision/perception, mais alors, si c'était le cas, les mots croisés relèveraient de la perception des mots écrits. Notons que le débat sur les influences

Il devient dès lors important de savoir si la parole naturelle est ambiguë ou non. En fait, on ne sait pas vraiment si c'est le cas : la parole enregistrée en studio est certainement peu ambiguë puisque V. Zue, après un entraînement (de 2500 heures), est parvenu à identifier des non-mots dans des spectrogrammes de parole continue (cf Cole, Rudnick, Zue, & Reddy, 1980). Pour ce qui est de la perception dans des conditions plus « écologiques » (dans un environnement bruyant, p.ex. au milieu d'autres conversations...), on ne sait pas exactement jusqu'à quel point la parole est ambiguë ou non. Bien entendu, si l'on extrait une étendue de deux cent millisecondes de signal de parole, il est probable que le sujet aura les plus grandes difficultés à l'identifier, mais cela est probablement dû au fait qu'on n'a pas donné suffisamment de contexte aux processus de séparations des sources sonores, d'adaptation à la réverbération du lieu...etc (Bregman, 1990). Quoiqu'il en soit, il demeure que nous n'hallucinons pas la majeure partie du temps et que, quelles que soient les influences de haut niveau, celles-ci ne peuvent s'exercer que sur une représentation calculée à partir de l'information présente dans le signal.

Les invariants : Les travaux de Liberman et al. ont largement répandu l'idée qu'il n'y a pas de caractéristique invariante définissant chaque phone, ni, a fortiori, chaque trait distinctif. Cependant, cette conviction se fonde en grande partie sur les expériences de synthèse de la parole. Or, certains chercheurs ont argumenté que les stimuli naturels sont beaucoup plus riches que les stimuli artificiels, et qu'ils pourraient contenir des invariants qui échappent à la simple inspection d'un spectrogramme. En conséquence, la recherche de modes de représentations du signal autres que le simple spectrogramme et qui soient plus à même de révéler les invariants, demeure active.

En particulier, beaucoup d'efforts sont consacrés à élucider les transformations non-linaires effectuées par le système auditif périphérique, dans l'espoir que les « neurospectogrammes » fourniront une représentation diminuant les caractéristiques variables et amplifiant les caractéristiques invariantes permettant de discriminer les sons de parole (Delgutte, 1981; Carlson & Granström, 1982). Pour la plupart des chercheurs, il fait peu de doute que les indices doivent être relativement abstraits, c'est à dire plus abstraits que « telle énergie dans telle bande de fréquence ». Ont été proposés : la forme globale du spectre à court-terme (Stevens & Blumstein, 1981), ou bien l'évolution dynamique de celui-ci (Kewley-Port & Luce, 1984), ainsi que des propositions plus exotiques que nous ne détaillerons pas ici.

---

de bas en haut (interactionisme) dans le traitement de la parole est loin d'être refermé. Le lecteur trouvera des points de références importants dans les articles de (par exemple) : Elman et McClelland (1988), Frauenfelder, Segui, et Dijkstra (1990), Massaro et Cohen (1991), McCelland (1991), McQueen (1991), Norris (1992), Samuel (1990).

L'unité de perception : Le principal problème pour des détecteurs de phones (ou de traits phonétiques) indépendants est la variabilité des réalisations acoustiques de ceux-ci (Repp & Liberman, 1987; Remez, 1987). Cette variabilité n'est cependant pas aléatoire car la réalisation d'un phone dépend en grande partie de ceux qui l'environnent. Par exemple, deux réalisations de /g/ devant /a/ sont acoustiquement beaucoup plus similaires entre-elles, que deux réalisations de /g/ devant des voyelles différentes. Une idée naturelle est alors de proposer une unité de reconnaissance de la parole prenant en compte une étendue acoustique plus large que le phones. Diverses unités de reconnaissance ont été proposées : le diphone (Klatt, 1980), la demi-syllabe (Fujimira, 1976; Samuel, 1989) et la syllabe (Massaro, 1974, 1987; Liberman & Streeter, 1978; Mehler, 1981; Segui, 1984). Parmi ces trois unités, la syllabe est celle qui a reçu le plus d'attention, et cela sans doute parce que c'est une unité linguistique et psychologique plus « naturelle » que le diphone ou la demi-syllabe. Comme le travail expérimental de notre thèse consistera à évaluer différentes prédictions de l'hypothèse « syllabique », selon laquelle l'unité de perception primaire est la syllabe, nous allons maintenant détailler les arguments en faveur de celle-ci.

En premier lieu, la syllabe, plutôt que le phonème, est la vraie unité de découpage séquentiel de la parole : alors qu'il existe de nombreuses contraintes sur les suites possibles de phonèmes, quasiment n'importe quelle suite de syllabes fournit un énoncé prononçable. Ceci est d'ailleurs l'une des raisons qui motivent l'introduction de la syllabe dans les théories linguistiques : elle permet de rendre compte d'une grande partie des contraintes phonotactiques (Goldsmith, 1990). Les suites possibles de phonèmes sont déjà nettement réduites quand on stipule que toute suite doit pouvoir être découpée en syllabes et qu'on a décrit les types de syllabes possibles dans la langue. Pour certaines langues, où seules les syllabes V, CV ou CVC sont possibles, cela réduit considérablement les possibilités. Le deuxième type de phénomène linguistique qui rend nécessaire le recours à la syllabe concerne les phénomènes accentuels : pour la plupart des langues, les règles qui régissent l'accentuation ne peuvent être formulées « simplement » qu'en faisant référence aux syllabes (Halle & Vergnaud, 1988). Par exemple, l'accent peut être sensible à la taille de la syllabe : il est « attiré » par les syllabes lourdes, i.e. celles qui contiennent plusieurs phonèmes dans la rime.[12] Bien entendu, les arguments linguistiques ne favorisent pas la syllabe par rapport au phonème : les deux sont des constructions extrêmement utiles pour rendre compte de phénomènes linguistiques mais cela ne préjuge pas de leur rôles respectifs dans la perception.

12. Les syllabes sont habituellement décrites comme étant composées d'une *attaque* (les consonnes qui la débute), et d'une *rime* (la voyelle et les consonnes qui la suivent). Dans « crac », l'attaque est « cr » et la rime « ac ». La rime se décompose elle-même en un noyau (la voyelle) et un coda (la ou les consonnes qui terminent éventuellement la syllabe).

Un argument plus psychologique souvent cité en faveur de la syllabe est que l'appréhension de la syllabe est plus naturelle, pour les enfants et les illettrés, que celle du phonème. Il est leur est beaucoup plus facile de manipuler des syllabes que de manipuler des phonèmes (Bertelson, de Gelder, Tfouni, & Morais, 1989; Liberman, Shankenweiler, Fisher, & Carter, 1974; Morais, Cary, Alegria, & Bertelson, 1979; Morais, Bertelson, Cary, & Alegria, 1986; Treiman & Breaux, 1982). Dans le même ordre d'idées, Bijeljac-Babic, Bertoncini, et Mehler (1993) ont montré que les nouveaux-nés étaient capables de distinguer des mots sur la base de leur nombre de syllabes (2 versus 3), mais pas sur la base de leur nombre de phonèmes. On peut aussi remarquer que dans l'histoire des systèmes d'écriture, les syllabaires ont précédé l'alphabet « phonémique » (cf Coulmas, 1989 pour une présentation et une généalogie des différents systèmes d'écriture).[13] Cependant le lien entre les intuitions métaphonologiques et les représentations calculées par le système perceptif n'est pas évident (Bertelson & de Gelder, 1991; Morais & Kolinsky, 1994). Par exemple, il est logiquement possible que la syllabe (et/ou le phonème) soit une unité de traitement utilisée pour produire la parole, mais pas pour la percevoir. Les tâches métalinguistiques nous renseigneraient alors sur les unités utilisées en production plutôt que sur celles utilisés en perception. Tout comme les arguments linguistiques, les arguments sur la « conscience métaphonologique » sont licites pour justifier de la réalité « psychologique » de telle ou telle unité, mais ils doivent être relativisés quand il s'agit de juger de leur réalité « perceptive ».[14]

Le premier type d'argument en faveur de la syllabe, plutôt que du phonème, comme unité primaire de perception est le fait que celle-ci est moins ambiguë que celui-là. En fait la variabilité acoustique des phonèmes est due dans une large mesure à l'interaction entre les consonnes et les voyelles : si bien que dans des syllabes CVC, l'information la plus fiable pour l'identification est fournie par les transitions entre la voyelle et les consonnes (Strange, 1987). L'interdépendance perceptive entre consonne et voyelle est illustrée dans l'expérience (op. cit.) de Mann et Repp (1980) où une fricative ambiguë est perçue comme /s/ devant /a/ et /ʃ/ devant /u/. Un autre exemple est celui de Miller et Liberman (1979), qui trouvent que la pente des formants de transitions qui détermine la perception de /b/ ou de /w/ est influencée par la durée de la voyelle qui suit. Autrement dit, une même transition formantique sera identifiée comme un /b/ quand la voyelle

13. De là à affirmer que le phonème est un simple artefact de l'écriture alphabétique (et que l'alphabet aurait été inventé plutôt que découvert), il y a un pas que ne franchissent pas les spécialistes de l'acquisition de l'écrit (cf Bertelson, 1986). Le point de vue généralement accepté est que l'apprentissage de l'alphabet *facilite* l'appréhension de la structure phonémique du langage.

14. Par exemple, dans leur défense de l'existence d'une représentation segmentale (i.e. phonétique) dans la perception de la parole, Pisoni et Luce (1987) ne citent en fait que des arguments pour la réalité psychologique des phonèmes, plutôt que pour leur réalité perceptive.

qui suit est longue, ou comme un /w/ quand elle est courte (Cela permettant, selon Miller et Liberman (1979), une adaptation au débit de parole). Finalement, une étude de Whalen (1989) montre encore plus clairement l'interdépendance des identifications d'une consonne et de la voyelle qui la suit : la quantité d'une grandeur acoustique utilisée pour identifier la première est « soustraite » de la grandeur utilisée pour identifier la seconde.

Ces faits laissent envisager qu'un système perceptif qui comparerait le signal avec des prototypes syllabiques (ou demi-syllabiques) aurait moins de difficultés qu'un système utilisant des prototypes de phonèmes. Certes, le nombre d'unités augmenterait largement (j'ai calculé qu'il serait de l'ordre de 3000 pour le français), mais la parcimonie n'est plus un argument très prisé de nos jours. En fait l'existence d'un syllabaire pour la production de la parole a été proposé par Levelt et Wheeldon (1994) qui s'appuient sur des effets de fréquence syllabique sur le temps de prononciation. Il n'y a donc rien de déraisonnable a priori à supposer l'existence d'un syllabaire pour la perception.[15]

Les effets expérimentaux les plus directs en faveur de la syllabe sont les effets de « congruence syllabique » (Mehler, Dommergues, Frauenfelder, & Segui, 1981), et de « complexité » (Segui, Dupoux, & Mehler, 1990). L'effet de congruence est observé dans une tâche où les sujets doivent détecter le plus rapidement possible si un stimulus auditif contient un fragment tel que, par exemple, /pa/ ou /pal/. Il s'avère que les temps de détection sont plus rapides pour un fragment qui correspond à la première syllabe du stimulus (p.ex. /pal/ dans /pal–mier/ et /pa/ dans /pa–lace/), que pour un fragment qui ne correspond pas à cette syllabe (p.ex. /pa/ dans /pal–mier/ et /pal/ dans /pa–lace/) (Mehler et al., 1981). L'interprétation proposée est la suivante : quand la cible ne correspond pas précisément à la structure syllabique du stimulus, les sujets sont ralentis car ils doivent, en quelque sorte, « briser » l'unité de perception qu'est la syllabe.

La même interprétation sous-tend l'effet de « complexité syllabique » qui apparaît lorsque des sujets doivent détecter un phonème placé au début d'un stimulus auditif : les temps de réaction dépendent alors du nombre de phonèmes contenus dans la première syllabe de celui-ci (Segui et al., 1990; Dupoux, en préparation). Les sujets détectent plus rapidement un phonème commençant une syllabe CV qu'un phonème commençant une syllabe CCV (ou CVC). Cela suggère que les phonèmes sont reconnus seulement après l'identification de la syllabe.

15. Toutefois, il faut souligner que l'invocation d'un syllabaire n'est pas la seule solution envisageable pour modéliser les effets de coarticulation. Par exemple, dans TRACE I, les effets de coarticulation sont « câblés » en autorisant des phonèmes à moduler les liens entre les traits acoustiques et d'autres phonèmes adjacents (Elman & McClelland, 1986). Toutefois, dans la discussion suivant ce chapitre, O. Fujimira fait remarquer que l'utilisation de demi-syllabes à la place des phonèmes serait tout de même plus économique que ce système de modulation et qu'il éviterait la compétition entre consonnes et voyelles.

Toutefois, une découverte de taille remet en question cette conclusion : ces effets de congruence et de complexité syllabiques, qui ont été obtenus avec des sujets français, n'ont pas été reproduits avec des sujets anglais (Cutler, Mehler, Norris, & Segui, 1986; Cutler, Butterfield, & Williams, 1987). Dans ces tâches, les Anglais ne sont apparemment pas sensibles à la structure syllabique des stimuli. De plus, d'autres auteurs (Marlsen-Wilson, 1984; Norris & Cutler, 1988; Pitt & Samuel, 1990) ont critiqué la notion que la syllabe est l'unité de décodage de la parole, sur la base d'autres expériences, menées également en anglais.[16] Ces résultats soulèvent la possibilité que les locuteurs de différentes langues utilisent différents types d'unités perceptives, hypothèse qui a engendré un ensemble de recherches comparant diverses langues (espagnol, portugais, hollandais, japonais... cf Cutler, 1993 pour une présentation de ces recherches).

Il faut souligner, cependant, qu'il est loin d'y avoir un consensus sur l'interprétation de ces différents résultats (on peut comparer, par exemple : Bradley, Sánchez-Casas, & García-Albea, 1993; Cutler et al., 1986; Dupoux, 1993; Norris & Cutler, 1985; Segui et al., 1990). La situation est particulièrement paradoxale en anglais où certains auteurs affirment que le phonème est l'unité de perception (p.ex. Pitt & Samuel, 1990), et d'autres, qu'il ne l'est pas (p.ex. Wood & Day, 1975). Au cours de cette thèse, nous soutiendrons que le débat sur l'unité de perception a été obscurci par la négligence des processus de décision mis en jeu dans les différentes tâches. Nous tâcherons d'éclaicir la situation dans les expériences qui suivent, et qui visaient toutes à tester la proposition que la syllabe est l'unité primaire de perception (plutôt que le phonème ou le trait distinctif), et ceci en utilisant des paradigmes expérimentaux innovateurs. Le premier d'entre eux est le paradigme d'interférence de Garner.

16. Ces expériences seront détaillées dans les chapitres expérimentaux.

Chapitre II

Interférences Syllabiques

Si les phonèmes sont perçus comme des unités indépendantes, alors on peut s'attendre à ce que les sujets puissent focaliser leur attention sur un phonème précis. Cette prédiction a été examinée par Wood et Day (1975) avec le paradigme d'interférence de Garner[1] : les sujets entendaient des listes de syllabes CV dont ils devaient déterminer le plus rapidement possible la consonne initiale (qui pouvait être soit `b`, soit `d`). Dans certaines listes dites « simples », la voyelle V était unique : à chaque essai, le sujet entendait soit /ba/, soit /da/. Dans d'autres listes, la voyelle pouvait varier d'un essai à l'autre : le sujet entendait /ba/, /bæ/, /da/, ou /dæ/ (listes dites « variées »). Si la consonne est perçue indépendamment de la voyelle, le sujet devrait pouvoir focaliser son attention sur elle et la variation de la voyelle ne devrait pas affecter son temps de décision. Au contraire, si la syllabe est identifiée avant le phonème alors on peut s'attendre à observer un ralentissement dans les listes « variées » où il faut reconnaître une syllabe parmi quatre par rapport aux listes « simples » où il faut reconnaître une syllabe parmi deux.

Wood et Day (1975) ont trouvé que les temps de réaction pour classifier la consonne initiale étaient systématiquement ralentis quand la voyelle variait, même lorsqu'on demandait explicitement aux sujets d'ignorer celle-ci.[2] Ils en ont conclu que la perception ne se faisait pas phonème par phonème mais que les sujets reconnaissaient d'abord la syllabe, puis ensuite les phonèmes.[3]

On pourrait penser que c'est le choix de plosives (/b/ et /d/) à classifier, qui est responsable de ce résultat. En effet, classiquement, les plosives sont considérées comme les phonèmes les plus « encodés », c'est à dire ceux dont la réalisation acoustique varie le plus avec la voyelle qui les suit (Liberman, Mattingly, & Tur-

1. Voir Garner (1974) et plus récemment le livre Lockhead et Pomerantz (1991) : *The Perception of Structure*, pour une présentation de nombreuses recherches engendrées par ce paradigme.

2. Wood et Day ont également montré que la classification de voyelle est affecté par la variation de la consonne initiale

3. Voir également Day et Wood (1972), Eimas, Tartter, et Miller (1981), Eimas, Tartter, Miller, et Keuthen (1978), Wood (1974, 1975).

vey, 1972). Si le sujet se focalisait essentiellement sur les aspects acoustiques du début des stimuli pour répondre, dans les listes variées, il y aurait quatre « images acoustiques » à comparer aux stimuli, et seulement deux dans les listes simples. L'effet de variabilité ne serait alors pas surprenant et n'impliquerait pas un traitement global de la syllabe. Toutefois, une remarquable étude de Tomiak, Mullenix, et Sawusch (1987) montre clairement que l'explication du phénomène ne réside pas dans la variabilité acoustique de la consonne : ils ont construit des syllabes synthétiques (/fæ/, /ʃæ/, /fu/, /ʃu/) en « collant » les fricatives aux voyelles de façon à ce qu'il n'y ait aucune coarticulation entre les deux segments acoustiques. En fait, le premier segment acoustique de 150 msec identifiait de manière non ambiguë la fricative et ne contenait aucune information sur la voyelle qui suivait. Pourtant la variation de la voyelle affectait encore largement les temps de classification de la consonne.

Ce résultat suggère que la perception « unitaire » de la syllabe ne reflète pas purement la structure physique du stimulus mais est bien le résultat d'une intégration mentale (l'effet est « dans la tête » des sujets). La suite de l'étude de Tomiak et al. (1987) le démontre encore plus clairement : les syllabes synthétiques étaient telles qu'il était possible de les entendre soit comme des syllabes, soit comme des «bruits». À la moitié des sujets elles étaient effectivement décrites comme des syllabes, alors qu'à l'autre moitié elles étaient décrites comme du « bruit ». C'est seulement quand elles étaient considérées comme des sons langagiers que la variation produisait un ralentissement ! Les mêmes stimuli physiques sont traités différemment selon qu'ils sont perçus comme de la parole ou comme du « bruit ». À notre avis, une interprétation plausible de ces résultats est que, dans le mode « linguistique », le modèle mental de la cible que se forment les sujets et qu'ils comparent à chaque essai au stimulus à classifier, doit nécessairement spécifier la voyelle ; il s'agirait donc d'une syllabe (ou au moins d'une demi-syllabe).

Deux autres études, utilisant la tâche de détection de phonème initial, nous semblent pouvoir être interprétées de la même façon : ce sont celles de Mills (1980) et Swinney et Prather (1980). Chez Mills (1980), les sujets entendaient des listes de sept syllabes CV dans lesquelles ils devaient détecter un phonème en position initiale. Le phonème cible (`/b/` ou `/s/`), était précisé en tête de chaque liste, de la façon suivante : « *vous devez détecter* `/b/` *comme dans* `/be/` », tandis que, dans la liste elle-même, il pouvait apparaître soit dans `/bo/`, soit dans `/be/`. Mills (1980) observe que les sujets sont plus rapides quand la syllable portant la cible correspond à celle de la consigne, et ceci, même quand on leur demande explicitement d'ignorer la voyelle. Cet effet suggère que le *modèle* de la cible que le sujet se forme pour effectuer la tâche est syllabique.

Swinney et Prather (1980) obtiennent un résultat comparable avec une méthode un peu différente. Ils utilisaient une tâche de détection mixte (phonème ou syllabe) : chaque bloc expérimental consistait en une série de petites listes de 5 à

12 monosyllabes CVC qui pouvaient contenir la cible (un phonème ou une syllabe). Il y avait trois blocs qui se différenciaient par la variabilité du contexte vocalique où pouvait apparaître la cible : dans un bloc, celle-ci était systématiquement suivie de la même voyelle ; dans un autre la cible était suivie par une voyelle parmi quatre ; dans le troisième bloc il y avait huit contextes vocaliques. Swinney et Prather (1980) ont observé que la variabilité ralentissait les sujets.[4]

En résumé, les études que nous venont de décrire montrent indubitablement que les sujets ne peuvent ignorer la voyelle qui suit une consonne sur laquelle ils doivent prendre une décision, et que ceci n'est probablement pas un effet acoustique de bas niveau. Une interprétation possible de ces résultats est que, conformément à un modèle syllabique (ou demi-bisyllabique) de la perception de la parole, les sujets identifient la syllabe CV en entier avant d'avoir accès à la consonne. Dans la condition non-variée, l'identification d'une syllabe parmi deux est suffisante pour décider de la réponse : il n'est pas nécessaire d'inspecter un code phonémique. On peut proposer deux causes du ralentissement dans la condition variée: (a) il faut identifier une syllabe parmi quatre, ou (b) il faut segmenter la syllabe en consonne-voyelle.

Cette interprétation n'est toutefois pas la seule qui puisse rendre compte de ces résultats. On peut proposer que les phonèmes sont extraits indépendamment les uns des autres mais que la variation d'un phonème adjacent ralentit *la décision*. Par exemple, Eriksen et Eriksen (1974) ont montré que les sujets qui devaient identifier une lettre (présentée visuellement), étaient incapables d'ignorer des lettres voisines, a priori non pertinentes pour effectuer la tâche. Plus précisement, une lettre cible était présentée au-dessus d'un point de fixation, et le sujet devait décider de son identité (en appuyant sur un bouton pour H ou K, et sur un autre pour S ou C). À gauche et à droite de la cible, apparaissaient d'autres lettres plus ou moins semblables à celle-ci, que le sujet avait pour instruction d'ignorer. Pourtant, ces lettres ralentissaient la réponse quand elles étaient dissimilaires à la cible ou prédisaient une réponse incompatible. Cela montre que les sujets ne pouvaient focaliser entièrement leur attention sur la lettre cible : les lettres adjacentes étaient identifiées *automatiquement* et interféraient dans les processus de décision sans que les sujets puissent l'empêcher (un peu comme un effet « Stroop »).

On peut donc imaginer qu'il en est de même dans la modalité auditive : lorsque le sujet entend une syllabe CV, les phonèmes sont identifiés *automatiquement*, et

4. En fait le ralentissement n'était observé que quand la cible était /b/ mais pas quand la cible était /s/ ; pour les syllabes, les temps sont constants. Signalons également que cette expérience présente un défaut méthodologique : dans le contexte à une voyelle, 1 seule voyelle était utilisée ; dans les contextes à 4 ou 8 voyelles, il y en avait d'autres. Swinney et Prather (1980) comparent les temps de réaction moyen sur toutes les syllabes de chaque bloc, donc à des syllabes différentes. Or on sait d'après des études utérieures l'influence de la voyelle sur le temps de détection de la consonne (Foss & Gernsbacher, 1983; Diehl, Kluender, Foss, Parker, & Gernsbacher, 1987).

leur variation gène la décision. Cette idée, selon laquelle c'est l'étape de décision qui est responsable du ralentisement observé entre les blocs expérimentaux et contrôles dans le paradigme de Garner, a été discutée par certains auteurs (p.ex. Miller, 1978, p.179). Le contre-argument classique est qu'il existe des cas où la variabilité ne provoque pas d'interférence. Par exemple, Wood (1974) a trouvé que les sujets n'étaient pas ralentis par une variation phonétique quand ils devaient classifier des syllabes (deux ou quatre) selon leur hauteur mélodique (« pitch »).[5] Cependant, dans les cas où l'on n'a pas observé d'interférence, la dimension à classifier était de nature différente de la dimension qui variait (par exemple, une dimension acoustique et une dimension phonétique). Quand la dimension à classer et la dimension qui varie appartiennent à un même niveau de représentation (ce qui est le cas pour les interférences entre phonèmes observées par Wood et Day), il se pourrait que l'interprétation décisionnelle soit correcte.

Quoiqu'il en soit, si du fait qu'il est impossible de focaliser sélectivement l'attention sur un phonème, on veut déduire que ce dernier n'est pas l'unité perceptive, alors on ne peut accorder ce statut à la syllabe que si l'on montre que l'attention peut être sélectivement focalisée sur elle. Nous avons donc réalisé une expérience utilisant le paradigme de Garner, semblable à celle de Wood et Day (1975), si ce n'est que les phonèmes étaient remplacés par des syllabes. Si l'on observe que les sujets peuvent classifier une syllabe sans être gênés par la variation dans une autre syllabe, alors la thèse « syllabe = unité perceptive » aura gagné du crédit. Si, au contraire, on observe une interférence entre syllabes, il faudra alors supposer : soit que l'unité perceptive est plus grande que la syllabe, soit que l'interprétation décisionnelle avancée plus haut est correcte.

EXPÉRIENCE 1 : CLASSIFICATION DE SYLLABE

Dans cette expérience, nous examinons si l'attention peut être focalisée sélectivement sur la première syllabe de stimuli bisyllabiques CV–CV. Les sujets doivent classifier la première syllabe et ignorer la seconde. Dans les blocs expérimentaux, cette dernière varie alors que dans les blocs contrôles elle reste la même.

5. Toutefois, ils étaient gênés pour effectuer une classification phonétique quand le pitch variait (cependant voir Miller, 1978). Ceci est un exemple d'interférence unidirectionnelle, phénonène particulièrement prisé par les chercheurs car il indiquerait que la dimension qui subit l'interférence est calculée *après* la dimension qui la produit (mais ne la subit pas).

DESCRIPTION

Matériel : Un locuteur masculin a enregistré les quatre stimuli /paku/, /toku/, /pagi/ et /togi/. Avec ceux-ci, nous avons constitué quatre listes aléatoires comprenant chacune 64 items. Deux listes, dites « contrôles », contenaient deux stimuli qui ne différaient que par la première syllabe (la première liste contenait 32 items /paku/ et 32 items /togu/; la seconde liste contenait 32 /pagi/ et 32 /togi/). Deux autres listes, dites « expérimentales », contenaient les quatre stimuli (chacun en 16 exemplaires) ; elles ne différaient que par l'ordre (aléatoire) des items.

Procédure : Chaque sujet était testé individuellement. L'expérience se déroulait entièrement sous le contrôle d'un ordinateur portable, le protocole expérimental étant décrit dans un langage de description d'expérience développé par l'auteur.[6] Les stimuli, stockés sur le disque dur du PC, étaient restitués au sujet par l'intermédiaire d'un casque stéréophonique de haute fidélité AKG.

Au début de l'expérience, le sujet entendait une fois chacun des quatre stimuli. Puis, il lisait les instructions affichées sur l'écran de l'ordinateur. Celles-ci précisaient qu'il entendrait des listes formées avec ces stimuli, qu'il devrait « *se concentrer sur la première syllabe de chaque «mot»* » et « *appuyer sur le bouton de réponse droit pour /PA/ et gauche pour /TO/* ». On lui demandait également de « *de répondre le plus rapidement possible* » mais d'« *éviter de commettre des erreurs* ». Après lecture des instructions, le sujet effectuait un entraînement à la tâche, sur un bloc de 16 essais comprenant les quatre stimuli.

Chaque essai débutait par la présentation auditive d'un stimulus et le sujet avait 1500 msec pour répondre. Puis, il voyait s'afficher son temps de réaction si la réponse était correcte, ou un rappel de l'assignation des cibles aux clés de réponse s'il avait commis une erreur. L'intervalle entre chaque essai était de 3 secondes (augmenté de 800 msec en cas de mauvaise réponse). À la suite de cet entraînement, la partie expérimentale proprement dite commençait : chaque sujet entendait quatre blocs successifs correspondant chacun à une des listes décrites dans la section précédente. À l'intérieur de chaque bloc, les essais avaient une structure identique à ceux du bloc d'entraînement (en particulier, les sujets recevaient du « feed-back » après chaque essai[7]). Au début de chaque bloc, le sujet était autorisé à se détendre pendant environ une minute ; les instructions étaient réaffichées à

6. Voici pour les détails techniques : le PC était un TOSHIBA T5200, équipé d'une carte de conversion analogique/digitale de haute qualité (OROS AU22 : fréquence d'échantillonnage 64 kHz sur 16 bits ; mais les fichiers contenant le signal sont stockés avec un sous-échantillonage d'un facteur 4, c'est à dire à 16 kHz) et d'une carte chronomètre (TC 24 de Real Time Device) mesurant les temps de réaction au cinquième de milliseconde près.

7. Si nous devions aujourd'hui refaire cette expérience, nous supprimerions le retour d'information sur les temps de réaction (« feed-back »), et demanderions aux sujets d'effectuer l'expérience avec les yeux fermés. Nous avons observé introspectivement, aprés avoir effectué une vingtaine de blocs, que l'affichage des temps de réaction pouvaient nous « distraire » et risquait de nuire à la concentration. Notre but, en affichant les temps de réaction, était d'accélérer les sujets.

l'écran et indiquaient quels stimuli allaient être présents dans le bloc à venir tout en rapellant qu'il fallait se concenter sur la première syllabe. Globalement, l'expérience durait environ une vingtaine de minutes. L'ordre des listes était contrebalancé entre les sujets comme indiqué sur la table II.1.

Sujets	Séquence			
1–2	C1	E1	C2	E2
3–4	E1	C1	E2	C2
5–6	C2	E2	C1	E1
7–8	E2	C2	E1	C1

E1=liste exp. 1 ; E2=liste exp. 2
C1=cont. (paku/toku) ; C2=cont. (pagi/togi)

TAB. II.1 – *Contre-balancement des blocs*

Sujets : Huit sujets étudiants de l'EPITA (une école d'ingénieurs en informatique située dans Paris XIII), d'âges compris entre 20 et 24 ans, de langue maternelle française et sans déficits avérés de l'audition, ont participé à cette expérience. Ils recevaient une rétribution de 20 FF pour leur participation.

RÉSULTATS

On a éliminé systématiquement les huit premiers temps de réaction de chaque bloc (considérés comme des « échauffements ») et on a pris en compte uniquement les cinquante-six suivants. Puis on a calculé, pour chaque sujet et dans chaque bloc, le taux d'erreurs (comptabilisant les absences de réponse et les appuis sur la mauvaise clé) et le temps de réaction moyen sur les réponses correctes.[8] Les moyennes sont présentées dans la table II.2.

Un simple t-test révèle que le ralentissement de 38 ms entre les blocs expérimentaux et les blocs contrôles est significatif ($t(7) = 4.5, p = .003$). Par contre, la différence de taux d'erreurs n'est pas significative ($t(7) = .3, p = .8$). Dans un deuxième temps, nous avons conduit une Anova sur les temps moyens de réaction, en incluant, en plus du facteur « Expérimental » (X) qui distingue les blocs expérimentaux et les blocs contrôles, deux autres facteurs :

– « Ordre » (O). Chaque sujet passait quatre blocs : deux expérimentaux et

8. Il est courant, avant de calculer des moyennes de temps de réaction, de supprimer les temps de réaction très rapides (typiquement, inférieurs à 100 msec), qui sont considérés comme des anticipations, et ceux particulièrement lents (typiquement, plus de 1500 msec), qui sont considérés comme résultant de défauts d'attention du sujet. Dans cette expérience, aucun temps de réaction n'était inférieur à 100 msec, et les temps supérieurs à 1500 msec n'étaient pas enregistrés (ils étaient dans la catégorie « non-réponse »).

	Type de bloc		
	Contrôle	Expérimental	Différence
Temps	390 ms	428 ms	38 ms
Erreurs	7.1 %	6.7 %	-0.4 %

Chaque cellule est la moyenne d'environ 896 données (8 sujets × 112 mesures moins les erreurs). L'écart-type entre sujets est de 74 msec; L'écart-type moyen à l'intérieur d'un bloc (intra-sujet) est de 119 msec.

TAB. II.2 – *Temps de Réaction (en msec) et Taux d'Erreurs (en %) — Classification de Syllabe initiale.*

deux contrôles. le facteur Ordre déterminait si le bloc appartenait à la première moitié de l'expérience (bloc 1 ou 2) ou à la seconde moitié de l'expérience (bloc 3 ou 4). Une interaction avec le facteur Expérimental pourrait révéler un effet diminuant avec l'entraînement.

– « Moitié de bloc » (M). Nous avons séparé, à l'intérieur de chaque bloc, les temps de réaction appartenant à la première et à la seconde moitié du bloc, et calculé les temps moyens sur chacune de ces moitiés. Si, à l'intérieur d'un bloc expérimental, le sujet devient de plus en plus capable de focaliser son attention sur la première syllabe, alors on pourrait observer une interaction Moitié × Expérimental.

L'Anova suivait un plan X2*O2*M2, les trois facteurs étant déclarés intra-sujet. Le tableau d'anova est détaillé plus bas.[9] L'effet du facteur Expérimental est significatif (conformément au t-test). Par contre, ni l'effet d'Ordre (17 ms, $F(1,7) = 1.9, p = .2$), ni l'effet de Moitié de bloc (15 ms, $F(1,7) = 2.6, p = .15$) n'atteignent la significativité. De plus, ils n'interagissent pas avec l'effet Expérimental. Les effets expérimentaux restreints respectivement à la première moitié et à la seconde moitié des blocs s'avèrent significatifs (X/M1 : 48 ms, $F(1,7) = 13.3, p = .008$; X/M2 : 38 ms, $F(1,7) = 6.0, p = .05$). L'effet expérimental est significatif dans la seconde moitié de l'expérience (pour les blocs 3–4 : X/O2 : 45 ms, $F(1,7) = 5.9, p = .05$) ; il ne l'est pas dans la première moitié (X/O1 :

9. Dans cette thèse, nous fournissons les tableaux d'anova (quand ils ne sont pas trop long) et ne mentionnons pas systématiquement les valeurs de F et p dans le texte. Cela rend celui-ci plus lisible, et fournit une information plus complète que la présentation habituelle qui ne détaille pas les effets non significatifs. Le carré moyen d'erreur (MSE, de l'anglais « Mean Square Error ») permet de se faire une idée sur la variabilité et de comparer celle-ci avec celle des autres tests. La notation est celle du logiciel VAR3 : l'effet d'un facteur (p.ex. 'X') est noté par le nom de celui-ci. Les interactions sont signalées par des points (p.ex. 'X.O'), et les restrictions, par le signe '/' : 'X/M1' signifie « effet du facteur X, restreint à la première modalité du facteur M ».

31 ms, $F(1,7) = 1.8, p = .22$), mais cela résulte vraisemblablement d'une plus grande variabilité (cf MSE). Une anova similaire sur les erreurs ne révèle aucun effet significatif.

```
Temps de réaction: PLAN  S8*M2*O2*X2
X = Expérimental
O = Ordre (O1=blocs 1-2 et O2=blocs 3-4)
M = Moitié de bloc (M1 = début et M2 = fin)
X              F(1,7)= 19.56    MSE=1192.28    p=0.0031
O              F(1,7)=  1.85    MSE=2512.64    p=0.2160
O.X            F(1,7)=  0.13    MSE=6028.32    p=0.2709
M              F(1,7)=  2.57    MSE=1466.79    p=0.1529
M.X            F(1,7)=  1.33    MSE=1262.68    p=0.2867
M.O            F(1,7)=  0.95    MSE=401.013    p=0.3622
M.O.X          F(1,7)=  0.01    MSE=755.376    p=0.0769
X/M1           F(1,7)= 13.30    MSE=1409.29    p=0.0082
X/M2           F(1,7)=  5.97    MSE=1045.67    p=0.0445
X/O1           F(1,7)=  1.76    MSE=4462.34    p=0.2263
X/O2           F(1,7)=  5.89    MSE=2758.27    p=0.0456
```

Discussion

Les sujets sont plus lents pour décider quelle est la première syllabe d'un stimulus CV–CV quand les blocs contenaient /paku/,/toku/, /pagi/ et /togi/, que quand ils contenaient, soit /paku/ et /toku/, soit /pagi/ et /togi/. Bien qu'on leur demande de se concentrer sur la première syllabe, les sujets ne semblent pas capables d'ignorer la variation de la seconde syllabe.[10] Cet effet ne semble pas diminuer avec l'entraînement (entre les blocs ou à l'intérieur de chaque bloc).[11]

Nous avons proposé dans l'introduction qu'on pouvait interpréter l'effet d'interférence Garner sur le phonème initial (chez Wood & Day et Tomiak et al.) de deux manières : (a) en supposant que l'unité perceptive était plus grande que le phonème ou, (b) en attribuant l'effet à l'étape de décision dans ce type de tâche, qui ferait qu'on ne peut focaliser son attention sur l'« objet » à classer quand un autre « objet » varie sur la même surface de représentation. Avec l'effet d'interférence sur la syllabe que nous venons de dévoiler, les deux raisonnements peuvent

10. Ils ne semblent pas non plus capables de se focaliser sur le phonème, puisque dans cette tâche, ils auraient pu répondre sur la base de la première consonne ou même de la première voyelle.

11. À titre anecdotique, signalons que nous nous sommes entraîné pendant une heure (16 blocs), atteignant un temps de réaction moyen de 240 msec, et que malgré cela, il demeurait une différence significative entre blocs expérimentaux et blocs contrôles.

encore être proposés : on peut dire (a) qu'il faut proposer une unité perceptive plus grande que la syllabe, ou bien (b) que c'est l'étape de décision qui est la principale responsable de l'interférence.

Toutefois, avant d'examiner cette alternative, il nous faut écarter une autre interprétation possible du résultat : on pourrait argumenter que les sujets arrivent effectivement à focaliser leur attention sur la première syllabe mais que c'est la *variabilité acoustique* de cette syllabe qui les ralentit. Par exemple, on sait que dans une tâche de comparaison (« matching »), le temps pour comparer deux syllabes est plus important si celles-ci sont physiquement différentes (Pisoni & Tash, 1974). Dans cette optique, il n'est même pas besoin d'invoquer la syllabe : on pourrait avancer que le sujet se forme des *images acoustiques* du début du signal pour chaque stimuli et compare celles-ci avec le stimulus présenté à chaque essai. Il y a alors deux images à comparer dans la condition contrôle et quatre dans la condition expérimentale, ce qui rendrait compte de l'effet. L'étude de Tomiak et al. citée plus haut a montré la limite des explications acoustiques dans la tâche « à la Garner ». Néanmoins, nous avons décidé de répliquer notre expérience en construisant des stimuli dont la première syllabe est physiquement identique à l'intérieur des paires (paku, pagi) et (toku, togi).

EXPÉRIENCE 2: CLASSIFICATION DE SYLLABES SANS VARIATION ACOUSTIQUE

DESCRIPTION

Matériel : Les stimuli de l'expérience précédente ont été utilisés. À l'aide d'un éditeur de son, on a extrait les syllabes : /pa/ de /paku/, /to/ de /toku/, /gi/ de /pagi/ et /ku/ de /toku/. En « recollant » les morceaux, on a obtenu quatre nouveaux stimuli /paku/, /toku/, /pagi/ et /togi/. Le collage n'a pas occasionné de clic et les stimuli paraissent naturels.

Procédure : Elle était en tous points identique à la précédente.

Sujets : Huit sujets étudiants, d'âges compris entre 23 et 27 ans, ont participé volontairement à cette expérience. Ils n'avaient pas participé à la précédente.

RÉSULTATS

Les données ont été préparées comme dans l'expérience précédente. Les résultats moyens figurent dans la table II.3.

Les t-tests révèlent que l'effet Expérimental est significatif sur les temps de réaction (47 ms, $t(7) = 3.8, p = .007$), mais pas sur les erreurs ($t(7) = .5, p = .6$). Des Anovas similaires à celles de l'expérience précédente ont été conduites, sur

	Type de bloc		
	Contrôle	Expérimental	Différence
Temps	321 ms	368 ms	47 ms
Erreurs	4.4 %	3.1 %	-1.3 %

Chaque cellule est la moyenne d'environ 896 données (8 sujets × 112 mesures moins les erreurs). L'écart-type entre sujets est de 60 msec; la moyenne des écart-types intra-sujet et intra-bloc est de 101 msec.

TAB. II.3 – *Temps de Réaction (en msec) et Taux d'Erreurs (en %) — Classification de Syllabe initiale.*

les temps de réaction et les erreurs, avec les trois facteurs intra-sujet : Expérimental, Ordre et Moitié de bloc. Elles révèlent que seul l'effet Expérimental est significatif dans les temps de réaction. Les facteurs Ordre et Moitié ne produisent ni effet ni interaction significative.

```
Temps de réaction
PLAN  S8*M2*O2*T2
X = Expérimental
O = Ordre
M = Moitié de bloc (M1 = début et M2 = fin)
X             F(1,7)= 15.26    MSE=2335.09    p=0.0059
O             F(1,7)=  2.97    MSE=6154.25    p=0.1285
O.X           F(1,7)=  1.87    MSE=1317.9     p=0.2138
M             F(1,7)=  0.06    MSE=1055.28    p=0.1865
M.X           F(1,7)=  0.77    MSE=297.997    p=0.4093
M.O           F(1,7)=  1.73    MSE=672.777    p=0.2299
M.O.X         F(1,7)=  0.14    MSE=999        p=0.2806
X/M1          F(1,7)= 14.49    MSE=1434.16    p=0.0067
X/M2          F(1,7)= 12.57    MSE=1198.92    p=0.0094
X/O1          F(1,7)=  8.90    MSE=3193.48    p=0.0204
X/O2          F(1,7)= 21.07    MSE=459.516    p=0.0025
```

Une Anova supplémentaire rassemblant les données de cette expérience et de la précédente et déclarant un facteur « Expérience », révèle que celui-ci ne produit aucune interaction significative, et que la différence de temps moyen entre les deux expériences (comparaison entre sujets) est marginale (63 ms, $F(1,14) = 3.68, p = .08$).[12] Dans cette analyse, le facteur Ordre devient significatif ($F(1,14) =$

12. Il ne s'agit pas d'une compensation avec les erreurs (« speed-accuracy tradeoff »),

$5.02, p = .04$), mais le facteur Moitié ne produit aucun effet. Aucune interaction n'est significative dans cette analyse et l'effet expérimental est significatif dans toutes les restrictions (début ou fin d'expérience, début ou fin de bloc). L'Anova détaillée se trouve en annexe (p.141).

DISCUSSION

Comme dans l'expérience précédente, les sujets sont gênés par la variation de la seconde syllabe quand ils doivent classifier la première syllabe de stimuli CV–CV. Cette expérience-ci montre que l'effet ne peut pas s'expliquer par la variabilité acoustique de la première syllabe, puisque il n'y avait qu'un unique exemplaire physique de /pa/ et qu'un unique exemplaire physique de /to/ dans chaque liste.

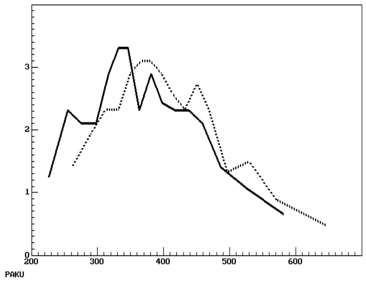
On pourrait imaginer que, dans les blocs expérimentaux, il soit possible de focaliser *quelquefois* son attention sur la première syllabe mais *pas dans tous les essais*. Selon cette hypothèse, il y aurait deux types d'essais : ceux où l'attention est correctement engagée et où le sujet est rapide, et ceux où l'attention n'est pas correctement engagée et où le sujet est lent. La différence entre les blocs expérimentaux et contrôles devrait donc être due essentiellement aux temps de réaction lents. Si c'est le cas, on s'attend à ce que les distributions de probabilité des temps de réaction soient dans la relation indiquée sur la figure 2a. Si, au contraire, le ralentissement est global, les distributions devraient plutôt ressembler à la figure 2b.

Fig. a Fig. b

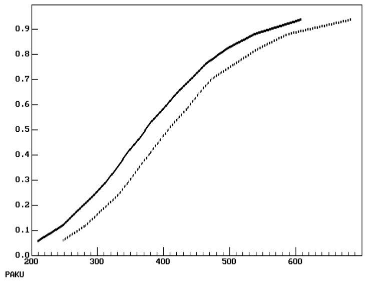
FIG. II.1 – *Distributions Théoriques des Temps de Réaction*

Nous avons donc tracé les distributions des temps de réaction *bruts* des blocs expérimentaux et contrôles des deux expériences (896 données par courbes). À gauche, on a indiqué la densité de probabilité des réponses en fonction du temps de réaction ; à droite, on a représenté la distribution cumulée correspondante, qui lisse les « accidents ».

puisque les sujets commettent également moins d'erreurs dans l'expérience 2 (quoique non significativement).

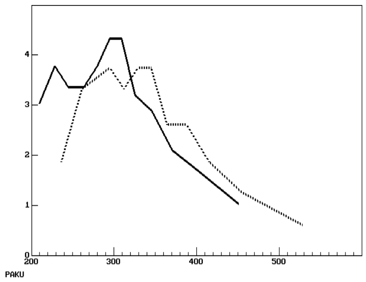


Densité

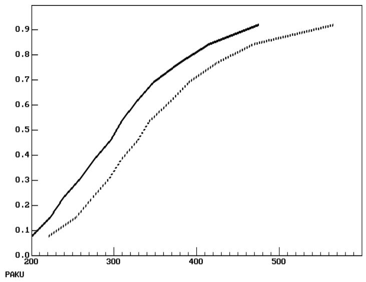


Dist. Cumulées

2



Densité



Dist. Cumulées

En abscisse : Temps de réaction; En ordonnée : densité de TR / msec. Le trait plein correspond aux blocs contrôles; le trait pointillé aux blocs expérimentaux.

FIG. II.2 – *Distributions des TR. Exp.1 et 2.*

Comme on peut le voir sur la figure II.2, la situation dans les deux expériences est plutôt celle de la fig.2b : la distribution dans la condition expérimentale est décalée *globalement* par rapport à la distribution dans la condition contrôle. L'effet ne provient pas uniquement des temps de décision lents : il est présent sur les temps les plus rapides qui sont de l'ordre de 200 msec. Par conséquent, l'hypothèse que le ralentissement dans les blocs expérimentaux serait dû à un échec attentionnel restreint seulement aux essais les plus lents, peut être écartée. [13]

Le résultat principal des expériences 1 et 2 est que les sujets ne sont pas parvenus à focaliser leur attention sur la première syllabe des stimuli. L'expérience

13. Le ralentissement est-il dû à l'identification de la seconde syllabe de chaque stimuli ? Ou bien correspond-t-il à un ralentissement global dans le bloc ? Aux temps les plus rapides, on peut se demander si les sujets ont entendu la seconde syllabe. En fait la première syllabe avait une durée de 100 msec dans l'expérience 2, et l'on peut penser que les sujets ont donc effectivement pu être influencés par la seconde syllabe (en leur accordant un temps moteur un peu inférieur à 100 msec). Pour répondre à la question posée plus haut, il faudrait refaire une expérience identique avec des stimuli dont la première syllabe serait nettement plus longue et voir si l'effet disparaît aux temps de réaction les plus rapides.

2 montre qu'ils n'ont pas utilisé une représentation de l'acoustique du début des stimuli. Le caractère remarquable de ce dernier résultat mérite d'être souligné : depuis l'étude de Pisoni et Tash (1974), où les sujets appariaient plus rapidement deux syllabes physiquement identiques que deux syllabes physiquement différentes, il est couramment admis que les sujets peuvent accéder à « de l'information acoustique de bas niveau en même temps qu'à une représentation phonétique plus abstraite » (Pisoni et Tash 1974, p. 290; voir également, par exemple, Samuel, 1977; Miller, 1994). On aurait pu s'attendre à ce que, comme un très faible nombre de stimuli est employé dans nos expériences, les sujets allaient pouvoir exploiter l'information acoustique de début de stimlus pour répondre. Ils ne l'ont apparemment pas fait (pas plus qu'ils ne pouvaient le faire dans l'étude de Tomiak et al., du moins en mode « parole »).

Pourquoi les sujets étaient-ils capables d'utiliser l'information acoustique dans l'étude de Pisoni et Tash mais pas dans la nôtre ? Une première remarque est que nos stimuli étaient bisyllabiques. On peut imaginer que la seconde syllabe masque rétroactivement la première et que les sujets ne pouvaient donc utiliser une représentation acoustique pour répondre. Cette hypothèse fait la prédiction intéressante que l'effet obtenu par Pisoni et Tash devrait disparaître si l'on demande à des sujets de comparer les syllabes initiales de stimuli multisyllabiques. Cependant, le masquage rétroactif ne peut expliquer l'interférence observée dans leurs expériences par Tomiak et al. (1987), qui utilisaient, eux, des monosyllabes.

Une seconde remarque est qu'il existe des cas où les sujets effectuent la tâche de comparaison sans, apparemment, utiliser de l'information de bas niveau. Forster (1979) a proposé un modèle pour expliquer les faits suivants :

– les sujets reconnaissent que deux formes visuelles sont la même lettre plus rapidement quand elles sont physiquement identiques (p.ex. AA) que quand elles ne le sont pas (p.ex. Aa) (Posner & Mitchell, 1967). L'interprétation classique est que la comparaison peut être effectuée au niveau, précoce, d'analyse des traits visuels quand les stimuli sont physiquement identiques, mais qu'il faut atteindre un niveau plus abstrait pour la comparaison "aA". D'ailleurs, Pisoni et Tash (1974) proposent une explication similaire pour expliquer que deux syllabes physiquement identiques sont appariées plus rapidement que deux syllabes physiquement différentes.
– quand il s'agit de comparer des suites de lettres, les sujets comparent plus rapidement des suites formant un mot (p.ex. DUST DUST), que des suites formant un non-mot (p.ex. FTKL FTKL). Or dans les deux cas, on se serait attendu à ce que l'appariement ait pu être effectué au niveau des traits visuels, et donc que les temps de décisions soient similaires.

Forster (1979) propose de résoudre cette contradiction apparente en « considérant comment la décision est effectuée : les objets sont comparés, simultanément, à différents niveaux d'analyse. Quand les formes sont assez simples (comme dans le cas des lettres individuelles), la comparaison peut être réalisée rapidement au niveau le plus bas des traits visuels, au point que les plus hauts niveaux d'analyse ont peu de chance de fournir une réponse avant ce niveau là. C'est donc le niveau des traits qui contrôle la décision. Par contre, si les stimuli sont plus complexes (p.ex. des mots), la comparaison au niveau des traits visuels devient vraisemblablement trop lente (car il y a de nombreux traits à comparer), au point que la réponse peut alors provenir des hauts niveaux, sur lesquels le nombre d'unités à comparer est inférieur. » (p.32–33, notre traduction). Une prédiction de ce modèle est que, dans la comparaison des suites de lettres, le niveau de traitement le plus bas pourrait quand même contrôler la réponse *pour les comparaisons négatives* : dès que deux traits physiques diffèrent, le sujet peut être assuré de la réponse (négative). En fait, il a été vérifié expérimentalement qu'il n'y a pas d'effet de lexicalité sur les temps de réaction des réponses « différent ». [14]

Dans nos expériences de classification de la première syllabe de stimuli CVCV, les sujets sont dans une situation proche de la tâche de comparaison. En effet, comme il y a peu de stimuli différents (2 ou 4), on peut concevoir que les sujets sont capables de les mémoriser tous, et, qu'à chaque essai, ils effectuent des comparaisons entre l'item présenté et les représentations mémorisées pour chaque item. Cependant, soit parce que les stimuli sont trop complexes acoustiquement, soit parce que les sujets ne peuvent en mémoriser tous les détails, la comparaison serait plus efficace à un niveau plus abstrait (i.e. où il y a moins d'unités à comparer) qu'à un niveau purement acoustique. Dans nos expériences, ce serait au niveau où le stimulus ne forme qu'une unité (bisyllabique) que la comparaison serait la plus efficace, et ce niveau contrôlerait la réponse. [15] Le modèle de Forster

14. C'est cela qui permet de conserver l'hypothèse que la représentation en traits est calculée *avant* des représentations plus abstraites, et qui distingue le modèle de Forster de modèles moins contraints qui supposent simplement que tous les niveaux de représentation influencent simultanément la décision.

15. Cette proposition rappelle celle qui a été avancée pour expliquer le fait que le temps de détection d'une unité linguistique (phonème, syllabe, mot) *diminue* quand *augmente* la taille de l'unité à détecter (TR mot < TR syllabe < TR phonème) (cf Savin & Bever, 1970; Foss & Jenkins, 1973; McNeill & Lindig, 1973). Une conception répandue est que cela serait dû au fait que les niveaux de traitement les plus abstraits sont accessibles plus facilement pour les décisions conscientes. Toutefois, rien ne force à cette interprétation. Une autre hypothèse, tout aussi vraisemblable, est que ce résultat peut être attribué à un effet de redondance : quand les sujets doivent détecter /pal/, ils peuvent, dans la plupart des expériences, répondre dès qu'ils détectent /p/ *ou* /a/ *ou* /l/, car il n'y a pas de pièges (des stimuli commençant, par exemple, par /pat/). On comprend alors que cette décision puisse être plus rapide que la detection de /p/. Norris et Cutler (1988) ont montré quand on force les sujets à faire une analyse complète de la syllabe (en introduisant des pièges ; les sujets doivent alors détecter (p.ex.) : /p/ *et* /a/ *et* /l/), les temps de détection de phonème peuvent devenir inférieurs aux temps de détection de syllabe.

permet donc de concilier nos résultats avec l'hypothèse que le système perceptif récupère les phonèmes ou les syllabes avant le stimulus « global ». Si ce modèle est correct, il prédit un ralentissement dans tous les blocs contenant quatre stimuli par rapport aux blocs contenant deux stimuli, et suggère que la tâche à la Garner ne permet pas de contraindre fortement les modèles de traitement de la parole.[16]

Il nous semble, cependant, que le paradigme de base peut être amélioré. Intuitivement, l'attention peut être focalisée sur un objet plus efficacement dans certains contextes plutôt que dans d'autres. Cela devrait se traduire par des coûts de variations différents selon les types d'interférences : *certaines variations devraient coûter plus que d'autres* à la prise décision. C'est cette intuition qui nous a conduit à réaliser l'expérience qui suit : si les phonèmes appartenant à la même syllabe sont « liés » entre-eux, il doit être plus difficile de focaliser son attention sur un phonème quand il y a une variation à l'intérieur de la syllabe qui le contient plutôt qu'à l'extérieur de celle-ci.

EXPÉRIENCE 3: INTERFÉRENCES EXTRA- ET INTRA-SYLLABIQUES EN CLASSIFICATION DE VOYELLE

La classification d'une voyelle est-elle affectée différemment quand on fait varier une consonne (a) dans la même syllabe et (b) dans une autre syllabe ? Dans cette expérience, les sujets doivent classifier soit la première, soit la seconde voyelle de stimuli VCCV. Ceux-ci peuvent appartenir à l'une des deux structures VC–CV (p.ex. /ac–ti/) ou V–CCV (p.ex. /a–cli/).

Notre manipulation consiste à faire varier ou non la première consonne du groupe CC central. Si les phonèmes appartenant à la même syllabe sont plus étroitement « liés » que des phonèmes appartenant à des syllabes différentes, alors on s'attend à ce que la variation *intra-syllabique* gêne plus les sujets que la variation *extra-syllabique* ; autrement dit, l'interférence devrait être maximale pour classer V_1 dans la structure V_1C–CV_2, et V_2 dans la structure V_1–CCV_2.

Il faut insister sur le fait qu'il est rationnel de la part des sujets d'utiliser la redondance dans les tâches de détection (quand il n'y pas de piège). Par contre, dans nos expériences, les sujets ne gagnent rien à prendre en compte la fin des stimuli : celui-ci ne sert a rien pour la réponse. Nos résultats ne peuvent donc s'expliquer par un effet de redondance.

16. En passant, on peut faire remarquer que le modèle de Forster permet d'expliquer l'effet de congruence syllabique en détection de fragment, sans supposer que la syllabe est *reconnue* avant le phonème (par le système de traitement). Rapellons que les sujets détectent plus facilement /pa/ dans /pa-lace/ que dans /pal-mier/, et inversement pour /pal/ (Mehler et al., 1981). Cela pourrait s'expliquer par le fait que la comparaison « gagne » au niveau syllabique qui contient moins d'unités que le niveau phonémique.

DESCRIPTION

Matériel : Tous les stimuli avaient le format VCCV. La classification portait soit sur la première, soit sur la seconde voyelle. On a choisi /a/ et /u/ comme voyelles à classifier. La variation à ignorer était localisée dans la première consonne, qui pouvait être, soit le coda de la première syllabe (structure VC–CV), soit l'attaque de la seconde syllabe (structure V–CCV). Le choix s'est porté sur les plosives /k/ et /p/ car elles peuvent apparaître aussi bien en coda de syllabe (kt/pt) qu'en début de groupe commençant une syllabe (kl/pl).[17] Le matériel est décrit dans la table II.4.

	Voyelle à Classer (a/u)	
Structure	V1	V2
VC.CV	acti/ucti	icta/ictu
	apti/upti	ipta/iptu
V.CCV	acli/ucli	icla/iclu
	apli/upli	ipla/iplu

Nous utilisons la lettre « c » plutôt que « k », respectant ainsi l'orthographe la plus fréquente en français.

TAB. II.4 – *Stimuli de classification de voyelle*

On a construit 8 listes contrôles et 4 listes expérimentales. Dans les listes contrôles, seule la voyelle variait ; ces listes étaient donc formées à partir de deux items (par exemple: acti/ucti). Dans les listes expérimentales, la consonne variait également (p ou c); ces listes étaient donc formées à partir de 4 items (acti/ucti/apti/upti). Toutes les listes possédaient 64 essais dans un ordre aléatoire. En résumé, il y avait :

– 4 listes expérimentales définies par (a) la position de la voyelle à classer, (b) la position de la consonne qui variait : coda de la première syllabe ou attaque de la seconde.
– 8 listes contrôles définies par les mêmes facteurs et également par la nature de la consonne post-vocalique (c ou p) maintenue constante.

Procédure : La procédure était identique à celle de l'expérience précédente.

Il y avait deux groupes de sujets : ceux qui classifiaient la première voyelle et ceux qui classifiaient la seconde voyelle. Pour chaque groupe, il y a quatre blocs contrôles et deux

17. La fricative /f/ peut également se trouver coda ou en début d'attaque complexe (mou–flon, caf–tière) ; mais /ch/ et /s/ ne peuvent pas être employés. /s/ peut se trouver en coda (plas-ma) mais il n'est pas évident qu'il puissent débuter un groupe CC de début de syllabe : « castor » doit-t-il être syllabifié en « ca-stor » ou bien en « cas-tor ». Notons que la syllabification des groupes /pt/ et /kt/ est quelquefois discutée par les linguistes (cf 108). Ici, nous nous fondons essentiellement sur les intuitions pour supposer que la coupe syllabique tombe entre les deux obstruentes.

expérimentaux. Pour limiter le nombre de blocs à quatre par sujet, chacun n'a effectué que deux blocs contrôles. Pour contre-balancer l'ordre de chaque bloc, on a eu recours au dessin décrit dans la table II.5.

Sujet	Séquence			
1	E1	C1	E2	P2
2	C1	E1	P2	E2
3	E2	P2	E1	C1
4	P2	E2	C1	E1
5	E1	P1	E2	C2
6	P1	E1	C2	E2
7	E2	C2	E1	P1
8	C2	E2	P1	E1

E=expérimental ; C=contrôle /c/ ; P=contrôle /p/ ;
1=VC–CV ; 2=V–CCV. Par exemple :
P2=(apli/upli), E1=(acti/upti/apti/upti).

TAB. II.5 – *Contrebalancement des blocs*

Sujets : Seize étudiants, d'âges compris entre 20 et 25 ans, ont participé à cette expérience, pour laquelle ils recevaient 20 FF.

RÉSULTATS

On a calculé les temps moyens et taux d'erreurs de chaque sujet dans chaque bloc, après avoir supprimé les huit premiers essais. La figure II.3 présente les coûts de variation obtenus en soustrayant les temps de réaction des conditions contrôles à ceux des conditions expérimentales. La table II.6 fournit les moyennes détaillées.

	$V_1C–CV_2$		$V_1–CCV_2$	
Voyelle classée	Cont	Expé	Cont	Expé
V_1	378 (2.0)	427 (2.9)	397 (1.8)	409 (2.5)
V_2	297 (2.5)	295 (2.8)	261 (5.4)	277 (4.5)

L'écart-type entre sujets est de 85 msec. La moyenne des écarts-types intra-sujet, intra-bloc est de 102 msec.

TAB. II.6 – *Temps de Réaction (et Pourcentage d'Erreurs) en Classification de Voyelle*

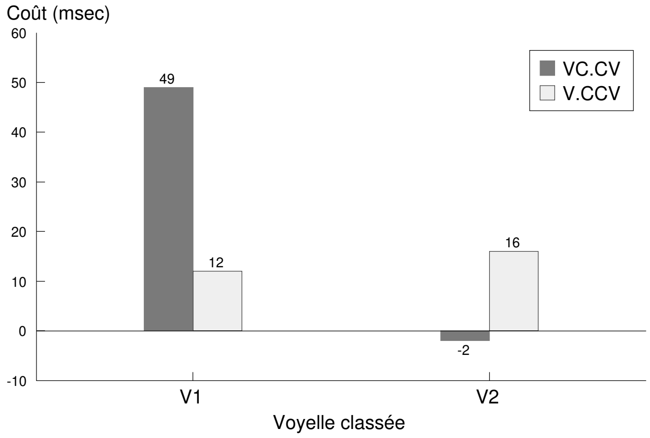


FIG. II.3 – *Coût de la variation de consonne en fonction de la voyelle classée et de sa position dans la structure syllabique*

Deux ANOVAs ont été effectuées : l'une sur les temps de réaction et l'autre sur les erreurs. Trois facteurs étaient définis : (a) Voyelle correspondant à la voyelle classée : V1 ou V2 (inter-sujets) ; (b) Structure : VC–CV ou V–CCV (intra-sujet) ; (c) Expérimental correspondant au type de bloc (Expérimental ou Contrôle) (intra-sujet).

Le facteur Expérimental produit un effet significatif : les sujets sont 19 msec plus lents dans les blocs expérimentaux que dans les blocs contrôles. L'effet de Voyelle est massif : les sujets qui classifient V_2 sont 120 msec plus rapides que ceux qui classifient V_1. Enfin, il y a une triple interaction significative Voyelle $\times$ Structure $\times$ Expérimental ; celle-ci signifie que, pour les coûts de la variabilité (égaux aux différences « expérimental - contrôle »), il y a une interaction entre Voyelle et Structure syllabique : les coûts sont plus importants quand la voyelle à classer se trouve dans la syllabe où la consonne varie. Les analyses restreintes à chaque groupe de sujets montrent que ceux qui classifiaient la première voyelle étaient gênés quand la consonne qui variait était dans la même syllabe (structure VC–CV), mais pas quand elle était dans la deuxième syllabe (V–CCV). À l'inverse, pour les sujets qui classifiaient la seconde voyelle, l'effet de la variation consonantique n'est pas significatif pour la structure VC–CV, mais il est marginal pour la structure V–CCV. Dans l'analyse des erreurs, la seule contribution significative provient d'une interaction Voyelle $\times$ Structure due au fait que les sujets classifiant la seconde voyelle font relativement plus d'erreurs quand la variation de la consonne est dans la seconde syllabe.

```
PLAN  S8<V2>*S2*T2
```

```
V = Voyelle classée (V1 ou V2)
S = Structure :  S1 = VC.CV;  S2 = V.CCV
X = Type de bloc (Expérimental vs controle)
temps de réaction
X           F(1,14)=  9.84    MSE=562.322    p=0.0073
S           F(1,14)=  1.94    MSE=1373.14    p=0.1854
S.X         F(1,14)=  0.84    MSE=439.624    p=0.3749
V           F(1,14)= 15.37    MSE=15016.3    p=0.0015
V.X         F(1,14)=  3.78    MSE=562.322    p=0.0722
V.S         F(1,14)=  2.15    MSE=1373.14    p=0.1647
V.S.X       F(1,14)=  7.06    MSE=439.624    p=0.0188
X/V1/S1     F(1,7)=   8.51    MSE=1121.59    p=0.0224
X/V1/S2     F(1,7)=   2.14    MSE=242.297    p=0.1869
X/V2/S1     F(1,7)=   0.05    MSE=358.973    p=0.1706
X/V2/S2     F(1,7)=   3.74    MSE=281.036    p=0.0944
Erreurs :
X           F(1,14)=  0.20    MSE=1.2634    p=0.3384
S           F(1,14)=  2.90    MSE=1.7455    p=0.1107
S.X         F(1,14)=  0.25    MSE=2.2634    p=0.3752
V           F(1,14)=  3.04    MSE=3.4777    p=0.1031
V.X         F(1,14)=  1.24    MSE=1.2634    p=0.2842
V.S         F(1,14)=  5.16    MSE=1.7455    p=0.0394
V.S.X       F(1,14)=  0.11    MSE=2.2634    p=0.2549
X/V1/S1     F(1,7)=   1.00    MSE=1         p=0.3506
X/V1/S2     F(1,7)=   0.80    MSE=0.7054    p=0.4008
X/V2/S1     F(1,7)=   0.13    MSE=0.4911    p=0.2709
X/V2/S2     F(1,7)=   0.21    MSE=4.8571    p=0.3393
```

Discussion

Le résultat principal de cette expérience est attesté par la triple interaction : les sujets sont plus gênés pour classifier une voyelle si l'on fait varier une consonne qui se trouve dans la même syllabe plutôt qu'une consonne qui se trouve dans une autre syllabe. L'interférence est relativement moins importante pour les sujets qui classifient la seconde syllabe que pour ceux qui classifient la première. Cela pourrait être dû à la distance en nombre de phonèmes entre le lieu de variation et le lieu de classification (2 phonèmes quand on classifie V_2 contre 1 phonème quand on classifie V_1). On peut également remarquer que V_2 est traitée nettement plus rapidement que V_1. Cela peut provenir de nombreux facteurs dont les plus plausibles sont : (a) un effet de baisse du critère de décision quand on progresse à l'intérieur des stimuli (Luce, 1986) ; (b) la présence d'indices acoustiques sur l'identité de la seconde voyelle, dus à une coarticulation anticipatrice, et présents avant le début de la périodicité vocalique (à partir duquel est mesuré le temps de réaction).

On ne peut pas rendre compte des interférences observées en supposant simplement que les sujets effectuent une comparaison globale des stimuli. Pour les expliquer, il faut faire référence à la composition interne des stimuli : dans les blocs expérimentaux avec variation « intra-syllabique », quatre syllabes pouvaient apparaître dans la position examinée par les sujets (p.ex. AC-ti, AP-ti, UC-ti, UP-ti) ; dans les blocs « extra-syllabiques », il n'y en avait que deux (p.ex A-cli, A-pli, U-cli, U-pli). On pourrait envisager que les sujets répondent en utilisant exclusivement la syllabe qui contient le phonème, mais cela n'explique pas pourquoi l'interférence paraît plus faible en seconde syllabe qu'en première [18], et surtout, cette hypothèse est démentie par les expériences précédentes : celles-ci montraient que l'attention ne peut pas être focalisée parfaitement sur une syllabe. Dans le cadre du modèle de Forster, le niveau syllabique serait donc un niveau parmi d'autres qui influencent la réponse. Cependant, c'est un niveau qui semble calculé *automatiquement*, puisqu'il influence les sujets dans une tâche qui ne requiert pas explicitement la manipulation de syllabe (exp.3). En cela, notre résultat est à rapprocher de l'effet de complexité syllabique en détection de phonème initial (Segui et al., 1990) : dans les deux cas, les sujets doivent effectuer une tâche sur des phonèmes et s'avèrent sensibles à la structure de la syllabe qui les contient.

Discussion Générale

Dans le paradigme de Garner, le sujet doit classifier des objets (p.ex. des syllabes CV) selon une dimension (p.ex. l'identité du premier phonème : la consonne C). Dans certains blocs d'essais, une seconde dimension (p.ex. la voyelle V) varie, alors que dans d'autres blocs, elle ne varie pas. Si les sujets sont ralentis par la variabilité de la seconde dimension, ou, en d'autres termes, s'ils sont incapables de focaliser leur attention sur la dimension à classifier, alors on considère généralement que cela est la preuve d'un traitement perceptif holistique des stimuli (Garner, 1974). Ainsi, le fait que les sujets soient gênés pour classifier une consonne quand la voyelle varie semblait un argument en faveur de la syllabe en tant qu'unité de perception (Wood & Day, 1975). Un tel raisonnement prédit que les sujets doivent pouvoir classifier une syllabe sans être dérangés par la variation d'une syllabe adjacente. En fait, l'expérience 1 montre que les sujets ne peuvent pas ignorer la variation de la seconde syllabe quand ils doivent classifier la première syllabe de stimuli CV–CV. L'expérience 2 montre que cela n'est pas dû à un effet de coarticulation entre la seconde et la première syllabe, et montre au pas-

18. Vers la fin du stimulus, l'analyse au niveau global pourrait être achevée et contrôler la réponse, faisant ainsi « disparaître » les effets syllabiques.

sage que les sujet ne peuvent se focaliser sur l'acoustique du signal pour effectuer la tâche.[19]

Notre interprétation favorise une explication décisionnelle des effets d'interférence : les sujets ne peuvent s'empêcher d'être distraits par la variation (cf discussion de l'expérience 2). Si l'on réalisait ce type d'expérience dans la modalité visuelle (en présentant sur un écran "PA", "TO"...), il est probable que la variation gênerait également les sujets, bien qu'il existe des invariants de forme qui caractérisent la première lettre. Apparemment, les sujets ne sont pas capables de focaliser leur attention sur une lettre, un phonème ou une syllabe. Cela ne prouve pas que ces objets ne sont pas des unités de traitement utilisées par le système perceptif.

Une alternative à l'explication en terme de distraction, est que les sujets utiliseraient spontanément une stratégie holistique (i.e. considèreraient le stimulus global) pour répondre : dans les blocs contrôles, il y a deux objets à discriminer, dans les blocs expérimentaux, il y en a quatre, ce qui expliquerait la différence de temps de réaction. Les deux types d'explication suggèrent que la découverte d'une différence de temps de réaction moyens entre des blocs contenant quatre stimuli et des blocs en contenant deux ne permet pas de conclure, ni sur le traitement ni sur la représentation utilisée par les sujets pour répondre.[20] Cependant, l'explication en terme de distraction soulève la possibilité que différents types de variations puissent produire des interférences plus ou moins grandes. C'est cette idée qui a présidé à la réalisation de l'expérience 3.

Si les sujets qui classifient des phonèmes utilisent une représentation qui est une simple concaténation de phonèmes, alors on s'attend à ce que l'interférence due à une variation soit indépendante de la structure syllabique des stimuli. Si, au contraire, cette représentation est structurée syllabiquement, c'est à dire, si les phonèmes appartenant à la même syllabe sont plus « liés » que les phonèmes appartenant à des syllabes différentes, alors on s'attend à ce que les variations intra-syllabiques « coûtent » plus que les variations extra-syllabiques. Dans l'expérience 3, les sujets devaient classifier une voyelle quand variait une consonne qui se trouvait, soit dans la même syllabe, soit dans une autre syllabe, que la voyelle à classer : l'interférence était maximale quand la variation était à l'inté-

19. Après avoir réalisé ces expériences, nous avons découvert que Shand (1976) avait également étudié le problème des interférences dues aux variations « hétéro-syllabiques ». Il avait observé un effet d'interférence en classification de phonème dû à la variation d'un syllabe adjacente. Cependant, dans son étude, les sujets devaient classifier un phonème (et non pas une syllabe), et de plus celui-ci se trouvait *après* la syllabe qui variait. Cet arrangement augmentait au maximum les chances d'observer de l'interférence. Au contraire, les expériences 1 et 2 plaçaient la syllabe dans une situation similaire à celle du phonème dans l'étude de Wood et Day (1975).

20. L'absence d'interférence, par contre, est nettement plus intéressante, car elle suggère une indépendance entre la dimension classée et celle qui varie.

rieur de la syllabe.

Il faut conclure que les phonèmes appartenant à la même syllabe sont plus « liés » que des phonèmes appartenant à des syllabes différentes. Ce résultat est important, car il montre la sensibilité des sujets à la structure syllabique dans une tâche qui ne requiert pas la manipulation explicite de syllabes, mais il faut souligner qu'il ne permet pas de déduire qui, de la syllabe, ou du phonème est récupéré en premier par le système perceptif. Pour des raisons inhérentes à la tâche, le système de décision pourrait accéder préférentiellement à une représentation syllabifiée du signal, bien que celle-ci ne soit pas la première extraite du signal. D'ailleurs, en employant un paradigme expérimental très différent de celui de Garner, Pitt et Samuel (1990) ont affirmé que les sujets pouvaient focaliser leur attention précisement sur un phonème, et ont conclu que celui-ci, plutôt que la syllabe, était l'unité de perception. Dans le chapitre suivant, nous allons examiner en détail leur proposition.

Chapitre III

Focalisation Attentionnelle dans la Structure Syllabique

Nous venons de voir que les sujets ne pouvaient apparemment pas focaliser leur attention sur le phonème ou la syllabe initiale de stimuli auditifs. Toutefois, cela pourrait être un artefact dû au paradigme expérimental où les stimuli sont peu nombreux et connus à l'avance par les sujets, ce qui pourrait encourager une stratégie de comparaison globale. Si les listes contiennent de nombreux items que les sujets ne connaissent pas à l'avance, alors cette stratégie ne peut plus s'appliquer. Cela est peut-être l'origine du contraste entre la conclusion de Wood et Day (1975) et celle de Pitt et Samuel (1990) : ces derniers ont argumenté que les sujets *pouvaient focaliser leur attention sur un phonème précis*, et ceci, en employant un paradigme où les stimuli sont très variés.

Pitt et Samuel (1990) se sont inspirés du paradigme de détection avec biais attentionnel inventé par Posner et Snyder (1975). Dans ce paradigme, la cible à détecter peut apparaître dans différentes positions, l'une d'entre elles étant plus probable que les autres. Le résultat typique est que le sujet est plus rapide quand la cible apparaît dans la position la plus probable, ce qui est généralement interprété comme un effet de préparation attentionnel. Pitt et Samuel ont adapté ce paradigme à la tâche de détection de phonème généralisée : leurs sujets devaient détecter des phonèmes qui pouvaient apparaître dans l'une des quatre positions consonantiques de mots ayant une structure C_1VC_2–C_3VC_4. Pour un premier groupe de sujets, la cible apparaissait le plus souvent en C_1 ; pour un second, en C_2 ; pour un troisième, en C_3 ; et pour un quatrième, en C_4. Pitt et Samuel ont comparé les temps de détection de chaque groupe dans chacune des positions. Les résultats sont clairs : chaque groupe est plus rapide (et fait moins d'erreurs) précisément sur la position favorisée par sa liste.

Selon Pitt et Samuel, ce résultat montre que les sujets peuvent focaliser leur attention sur un phonème précis, et donc que le phonème est l'« unité de perception » de la parole. Si celle-ci était la syllabe, raisonnent-ils, on devrait observer un avantage attentionnel étendu à tous les phonèmes de la même syllabe. Par exemple, si les sujets sont habitués à détecter la cible en C_2, ils devraient également être rapides pour la détecter en C_1, puisqu'elle est dans la même syllabe.[1] Ce n'est pas ce que Pitt et Samuel observent et, par conséquent, ils concluent que l'unité perceptive est le phonème. C'est toutefois une conclusion un peu hâtive. Les temps de décision sont très lents (environ 800 msec, ce qui est similaire à ce qu'on observe typiquement en décision lexicale[2]). Les sujets pourraient fort bien avoir avoir identifié la syllabe dans un premier temps, puis l'avoir ensuite décomposée en phonèmes.

Le but principal de Pitt et Samuel était de déterminer la *taille* du faisceau attentionnel. Pour notre part, nous nous demandons plutôt *quelle est la propriété sur laquelle se focalise l'attention*. Si le signal est perçu purement comme une séquence de phonèmes, ainsi que le suggèrent Pitt & Samuel, leur résultat s'interprète comme montrant que le premier groupe porte son attention sur le premier phonème de chaque stimuli ; le second groupe, sur le troisième phonème (C_2), et ainsi de suite. C'est la propriété de « n^e^ phonème » qui serait donc pertinente.

Une interprétation alternative est que les sujets s'habituent à détecter un phonème dans une position définie *syllabiquement*. Ainsi, le premier groupe porterait son attention sur l'attaque de la première syllabe, le second groupe sur le coda de la première syllabe, le troisième sur l'attaque de la deuxième syllabe et le quatrième sur le coda de la deuxième syllabe. Notons, que cela ne signifie pas nécessairement que la syllabe doive être reconnue avant les phonèmes. Au moment où la réponse est effectuée, les deux types d'unités peuvent fort bien avoir été identifiées par le système perceptif. Ce modèle propose que la représentation utilisée par les sujets pour effectuer leur réponse est plus qu'une simple chaîne linéaire de phonème : elle est structurée syllabiquement. Pour différencier ce modèle d'un simple modèle phonémique séquentiel, il faut décorréler les deux types de positions et déterminer si les sujets peuvent focaliser leur attention sur une position syllabique indépendamment de la position séquentielle.

Les expériences qui suivent ont pour but de déterminer si les sujets peuvent focaliser leur attention sur une position précise dans la structure syllabique des mots. Deux positions syllabiques ont été considérées : le coda de la première syllabe de mots bisyllabiques (caP–tif) et l'attaque de la seconde syllabe (ca–Price). Dans les deux cas, le phonème cible est en troisième position séquentielle. Deux listes sont construites où l'une et l'autre des deux positions syllabiques sont, res-

1. Ou alors l'avantage attentionnel pourrait s'étendre aux phonèmes placés dans des syllabes différentes mais dans une position syllabique interne identique, attaque ou coda.

2. Nous ne voulons pas suggérer que ces résultats sont post-lexicaux car Pitt & Samuel l'observent également sur des pseudo-mots. Mais à des temps aussi « lents » de nombreux processus post-perceptuels peuvent jouer.

pectivement, favorisées ; c'est à dire que dans une liste, la cible est plus souvent placée dans le coda de la première syllabe, alors que dans l'autre, elle est plus souvent dans l'attaque de la seconde syllabe. C'est seulement si les sujets ont accès à une représentation syllabifiée qu'ils pourront tirer avantage de ce biais. Si, pour effectuer leur décision, ils n'ont accés qu'à une représentation linéaire, alors on ne s'attend pas à observer d'effet de liste. Néanmoins, il se pourrait que les deux types de représentations (linéaire et syllabique) coexistent et jouent un rôle simultanément ; l'attention pourrait alors être focalisée sur une position définie à la fois séquentiellement et syllabiquement. Pour évaluer cette troisième possibilité, on examine la généralisation sur la quatrième position séquentielle après un entraînement sur la troisième position.

EXPÉRIENCE 4: INDUCTION SYLLABIQUE

La méthode employée suit précisément celle de Pitt et Samuel : la tâche, une décision, les temps de présentation des stimuli et les intervalles inter-stimuli, la consigne, sont aussi identiques que possible à ceux décrits dans leur article. Toutefois, en raison des contraintes lexicales du français, il a fallu restreindre nettement le nombre d'essais : les listes expérimentales possèdent 114 stimuli au lieu de 480 dans l'expérience originale. Cela entraîne le risque que les sujets n'aient pas le temps de découvrir la régularité dans la position de la cible.

DESCRIPTION

Matériel : Tous les mots (164 au total) étaient bisyllabiques et possédaient l'une des quatre structures CVC–CV, CV–CCV, CCVC–CV, ou CCV–CCV (en négligeant d'éventuelles consonnes finales). Le phonème cible était systématiquement la consonne qui suivait la première voyelle et pouvait donc se trouver soit dans la première syllabe, soit dans la seconde. Tous les stimuli possédaient un groupe de consonnes central. En effet, une suite VCV étant nécessairement syllabifiée comme V–CV en français[3], un coda d'une syllabe non finale se trouve obligatoirement dans un groupe de consonnes. Il a été décidé de conserver cette contrainte pour les mots possédant une cible en attaque de seconde syllabe afin que cette propriété (présence d'un groupe consonantique central, ou non) ne permette pas de distinguer les deux types de mots.

Deux listes de 114 mots ont été construites de la façon suivante :

– 50 mots comprenaient un phonème cible en troisième position séquentielle. Dans la première liste, ce phonème apparaissait toujours en *coda* de la première syllabe

3. Cela est une conséquence du « principe de l'attaque maximale », et du fait qu'en français toute consonne peut apparaître en début de mot: par conséquent, une consonne intervocalique appartient nécessairement à la syllabe ayant pour noyau la seconde voyelle.

(ex: caP.ture). Dans la seconde liste, il apparaissait dans l'*attaque* de la seconde syllabe (ex: ca.Price). Ces mots, différents pour les deux listes, avaient pour but d'attirer l'attention des sujets sur une position structurale particulière. Pour cette raison, on les appellera mots « inducteurs ».

– 32 mots « distracteurs » ne comprenaient pas la cible. Ils servent à empêcher que le sujet n'anticipe systématiquement la présence de la cible, stratégie évidente à adopter si tous les mots avaient possédé la cible. Les distracteurs sont les mêmes dans les deux listes.
– 32 mots « tests », partagés par les deux listes, appartenaient à l'une des quatre catégories obtenues par le croisement des deux facteurs : position séquentielle (troisième ou quatrième) et position syllabique (coda de la première syllabe ou attaque de la seconde syllabe). 8 mots de chaque catégorie appartenaient aux mots tests (cf table III.1).

Position de la cible			
3ième		4ième	
coda	attaque	coda	attaque
caP–tif	cy–Clone	staG–ner	prê–Tresse
doC–teur	ca–Price	fluC–tue	tri–Pler
seG–ment	no–Blesse	fraG–ment	cri–Bler
symP–tôme	di–Plôme	stiG–mate	fla–Grant
faC–teur	mi–Graine	traC–teur	pro–Blème
suB–til	su–Blime	cryP–ter	pla–Trer
reP–tile	nom–Breux	planC–ton	trem–Bler
leC–ture	pa–Trie	fraC–ture	pro–Grès

TAB. III.1 – *Matériel d'induction syllabique : Mots tests (le phonème cible est en majuscule)*

Les deux listes ne différaient que par les mots inducteurs. L'ordre des mots, identique dans les deux listes, a été tiré aléatoirement, avec pour seules contraintes que (a) chaque mot test soit précédé d'un ou deux mots inducteurs et (b) que le phonème cible soit toujours différent dans deux essais successifs. Dans chacune des listes, la cible apparaissait dans 80 % des cas (50+8+8 sur 82 positifs) dans la même position syllabique (par ex. coda), et dans 20 % des cas dans l'autre position syllabique (attaque). Dans les deux listes, elle apparaissait en troisième position séquentielle dans 80 % des cas, et en quatrième position dans les 20 % restants. Enfin, il y avait 72 % d'essais positifs, et 28 % d'essais négatifs.

L'identité des phonèmes cibles a été variée autant que possible, dans les inducteurs comme dans les mots tests. Dans ces derniers, les cibles étaient toutes des occlusives (p,t,k,b,d,g) pour lesquelles on peut définir de façon assez précise un début (conventionnellement le début est placé à l'explosion). Dans les inducteurs, d'autres phonèmes ont été employés, en particulier des liquides pour les inducteurs de type 'coda' (ex: carton), et des fricatives pour les inducteurs de type 'attaque' (ex: névrose).

Les stimuli, lus par une voix féminine, ont été enregistrés digitalement (fréquence d'échantillonnage : 16Khz) sur le disque dur d'un PDP 11/73. A l'aide d'un éditeur de parole (EDISON), chaque stimulus a été extrait et une marque a été placée sur le début de l'explosion des cibles occlusives pour les 32 mots tests. Deux cassettes audio ont ensuite été enregistrées, avec les stimuli sur la piste 1, les tops de synchronisations des essais et les marques des cibles, sur la piste 2.

Procédure : Durant toute l'expérience, le sujet est assis devant un ordinateur et porte un casque haute-fidélité dans lequel sont présentés binauralement les stimuli. Devant lui, deux boutons de réponses (clés morses) reliés à l'ordinateur permettent de mesurer ses temps de réaction avec une précision inférieure à la milliseconde. L'expérience est une succession d'essais qui durent chacun 5 secondes. Au début de chaque essai, un phonème est présenté visuellement, au centre de l'écran de l'ordinateur pendant 1 seconde ; l'écran s'efface ensuite pendant une autre seconde au terme de laquelle un mot est présenté dans les écouteurs ; le sujet doit alors « appuyer aussi vite que possible sur le bouton OUI si le mot contient la cible, sur le bouton NON si le mot ne la contient pas » (au début de l'expérience, le sujet a pu choisir le côté correspondant à OUI, l'autre étant automatiquement assigné à NON).

En expliquant les instructions, l'expérimentateur n'employait pas le terme « phonème » mais demandait au sujet de « se représenter AUDITIVEMENT le son correspondant à la lettre présentée à l'écran » et lui fournissait des exemples : « K comme dans 'quand' ou 'clé'; R comme dans 'France' ou 'montre' ». Par ailleurs, les sujets n'étaient pas informés du fait que, dans la liste, la cible apparaîtrait plus souvent dans une position plutôt qu'une autre. Insistons sur le fait qu'à la différence des expériences « classiques » de détection de phonème, la cible était un phonème différent à chaque essai. D'autre part, la tâche est ici une décision plutôt qu'une simple détection, puisque le sujet doit également donner une réponse quand le mot ne contient pas la cible.

Sujets : Vingt sujets volontaires, étudiants dans diverses universités parisiennes, ont participé à l'expérience. Ils étaient tous de langue maternelle française. Aucun ne connaissait le but de l'expérience. Testés individuellement, ils ont été assigné en fonction de leur ordre d'arrivée à l'une des deux listes expérimentales. Trois sujets supplémentaires ont été remplacés pour avoir commis plus de 15 % de fausses alarmes (c'est à dire plus de 5 réponses oui sur les 32 distracteurs).

RÉSULTATS

Le taux global de fausses alarmes était de 5.3 %. Les omissions (3 % des données) et les temps de réactions inférieurs à 100 ms ou supérieurs à 1500 ms (1.5 % des données) ont été remplacés, pour chaque sujet, par son temps de réaction moyen sur les items dans la même condition. La table III.2 présente les temps de réaction moyens et les taux d'erreur des deux groupes de sujets sur les quatre types de mots tests. La figure III.1 présente graphiquement les temps de réaction.

	Position de la cible			
Groupe	3ième		4ième	
d'induction	coda	attaque	coda	attaque
coda	473 (5%)	506 (3%)	489 (3%)	525 (5%)
attaque	515 (6%)	457 (1%)	518 (4%)	490 (5%)

Chaque cellule est la moyenne de 80 mesures. L'écart-type entre sujets des temps de réaction moyens est de 99 msec ; les moyennes des écarts-types intra-condition sont de 125 msec par sujet et de 53 msec par items.

TAB. III.2 – *Temps de décision moyens en ms et erreurs en %.*

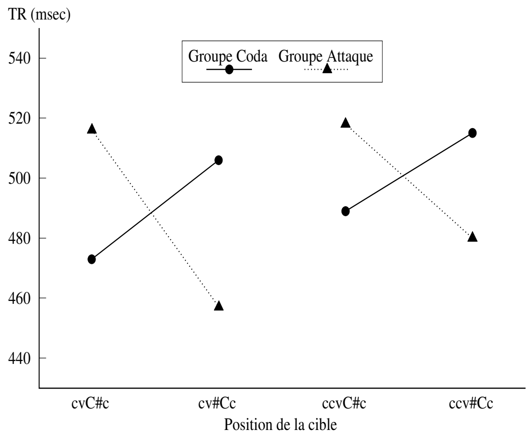


FIG. III.1 – *Détection de phonème généralisée avec biais syllabique*

Quatre analyses de variance, avec les sujets puis les items déclarés comme facteurs aléatoires, sur les temps de réaction et les erreurs, ont été effectuées. Trois facteurs binaires étaient déclarés : position séquentielle (P) de la cible (3 ou 4), position syllabique (S) (coda ou attaque), groupe d'induction (G) (coda ou attaque, dépendant de la liste expérimentale).

```
Analyse par sujets: PLAN  S10<G2>*S2*P2
G = Groupe d'Induction
S = Position Syllabique (Coda vs Attaque)
```

```
P = Position Phonèmique (3ième vs 4ième)
Temps de réaction:
P               F(1,18)=  4.22    MSE=1483.5      p=0.0548
S               F(1,18)=  0.09    MSE=4057.54     p=0.2324
S.P             F(1,18)=  1.10    MSE=1248.51     p=0.3081
G               F(1,18)=  0.00    MSE=41148.3     p=1.0000
G.P             F(1,18)=  0.00    MSE=1483.5      p=1.0000
G.S             F(1,18)=  7.54    MSE=4057.54     p=0.0133
G.S.P           F(1,18)=  0.68    MSE=1248.51     p=0.4204
G.S/P1          F(1,18)=  6.83    MSE=3046.09     p=0.0176
G.S/P2          F(1,18)=  4.71    MSE=2259.96     p=0.0436
S/G1            F(1,9)=   2.21    MSE=5496.56     p=0.1713
S/G2            F(1,9)=   7.19    MSE=2618.52     p=0.0251
Erreurs:
P               F(1,18)=  0.17    MSE=0.3028     p=0.3150
S               F(1,18)=  0.69    MSE=0.2917     p=0.4170
S.P             F(1,18)=  3.10    MSE=0.2583     p=0.0953
G               F(1,18)=  0.00    MSE=0.2806     p=1.0000
G.P             F(1,18)=  0.00    MSE=0.3028     p=1.0000
G.S             F(1,18)=  0.17    MSE=0.2917     p=0.3150
G.S.P           F(1,18)=  0.19    MSE=0.2583     p=0.3319
G.S/P1          F(1,18)=  0.36    MSE=0.2778     p=0.4440
G.S/P2          F(1,18)=  0.00    MSE=0.2722     p=1.0000
S/G1            F(1,9)=   0.06    MSE=0.4139     p=0.1880
S/G2            F(1,9)=   1.33    MSE=0.1694     p=0.2785
Analyse par items: PLAN  S8<S2*P2>*G2
Temps de réaction:
P               F(1,28)=  0.50    MSE=7385.04     p=0.4853
S               F(1,28)=  0.26    MSE=7385.04     p=0.3859
S.P             F(1,28)=  0.16    MSE=7385.04     p=0.3078
G               F(1,28)=  0.01    MSE=3479.71     p=0.0789
G.P             F(1,28)=  0.14    MSE=3479.71     p=0.2889
G.S             F(1,28)=  7.55    MSE=3479.71     p=0.0104
G.S.P           F(1,28)=  0.79    MSE=3479.71     p=0.3817
G.S/P1          F(1,14)=  5.19    MSE=4426.16     p=0.0389
G.S/P2          F(1,14)=  2.38    MSE=2533.27     p=0.1452
S/G1            F(1,28)=  1.24    MSE=5644.49     p=0.2749
S/G2            F(1,28)=  4.05    MSE=5220.27     p=0.0539
Erreurs
P               F(1,28)=  0.08    MSE=0.8259     p=0.2206
S               F(1,28)=  0.30    MSE=0.8259     p=0.4118
S.P             F(1,28)=  1.21    MSE=0.8259     p=0.2807
G               F(1,28)=  0.00    MSE=0.4598     p=1.0000
```

G.P	F(1,28)=	0.00	MSE=0.4598	p=1.0000
G.S	F(1,28)=	0.14	MSE=0.4598	p=0.2889
G.S.P	F(1,28)=	0.14	MSE=0.4598	p=0.2889
G.S/P1	F(1,14)=	0.45	MSE=0.2768	p=0.4867
G.S/P2	F(1,14)=	0.00	MSE=0.6429	p=1.0000
S/G1	F(1,28)=	0.06	MSE=0.5491	p=0.1917
S/G2	F(1,28)=	0.38	MSE=0.7366	p=0.4574

Aucun effet principal ou interaction n'est significatif dans les analyses des erreurs. Dans les analyses canoniques des temps de réaction, seule l'interaction entre groupe d'induction et position syllabique est significative. La position séquentielle produit un effet de 18 msec qui atteint juste le seuil de significativité (p=.05) mais seulement dans l'analyse par sujets. Dans des comparaisons restreintes à chaque groupe, il apparaît que l'effet de position syllabique n'est significatif que pour le groupe induit en Attaque (G2). Dans des comparaisons restreintes à chaque position séquentielle, l'interaction entre Induction et Position Syllabique, restreinte à la position phonémique « 3 », est de 45 msec (significatif) ; celle restreinte à la position « 4 » est de 32 msec (non significatif). Toutefois il n'y a pas de triple interaction et la non-significativité de l'interaction sur la quatrième position reflète probablement l'assez faible puissance statistique.

Discussion

Comme l'atteste l'interaction obtenue entre induction et position syllabique dans l'analyse des temps de réaction, les deux groupes de sujets ont eu des comportements différents : chaque groupe était avantagé quand les cibles apparaissaient dans la position syllabique la plus probable de sa liste. Il est vrai que cet avantage n'est significatif que pour le groupe habitué en attaque, mais cette asymétrie pourrait être le résultat d'une différence de difficulté entre les stimuli coda et attaque. Seul un groupe contrôle, neutre sur le plan de l'induction syllabique, permettrait de déterminer si il existe une telle différence entre les deux types de stimuli. Toutefois, l'interaction est la mesure la plus pertinente : elle montre que les sujets ont utilisé la structure syllabique des stimuli et ont optimisé leur comportement en fonction de celle-ci. Pour que cela soit le cas, il faut qu'ils aient eu accès à une telle information syllabique.

Les sujets recevaient 80 % des cibles en troisième position séquentielle mais il n'y a qu'un faible effet de position séquentielle (18 msec, significatif seulement par sujets). On ne peut donc pas exclure l'éventualité que les sujets aient subit une habituation séquentielle en plus d'une habituation syllabique. Néanmoins, il faut souligner que l'induction séquentielle, si elle est réelle, ne suffit pas à faire disparaître l'induction syllabique. Les expériences suivantes fourniront l'occasion de réexaminer la réalité de l'induction séquentielle.

Une autre question est celle de l'origine de l'effet syllabique observé. Il est assez couramment admis que la détection de phonème peut utiliser deux sources d'information : d'une part une représentation pré-lexicale, d'autre part une représentation post-lexicale, c'est à dire récupérée après l'identification du mot (Foss & Blank, 1980; Cutler & Fodor, 1979; Cutler et al., 1987; Eimas, Hornstein, & Payton, 1990). On peut donc se demander si l'information syllabique qui influence les sujets dans notre expérience est d'origine pré-lexicale ou post-lexicale. La considération des temps de réaction moyens offre un premier ordre d'idées : ceux-ci sont de l'ordre de 500 ms, et ils sont mesurés, rappelons-le, à partir du début du phonème cible, qui se trouve en moyenne à 300 ms à l'intérieur du mot. Au moment où il effectue sa réponse, le sujet est donc 800 msec après le début du stimulus. Ces temps sont très comparables aux 770 msec obtenus en décision lexicale auditive, avec des sujets français et des stimuli bisyllabiques par Dupoux (1989, exp.6). Cela n'est pas en soi la preuve définitive que les sujets ont utilisé des informations post-lexicales, puisque il y a des cas ou malgré la lenteur des réponses, les sujets ne sont pas influencés par le lexique (Eimas et al., 1990).

Les études les plus comparables à la nôtre sont celles de Frauenfelder et Segui (1989) et Frauenfelder et al. (1990) car toutes deux utilisaient la détection de phonème généralisée (mais sans biais attentionnel). Frauenfelder et Segui (1989) observent un effet *d'amorçage sémantique* entre des mots successifs. Cela suggère que les sujets utilisent l'information lexicale pour répondre ; toutefois cela pourrait être le résultat d'une stratégie propre à la situation d'amorçage sémantique. Frauenfelder et al. (1990) comparent la détection de phonèmes dans des mots et des non-mots ; la détection est plus rapide dans les mots, mais seulement quand le phonème cible se trouve après le point d'unicité du mot.[4]. Dans le cas de nos stimuli, l'examen de leur point d'unicité (grâce à *BRULEX* ; cf Content 1990), révèle que celui-ci est toujours situé *après* le phonème cible.

L'une des variables connues pour affecter le plus le temps d'accès au lexique

4. Le point d'unicité est le phonème à partir duquel le mot se distingue de tous les autres dans le lexique. Si la recherche lexicale se fait selon un arbre de décision où chaque noeud possédent une branche par phonème alors un mot peut être identifié dès que son « point d'unicité » est atteint. Marlsen-Wilson a proposé un tel modèle pour la reconnaissance des mots auditifs ; un des arguments expérimentaux qu'il avance est la corrélation quasi parfaite entre les temps de décision lexicale et la position du point d'unicité qu'il obtient dans Marlsen-Wilson (1984) (voir cependant Goodman & Huttenlocher, 1988 et Taft & Hambly, 1986).

Remarque méthodologique : il ne me semble pas évident que l'étude de Frauenfelder et al. (1990) démontre parfaitement un effet de point d'unicité. Ceux-ci comparent des conditions « avant point d'unicité » et « après point d'unicité ». Mais les stimuli de ces deux catégories ne sont apparemment pas contrôlés pour la longueur des mots. Dans l'exemple qu'ils fournissent (avant PU : /oPeratie/, apres /olymPiade/), la cible (/P/) est plus proche de la fin du stimulus dans la condition «après PU». Il faudrait comparer l'effet lexical en détection de phonème avant et après le point d'unicité dans des mots *de même longueur* à point d'unicité précoce vs tardif.

étant la fréquence d'occurrence des mots (les mots fréquents sont reconnus plus rapidement que les mots peu fréquents), nous avons examiné le coefficient de corrélation entre le temps de réponse moyen et la fréquence lexicale de chaque item (les logarithmes plutôt que les fréquences brutes ont en fait été utilisés). Cette analyse révèle qu'il n'y a pas de corrélation globale entre temps de réaction et fréquence lexicale (r=-0.22, F(1,30)=1.6, p=.22). Toutefois, des analyses restreintes, présentées dans la table III.3, montrent qu'il y a une accélération significative due à la fréquence lexicale pour le groupe « coda » quand la cible apparaît en position « attaque ».

	Cible	
Induction	Coda	Attaque
Coda	-0.30 (p=0.25)	-0.51 (p=0.04)
Attaque	0.18 (p=0.5)	-0.18 (p=0.5)

TAB. III.3 – *Coefficients de Corrélation (et significativité) entre Temps de Réaction et Fréquence Lexicale*

On pourrait être tenté d'interpréter cela en supposant que la « surprise » entraîne l'utilisation d'une stratégie lexicale. Ces données sont toutefois trop succinctes pour permettre une conclusion catégorique ; de plus, on n'explique pas l'absence d'effet pour le groupe « Attaque ». Toutefois, la corrélation obtenue dans un des cas montre la légitimité de la question de l'origine des réponses. L'expérience qui suit a pour but d'éclairer ce problème de la nature lexicale ou pré-lexicale des réponses.

EXPÉRIENCE 4BIS : INDUCTION SYLLABIQUE ACCÉLÉRÉE

Afin de diminuer la possibilité que les sujets emploient une stratégie lexicale, nous avons décidé de refaire la même expérience en modifiant la tâche des sujets. Ceux-ci doivent maintenant effectuer une véritable détection plutôt qu'une décision : ils ne disposent plus que d'un unique bouton de réponse pour signaler qu'ils ont entendu le phonème cible ; si la cible n'apparaît pas, ils ne doivent rien faire. D'autre part, la consigne a été modifiée en portant l'emphase sur la vitesse : on demande aux sujets « d'appuyer sur le bouton, aussitôt qu'ils entendent le phonème cible et de ne pas attendre la fin du mot ». On espère ainsi diminuer les temps de réaction et, par là même, les possibilités d'emploi de stratégies lexicales.

Plusieurs études montrent que de tels changements de consigne peuvent affecter les résultats en faisant apparaître ou disparaître des effets (p.ex. Dupoux, 1989 et Sebastian-Gallés, Dupoux, Segui, & Mehler, 1992). Ainsi, Cutler, Norris, et Williams (1987) ont argumenté que c'est l'emploi d'une tâche de décision (choix « oui-non »), plutôt que d'une tâche de détection (paradigme « go-no ») qui était responsable des effets lexicaux observés par Taft et Hambly (1985) en détection de syllabe.

DESCRIPTION

Procédure : Cette expérience se déroulait dans des conditions identiques à la précédente, à l'exception des modifications citées en introduction : les sujets disposaient d'un unique bouton (clé morse) pour répondre, et la consigne, qui apparaissait cette fois sur l'écran de l'ordinateur au début de la passation, insistait plus particulièrement sur la rapidité des réponses (on précisait aux sujets que leurs temps de réaction étaient mesurés et qu'ils ne devaient pas spécialement attendre la fin du stimulus pour répondre).

Signalons que les moyens techniques étaient différents : plutôt que d'utiliser des cassettes audio, les stimuli (identiques à ceux de l'expérience précédente), étaient directement restitués (à 16 Khz) à partir du disque dur du PC chargé de présenter les phonèmes cibles, et de recueillir les temps de réaction.

Sujets : Quarante sujets, étudiants de diverses universités parisiennes ou employés de l'Ecole de Hautes Etudes, ont participé à cette expérience. Six sujets supplémentaires ont été remplacés pour avoir dépassé un seuil de fausses alarmes fixé à 15 % .

RÉSULTATS

Le taux global de fausses-alarmes était de 6.3 %. Les mêmes critères que dans l'expérience précédente ont conduit au rejet de 2.3 % des données brutes.

	Position de la cible			
Groupe	3ième		4ième	
d'induction	coda	attaque	coda	attaque
coda	323 (4%)	346 (4%)	341 (4%)	362 (3%)
attaque	382 (6%)	339 (1%)	358 (4%)	349 (3%)

Chaque cellule est la moyenne de 160 mesures. L'écart-type des temps de réaction moyens entre sujets est de 77 msec ; les moyennes des écarts-types intra-condition sont de 108 msec par sujets et de 49 msec par items.

TAB. III.4 – *Temps de décision moyens en msec et erreurs en %*

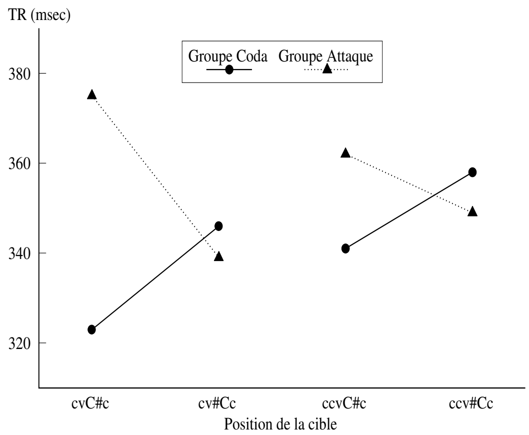


FIG. III.2 – *Détection de phonème généralisée avec biais syllabique*

Des Anovas identiques à celles de l'expérience précédente ont été menées (cf plus bas). Le seul effet significatif à la fois dans l'analyse par items et dans l'analyse par sujets (des temps de réaction) provient de l'interaction entre Groupe d'Induction et Position Syllabique. L'effet de position phonémique est de 5 msec (non significatif). Les analyses restreintes par position phonémique montrent un effet significatif par sujets et marginal par items en position « 3 », et marginal dans les deux cas en position « 4 ». Dans les analyses des erreurs, il y a une triple interaction marginale, dont les analyses restreintes aux positions séquentielles « 3 » et « 4 » révèlent qu'elle provient d'une interaction Induction × Syllabe (dans le sens prédit par l'induction) sur la position 3 mais pas sur la position 4.

```
Analyse par sujets: PLAN  S20<G2>*S2*P2
G = Groupe d'Induction
S = Position Syllabique (Coda vs Attaque)
P = Position Phonémique (3ième vs 4ième)
Temps de réaction:
P              F(1,38)=  0.28    MSE=3241.02    p=0.4002
S              F(1,38)=  0.06    MSE=4282.79    p=0.1922
S.P            F(1,38)=  1.24    MSE=2026.69    p=0.2725
G              F(1,38)=  0.33    MSE=24220.4    p=0.4310
G.P            F(1,38)=  1.76    MSE=3241.02    p=0.1925
G.S            F(1,38)=  5.41    MSE=4282.79    p=0.0255
G.S.P          F(1,38)=  1.50    MSE=2026.69    p=0.2282
G.S/P1         F(1,38)=  5.02    MSE=4283.19    p=0.0310
G.S/P2         F(1,38)=  2.33    MSE=2026.3     p=0.1352
S/G1           F(1,19)=  1.87    MSE=5013.2     p=0.1874
```

S/G2	F(1,19)=	3.95	MSE=3552.39	p=0.0615
Erreurs :				
P	F(1,38)=	0.00	MSE=0.2763	p=1.0000
S	F(1,38)=	4.63	MSE=0.2645	p=0.0378
S.P	F(1,38)=	0.07	MSE=0.3487	p=0.2072
G	F(1,38)=	0.00	MSE=0.1947	p=1.0000
G.P	F(1,38)=	0.00	MSE=0.2763	p=1.0000
G.S	F(1,38)=	0.85	MSE=0.2645	p=0.3624
G.S.P	F(1,38)=	3.51	MSE=0.3487	p=0.0687
G.S/P1	F(1,38)=	4.34	MSE=0.2882	p=0.0440
G.S/P2	F(1,38)=	0.62	MSE=0.325	p=0.4359
S/G1	F(1,19)=	0.88	MSE=0.2263	p=0.3600
S/G2	F(1,19)=	4.13	MSE=0.3026	p=0.0563
Analyse par items : PLAN S8<S2*P2>*G2				
Temps de réaction :				
P	F(1,28)=	0.02	MSE=4801.03	p=0.1115
S	F(1,28)=	0.04	MSE=4801.03	p=0.1571
S.P	F(1,28)=	0.21	MSE=4801.03	p=0.3497
G	F(1,28)=	1.82	MSE=1937.9	p=0.1881
G.P	F(1,28)=	1.10	MSE=1937.9	p=0.3032
G.S	F(1,28)=	5.37	MSE=1937.9	p=0.0280
G.S.P	F(1,28)=	0.66	MSE=1937.9	p=0.4234
G.S/P1	F(1,14)=	2.85	MSE=3325.36	p=0.1135
G.S/P2	F(1,14)=	3.99	MSE=550.435	p=0.0656
S/G1	F(1,28)=	1.41	MSE=2713.29	p=0.2450
S/G2	F(1,28)=	1.69	MSE=4025.64	p=0.2042
Erreurs :				
P	F(1,28)=	0.00	MSE=2.0223	p=1.0000
S	F(1,28)=	1.51	MSE=2.0223	p=0.2294
S.P	F(1,28)=	0.03	MSE=2.0223	p=0.1363
G	F(1,28)=	0.00	MSE=0.7991	p=1.0000
G.P	F(1,28)=	0.00	MSE=0.7991	p=1.0000
G.S	F(1,28)=	0.70	MSE=0.7991	p=0.4099
G.S.P	F(1,28)=	3.83	MSE=0.7991	p=0.0604
G.S/P1	F(1,14)=	3.40	MSE=0.9196	p=0.0865
G.S/P2	F(1,14)=	0.74	MSE=0.6786	p=0.4042
S/G1	F(1,28)=	0.51	MSE=0.9732	p=0.4811
S/G2	F(1,28)=	1.69	MSE=1.8482	p=0.2042

DISCUSSION

Tout d'abord, les changements de tâche et de consigne ont eu l'effet escompté de diminuer le temps de réaction moyen : celui-ci est passé d'environ 500 ms dans la première expérience à 350 ms dans celle-ci.[5] Malgré cela, les résultats demeurent qualitativement les mêmes que dans l'expérience 1 : les sujets sont plus rapides quand les cibles apparaissent dans la position syllabique favorisée par leur liste, comme l'atteste l'interaction entre induction et structure syllabique. D'autre part, il n'y pas de trace d'effets de la position séquentielle. L'expérience 4 est donc répliquée, et ce, dans des conditions favorisant moins l'emploi de stratégies lexicales. Les temps de réactions sont particulièrement rapides: ils sont, par exemple, plus de deux fois inférieurs à ceux obtenus par Pitt & Samuel (1990) (Nos taux d'erreurs sont également nettement plus faibles).

Dans des conditions assez semblables aux nôtres (sujets et stimuli français), Frauenfelder et Segui (1989, exp.2) avaient obtenus un effet lexical – d'amorçage sémantique – avec la détection de phonème généralisée. Mais leur temps de réaction moyen (de l'ordre de l'ordre de 450 msec pour les cibles placées en milieu de mot), est plus comparable à celui de l'expérience 4 qu'à celui de l'expérience 4bis. Dans une autre étude, Frauenfelder et al. (1990) trouvaient un effet de lexcialité en détection de phonème généralisée, mais seulement au-delà du point d'unicité. Or, dans nos stimuli, le point d'unicité se trouve toujours après le phonème cible (au milieu ou à la fin du groupe CC central). Les comparaisons avec des détections de phonème initial sont plus délicates mais il n'y a pas, à notre connaissance, de résultat montrant des effets lexicaux (fréquence ou status lexical) avec des temps de détection inférieurs à 400 msec (cf, par exemple, Cutler et al., 1987; Dupoux, 1989, exp 3). Il faut également noter que nous n'avons plus trouvé d'effet de fréquence dans les temps de réaction des sujets (globalement : r=-0,04, F(1,30)=0.05, p=.8 ; et la corrélation restreinte par groupe la plus importante est r=.29, p=.27). Pour conclure sur l'hypothèse d'une origine lexicale de l'information syllabique, signalons (en anticipant sur le prochain chapitre) que N. Sebastian-Gallés a montré, avec des expériences similaires aux nôtres, que les sujets étaient influencés par la structure syllabique quand les stimuli étaient des *non-mots* autant que quand ils

---

5. Pour comparer cette expérience à la précédente, nous avons conduit une analyse de variance supplémentaire en rassemblant ces données et les précédentes, et en déclarant une variable supplémentaire « Expérience ». Celle-ci produit un effet massif (147 msec ; $F_1(1,56)=37$), mais n'interagit avec aucun autre facteur (en particulier il n'interagit pas avec l'« interaction syllabique » : G×S×Exp : $F_1(1,56)< 1$). L'augmentation de puissance statistique qui résulte de l'augmentation du nombre de degrés de liberté améliore la significativité de l'interaction globale Induction × Position Syllabique ($F_1(1,56)=12$; p=.001), ainsi que ses restrictions aux positions phonémiques « 3 » ($F_1(1,56)=10.6$; p=.002) et « 4 » ($F_1(1,56)=6.4$; p=.02). Une telle analyse doit cependant être interprétée avec précaution étant donné que l'hypothèse d'homogénéité des variances n'est vraisemblablement pas remplie.

étaient des mots.

L'expérience qui suit a pour but de déterminer si, en l'absence de régularité syllabique, les sujets peuvent focaliser leur attention sur une position phonémique séquentielle.

Expérience 5: Induction séquentielle

Les deux expériences qui précèdent montrent que les sujets peuvent tirer parti du fait que le phonème cible apparaît plus souvent, soit dans la première, soit dans la seconde syllabe. Il n'y avait pratiquement pas de trace d'induction séquentielle : les sujets n'étaient pas ralentis quand la cible apparaissait en quatrième position bien que la troisième position soit plus probable. Cela signifie-t-il qu'il ne peut y avoir d'induction séquentielle ? Il est possible qu'on puisse observer une induction séquentielle *en l'absence d'induction syllabique*. Nous avons donc construit avec le matériel précédent des listes sans biais syllabique mais avec un fort biais séquentiel : la cible apparaissait en troisième position dans 80 % des cas.

Description

Matériel : Les stimuli des deux expériences précédentes ont été réutilisés. On a construit deux nouvelles listes expérimentales en échangeant alternativement un inducteur sur deux entre les listes précédentes. Les listes ainsi obtenues contenaient chacune 25 inducteurs « coda » et 25 inducteurs « attaque ». Les mots tests et les distracteurs n'ont pas été modifiés. On a fait en sorte que la moitié des mots tests soient précédés d'un inducteur de type « coda » et l'autre moitié, d'un inducteur de type « attaque ». Finalement, il n'y a pas de biais global favorisant un type de structure syllabique ; par contre, 66 mots possèdent la cible en troisième position et 16 seulement, en quatrième position.

Procédure : Elle était exactement identique à celle de l'expérience précédente : la tâche était une détection (paradigme « go–no go »), et les instructions insistaient sur la rapidité des réponses.

Sujets : Vingt sujets volontaires, tous étudiants, ont participé à cette expérience. Dix ont été assignés à chaque liste. Trois sujets supplémentaires ont été remplacés pour avoir fait trop d'erreurs.

Résultats

Les mêmes critères que dans les expériences précédentes ont été appliqués conduisant au remplacement de 3.7 % des données (2.7 % d'ommissions et 1 % de « hors-limites »). Les résultats sont présentés tab. III.5.

Position de la cible			
3ième		4ième	
coda	attaque	coda	attaque
300 (8%)	321 (4%)	319 (3%)	320 (3%)

Chaque cellule est la moyenne de 160 mesures. L'écart-type des temps de réaction moyens entre sujets est de 52 msec ; les moyennes des écarts-types intra-condition sont de 92 msec par sujets et de 37 msec par items.

TAB. III.5 – *Temps de décision moyens en msec et erreurs en %*

Dans les analyses de variance, trois facteurs étaient déclarés: liste expérimentale (entre-sujets), position séquentielle (intra-sujets) et position syllabique (intra-sujets) .

```
Analyse par sujets: PLAN  S10<L2>*S2*P2
L = Liste
S = Position Syllabique (coda ou attaque)
P = Position Phonémique (3 ou 4)
Temps de réaction:
L              F(1,18)=  0.03    MSE=11183.9    p=0.1356
P              F(1,18)=  1.21    MSE=1567.6     p=0.2858
P.L            F(1,18)=  0.19    MSE=1567.6     p=0.3319
S              F(1,18)=  1.02    MSE=2605.29    p=0.3259
S.L            F(1,18)=  0.18    MSE=2605.29    p=0.3236
S.P            F(1,18)=  1.06    MSE=1452.43    p=0.3169
S.P.L          F(1,18)=  0.16    MSE=1452.43    p=0.3061
Erreurs:
L              F(1,18)=  1.67    MSE=0.2694    p=0.2126
P              F(1,18)=  4.46    MSE=0.2806    p=0.0489
P.L            F(1,18)=  0.71    MSE=0.2806    p=0.4105
S              F(1,18)=  3.10    MSE=0.2583    p=0.0953
S.L            F(1,18)=  0.19    MSE=0.2583    p=0.3319
S.P            F(1,18)=  2.66    MSE=0.1694    p=0.1203
S.P.L          F(1,18)=  0.00    MSE=0.1694    p=1.0000
Analyse par items: PLAN  S8<S2*P2>*L2
Temps de réaction:
L              F(1,28)=  0.26    MSE=1262.68    p=0.3859
P              F(1,28)=  0.37    MSE=3689.01    p=0.4521
P.L            F(1,28)=  0.24    MSE=1262.68    p=0.3720
S              F(1,28)=  0.52    MSE=3689.01    p=0.4768
S.L            F(1,28)=  0.37    MSE=1262.68    p=0.4521
```

S.P	F(1,28)=	0.38	MSE=3689.01	p=0.4574
S.P.L	F(1,28)=	0.20	MSE=1262.68	p=0.3418
Erreurs :				
L	F(1,28)=	3.82	MSE=0.1473	p=0.0607
P	F(1,28)=	1.13	MSE=1.3884	p=0.2969
P.L	F(1,28)=	1.70	MSE=0.1473	p=0.2029
S	F(1,28)=	0.72	MSE=1.3884	p=0.4033
S.L	F(1,28)=	0.42	MSE=0.1473	p=0.4778
S.P	F(1,28)=	0.41	MSE=1.3884	p=0.4728
S.P.L	F(1,28)=	0.00	MSE=0.1473	p=1.0000

Le seul effet significatif est celui de Position Phonémique (P) dans l'analyse des erreurs par sujets : les sujets commettent *plus* d'erreurs en troisième qu'en quatrième position. Une inspection de la table III.5 révèle que cela est dû à un effet de compensation entre la précision et le temps de réaction (« speed-accuracy tradeoff »). Ajoutons que les sujets étaient plus rapides de 25.6 msec que ceux dans l'expérience précédente, mais que cette différence n'est pas significative (F(1,58)=1.83 ; p=0.18). Signalons enfin que nous avons également conduit une analyse de variance pour déterminer si la position syllabique de la cible dans l'item précédent (l'« induction locale »), avait un effet sur les temps de réaction. Cette analyse n'a révélé aucun effet significatif.

Discussion

Bien que les sujets aient reçu 80 % des phonèmes cibles en troisième position séquentielle, ils ont détecté ceux qui apparaissaient en quatrième position sans être ralentis. C'était déjà le cas dans l'expérience précédente, mais il y avait alors la « concurrence » de l'induction syllabique. Ici, le statut syllabique des cibles étant varié systématiquement, la seule régularité globale qu'auraient pu détecter les sujets était le biais vers la troisième position. Ils ne l'ont pas fait. Cela prouve que la position séquentielle d'un phonème n'est pas pertinente perceptuellement. Si le phonème était l'unité primaire de perception, on aurait pu s'attendre à ce qu'il soit possible de les « compter ». Cela ne semble pas être le cas, du moins avec cette tâche.

Nous avons interprété les résultats des expériences 4 et 4bis en affirmant que les sujets focalisaient leur attention, soit sur une position coda de la première syllabe, soit sur une position attaque de la seconde syllabe. Mais, il est possible que les sujets aient en fait focalisé leur attention sur la syllabe *entière*. Pour écarter cette possibilité, il aurait fallu mesurer les temps de réaction sur d'autres phonèmes, appartenant à la même syllabe. Par exemple, si les sujets qui détectent une majorité des phonèmes cibles en coda de la première syllabe portent leur attention sur cette syllabe globalement, ils devraient être plus rapides que l'autre groupe pour détecter un phonème au début de celle-ci.

Ce sont les résultats de Pitt et Samuel (1990) qui nous incitent à penser que cela n'est pas le cas : leurs sujets avaient un avantage attentionnel restreint précisément au phonème favorisé par leur liste. Cependant, comme il existe dans la littérature la proposition que les Français pourraient utiliser la syllabes dans des tâches où les Anglais préfèrent employer le phonème (cf Cutler, Mehler, Norris et Segui, 1986 ; Norris et Cutler, 1988 ; et le chapitre IV), nous avons décidé de reproduire une expérience similaire à celle de Pitt et Samuel, mais cette fois-ci, en français.

EXPÉRIENCE 6: INDUCTION INTRA-SYLLABIQUE

Cette expérience est une reproduction de Pitt et Samuel (1990) en français : on compare quatre groupes de sujets habitués à détecter, respectivement, la première, seconde, troisième et quatrième consonne de mots de structure CVC–CVC.

DESCRIPTION

Matériel : On a sélectionné 147 mots français ayant tous une structure C_1VC_2–C_3VC_4, ce qui fournit quatre positions consonantiques potentielles pour la détection de phonème. On a ensuite construit une liste de mots et quatre listes de phonèmes cibles appariées. Ces listes définissent quatre suites de 147 essais « mot — phonème cible ». Il y avait trois catégories d'essais : les « tests » (40 essais), les « inducteurs » (70 essais), et les « distracteurs » (37 essais). Les phonèmes cibles des essais « tests » et « distracteurs » étaient identiques pour les quatre listes. Les « distracteurs » étaient des essais où le phonème cible n'apparaissait pas dans le mot. Les essais « tests » appartenaient à quatre catégories en fonction de la position de la consonne cible (cf tab.III.6).

Les listes ne différaient finalement que par les essais « inducteurs » : dans la première liste, le phonème cible était systématiquement en C_1, dans la seconde il était en C_2 et ainsi de suite. Les listes ont été construites de façon à ce que chaque essai « test » soit toujours précédé d'au moins un essai « inducteur ». On a également fait en sorte qu'un phonème cible particulier (p.ex. un /p/) n'apparaisse jamais dans deux essais successifs. La table III.7 fournit un exemple de quelques essais successifs :

Procédure : La procédure expérimentale était globalement similaire à celle des expériences précédentes. La structure des essais était identique si ce n'est que les sujets étaient informés en cas d'erreur (fausse alarme ou manqué) ; l'ordinateur affichait alors un message d'avertissement et le même essai était « rejoué » immédiatement. Cette modification de la procédure avait pour but de permettre au sujet de se reconcentrer rapidement après une erreur. Les temps obtenus dans les essais « rejoués » n'étaient pas pris en compte dans les analyses de temps de réaction.

C1	C2	C3	C4
Cortège	neCtar	bisCotte	pastèQUe
Discours	gaDget	sarDine	tornaDe
Tournage	ryThmique	facTure	baskeT
Turbine	fooTball	culTure	despoTe
Garnir	maGnum	vulGaire	mustanG
Lapsus	caLmar	carLingue	cerveLLe
Largeur	soLfège	guirLande	survoL
Moustique	gyMnase	kerMesse	costuMe
Reptile	caRtouche	boomeRang	mystèRe
Rustine	feRtile	balleRine	castoR

TAB. III.6 – *Mots Tests (phonème cible en majuscule)*

		Liste			
Type d'essai	Stimulus	1	2	3	4
inducteur	cordage	K	R	D	J
distract.	courgette	F	F	F	F
inducteur	calmante	K	L	M	T
inducteur	mortel	M	R	T	L
test 3	biscotte	K	K	K	K
distract.	charmante	B	B	B	B
inducteur	victime	V	K	T	M
inducteur	servile	S	R	V	L
Test 4	tornade	D	D	D	D

TAB. III.7 – *Quelques Essais Successifs*

L'expérience durait environ un quart d'heure et était donc nettement plus courte que celle de Pitt et Samuel (1990). On peut noter trois autres différences entre notre procédure et la leur : (a) dans notre cas il y a 75 % des essais où la cible apparaît dans le mot (50 % chez eux) ; (b) nous employons une tâche de détection plutôt qu'une décision oui-non ; (c) nous donnons une correction (« feed-back ») aux sujets sur leurs erreurs.

Sujets : Quarante étudiants de diverses universités parisiennes, tous de langue maternelle française ont participé. Ils recevaient 20 francs pour leur participation.

RÉSULTATS

Les taux d'erreurs étaient plutôt faibles (2.7 % de fausses alarmes ; 1.3 % de manqués). On a éliminé les temps de réaction inférieurs à 100 msec (2.4 %) et ceux supérieurs à 1500 msec (0.5 %).[6]

Les figures III.3 et III.4 présentent des analyses « côuts/bénéfices», c'est à dire, pour chaque groupe et chaque position, la différence entre le score de ce groupe sur cette position et la moyenne des autres groupes sur cette position. La table III.8 présente les moyennes détaillées.

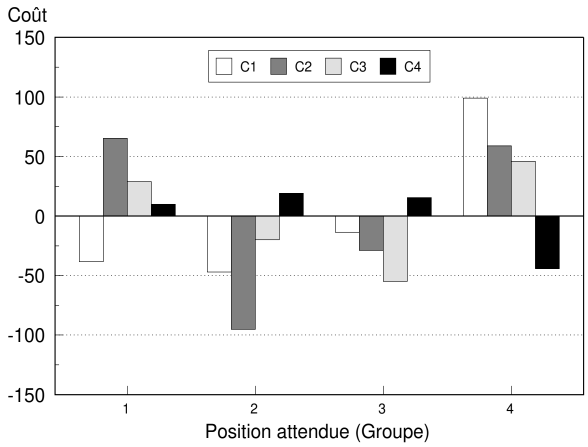


FIG. III.3 – *Coûts/Bénéfices sur les temps de détection de phonème dans* C_1VC_2–C_3VC_4

Les temps de réaction et les erreurs ont été soumis à des analyses de variance (détaillée en annexe, p.142) où deux facteurs étaient déclarés : Induction (biais sur C1, C2, C3 ou C4 dépendant du groupe de sujets) et Position de la cible (C1, C2, C3 ou C4). Tout d'abord, le facteur Position produit un effet significatif : les sujets sont plus rapides pour répondre vers la fin des mots. L'interaction globale Position $\times$ Induction est significative, ce qui montre que les groupes de sujets n'ont pas eu le même comportement. Les significativités des coûts/bénéfices ont été évaluées en examinant toutes les interactions Induction $\times$ Position opposant deux à deux

6. De plus, nous avons éliminé un item (despote) car il avait été mal prononcé « tespote », et les sujets détectaient le premier /t/, alors qu'il aurait fallu que ce soit le second (les temps de réaction étaient tous négatifs). Les conclusions des ANOVAs sont les mêmes, avec ou sans cet item.

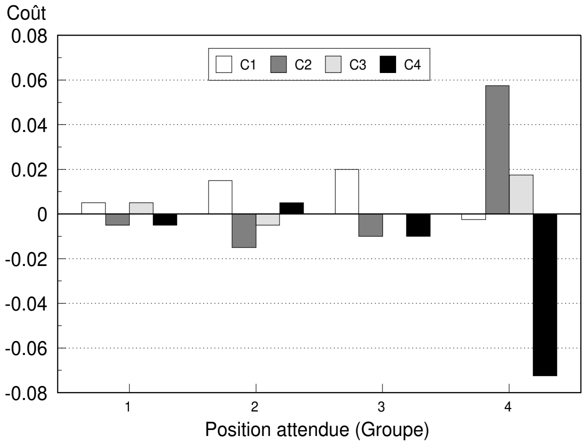


FIG. III.4 – *Coûts/Bénéfices sur les taux d'erreurs en détection de phonème dans* C_1VC_2–C_3VC_4

Position de la cible				
Groupe	C1	C2	C3	C4
1	503 (1%)	596 (3%)	488 (2%)	344 (1%)
2	494 (0%)	436 (1%)	439 (1%)	353 (7%)
3	527 (1%)	502 (2%)	404 (0%)	350 (4%)
4	640 (0%)	590 (3%)	505 (1%)	290 (1%)

[a] Chaque cellule est la moyenne de 100 mesures. (10 Ss $\times$ 10 items). L'ecart-type entre sujets est de 104 msec, entre items, il est de 103 msec ; la moyenne des écart-types intra-sujet, intra-condition est de 104 msec.

TAB. III.8 – *Temps de Réaction (en ms) et Taux d'Erreurs pour chaque Groupe dans chaque position*

des groupes d'induction sur leurs positions favorisées.[7] Par exemple, dans les anovas qui suivent « `I1 I2.P1 P2` » désigne l'interaction restreinte aux groupes 1 et 2 et aux positions C1 et C2. Dans les analyses par items et par sujet des temps de réaction, toutes ces interactions sont significatives. On a également opposé chaque groupe d'induction aux trois autres pris ensemble, avec des interactions de type `I1 vs I2 I3 I4 . P1 vs P2 P3 P4` ; encore une fois ces interactions étaient significatives. On a également testé des interactions de type `I1 I3.P2`

7. Les interactions permettent de retirer les composantes principales dues à la position et aux différences entre sujets.

`P4` pour examiner si un groupe était également avantagé sur des positions autres que celle sur laquelle il était habitué. Par exemple, s'il y avait un avantage sur la syllabe entière, les sujets du groupe 1 pourraient être meilleurs sur la position 2 que les sujets du groupe 4. En fait, ces analyses révèlent que les groupes 2 et 3 sont plus proches l'un de l'autre que ne le sont tous les autres couples.

Discussion

Les sujets français se conduisent comme les sujets américains de Pitt et Samuel (1990) : ils focalisent leur attention sur un phonème précis plutôt que sur une syllabe entière. Il est donc correct d'affirmer que les sujets, même français, focalisent leur attention sur une position phonémique définie syllabiquement, plutôt que sur une syllabe entière.

Dans cette expérience, comme dans les précédentes, les sujets n'étaient pas avertis à l'avance que le phonème apparaîtrait plus souvent dans une position que dans une autre. Interrogés de façon informelle à la fin de la passation, la plupart des sujets avaient remarqué que la cible apparaissait plus souvent soit en début, soit au milieu, soit à la fin des mots. Mais, comme dans les précédentes expériences, les sujets induits en milieu de mots, n'avaient pas remarqué que la cible arrivait plus souvent soit dans la première, soit dans la seconde syllabe.

L'emploi du terme « focalisation de l'attention » provient du paradigme de Posner qui a inspiré Pitt et Samuel. Cependant, ici, les sujets ne sont apparemment pas conscients de la manipulation probabiliste des listes. Il est possible que ces résultats s'apparentent davantage à l'apprentissage implicite d'une régularité, plutôt qu'à une focalisation de l'attention. La plupart des études sur l'apprentissage implicite (cf Reber 1989 pour une revue) ont examiné l'acquisition des *grammaires*, et montrent que le sujet a besoin d'un nombre assez important d'essais pour apprendre la régularité. Nous avons donc examiné si l'avantage « attentionnel » augmentait au cours de nos expériences, à l'aide de diverses mesures : d'une part, nous avons comparé l'effet entre la première et la deuxième moitié des listes ; d'autre part, nous avons examiné la corrélation entre la taille de l'effet et la position des mots tests dans la liste. Aucune de ces analyses n'a révélé d'effet significatif. En conclusion, il ne semble pas que l'effet nécessite un long entraînement. La régularité que les sujets « apprennent » est donc plus facile que l'apprentissage d'une grammaire.

Discussion Générale

Pitt et Samuel pensaient que le fait que les sujets puissent focaliser leur attention sur un phonème précis, démontrait que celui-ci, plutôt que la syllabe, était l'unité de perception. Cependant leurs temps de réactions étaient tels que le phonème pouvait très bien avoir été identifié après la syllabe.

Pitt et Samuel s'intéressaient essentiellement à la « taille du faisceau attentionnel » et ne discutent pas de la propriété sur laquelle les sujets « focalisent leur attention ». Leur article suggérait, implicitement, que les sujets peuvent s'habituer à détecter un phonème dans une position phonémique séquentielle, c'est à dire que la propriété pertinente est celle de « n^e phonème ». Dans les études de ce chapitre, nous avons entrepris de montrer que les sujets s'habituent en fait à détecter des phonèmes placés dans une position précise de la structure syllabique des stimuli.

Dans l'expérience 4, nous avons fait varier la position syllabique du phonème cible plutôt que sa position séquentielle : un groupe de sujets recevait ses cibles plus souvent en coda de la première syllabe de mots bisyllabiques (ex. « caP–tif »), alors qu'un second groupe les recevait plus souvent dans l'attaque de la seconde syllabe (ex. « ca–Price ») ; dans les deux cas, la cible était en troisième position séquentielle. Les deux groupes de sujets étaient testés sur un sous-ensemble de stimuli (identiques pour les deux groupes) où le phonème cible apparaissait soit dans l'une, soit dans l'autre des positions syllabiques. On a observé que les sujets ont détecté plus rapidement les cibles qui apparaissaient dans la position syllabique favorisée par leur liste, ce qui s'est traduit par une interaction entre l'induction attentionnelle des sujets et la position syllabique de la cible. Ceci montre que les sujets ont utilisé une représentation syllabique des stimuli pour effectuer leur réponse.

Cette représentation provenait-elle d'un traitement pré-lexical ou bien avait-elle une origine post-lexicale, c'est à dire, était-elle « récupérée » seulement après l'identification des mots ? La taille des temps de réactions (800 ms depuis le début du mot) et des traces de corrélation entre temps de réponse et fréquence lexicale des stimuli pour un des deux groupes rendaient l'hypothèse post-lexicale envisageable. L'expérience 4bis avait pour but de diminuer les chances d'emploi d'une stratégie lexicale par les sujets en accélérant leurs réponses. A cette fin, la tâche a été modifiée : plutôt qu'une décision oui-non, les sujets devaient effectuer une détection (paradigme « go-no go »). Bien que les temps de réaction aient diminué de 150 ms (350 au lieu de 500 ms), l'effet syllabique demeure. Ce résultat rend improbable, sinon impossible, l'emploi par les sujets d'informations post-lexicales (le chapitre suivant renforcera cette conclusion en présentant une réplication avec des stimuli « non-mots »).

L'apparition d'un effet syllabique dans les expériences 4 et 4bis est d'autant plus remarquable que rien dans la tâche de détection de phonème n'encourage explicitement le sujet à utiliser une information syllabique (De plus les sujets, interrogés à la fin de l'expérience, n'avaient pas remarqué « consciemment » la régularité syllabique). Ceci oppose notre méthode avec la tâche de détection de fragments de Mehler et al. (1981), où le sujet manipule nécessairement des syllabes (puisqu'il doit détecter, typiquement, /pa/ ou /pal/), fait qui pouvait encourager l'emploi d'une « stratégie » syllabique.

Néanmoins, on pourrait imaginer que, dans nos expériences, c'est l'existence d'une régularité syllabique qui encourage les sujets à utiliser une représentation faisant référence aux syllabes.n En l'absence d'une telle régularité, les sujets pourraient-ils utiliser une représentation purement phonémique (i.e une simple séquence de phonèmes) ? C'est ce que nous avons examiné dans l'expérience 5 : les sujets détectaient des phonèmes cibles qui apparaissaient avec des fréquences égales dans les deux positions syllabiques (coda ou attaque), mais beaucoup plus souvent en troisième position séquentielle (p.ex. « caPtif ») qu'en quatrième (p.ex. « proBlème »). Pourtant, les sujets étaient aussi rapides pour détecter un phonème en quatrième position séquentielle qu'un phonème en troisième position ; ils ne pouvaient donc apparemment pas utiliser une régularité purement séquentielle, même en l'absence de régularité syllabique.

Comme elles ne testent qu'une seule position phonémique à l'intérieur de chaque syllabe, les expériences 4 et 4bis ne permettent pas de savoir si les sujets focalisaient leur attention sur toute la syllabe ou bien, plus précisément, sur une seule position phonémique dans la structure syllabique. L'expérience de Pitt et Samuel (1990) favorisait plutôt la seconde alternative puisque les sujets habitués à détecter un phonème dans une syllabe ne transféraient pas leur avantage à un autre phonème appartenant à la même syllabe. Toutefois, comme il a été proposé que les sujets anglais sont plus enclins à utiliser des « stratégies phonémiques », et les sujets français « des stratégies syllabiques » (Cutler et al., 1986; Norris & Cutler, 1988), il importait de savoir si les Français focalisent leur attention sur la syllabe entière ou sur un phonème précis. L'expérience 6 reproduit celle de Pitt et Samuel et montre que les sujets français, habitués à détecter un phonème dans une syllabe, ne transfèrent pas leur avantage attentionnel à un autre phonème situé dans la même syllabe. En conclusion, on peut affirmer que les sujets focalisent leur attention sur *une position phonémique précise dans la structure syllabique* des stimuli.

Ces expériences montrent que les sujets engagés dans une tâche de détection de phonème utilisent une représentation syllabifiée plutôt qu'une représentation purement séquentielle. Cela est consistent avec la conclusion de l'expérience 3 du chapitre précédent. Ces deux types de tâches, cependant, permettent difficilement de conclure quant aux relations temporelles de l'identification (par le système perceptif) des différentes unités (syllabe ou phonème). Pourtant, un de nos résultats ne s'accorde pas avec un modèle syllabique où l'identification de la syllabe serait l'étape limitante après laquelle, seulement, les phonèmes deviendraient *immédiatement* accessibles : s'il fallait attendre la fin d'une syllabe pour avoir accés aux phonèmes qu'elle contient, on s'attendrait à ce qu'un coda soit détecté beaucoup plus rapidement qu'une attaque. Or, on a observé que les sujets étaient aussi rapides pour détecter un phonème, qu'il soit placé en début de seconde syllabe ou en fin de première syllabe. Nous reviendrons sur cette observation dans le chapitre de conclusion (chap.VII) ; de plus, le chapitre V décrira des expériences tentant plus directement d'examiner la question de la primauté temporelle de l'identification du phonème (ou du trait phonétique) sur la syllabe.

Finalement, les expériences de ce chapitre renforcent la thèse que les sujets *français* sont sensibles à la structure syllabique des stimuli (Cutler et al., 1986; Mehler et al., 1981; Segui, 1984; Segui et al., 1990). Le paradigme attentionnel semble offrir un instrument de choix pour examiner la sensibilité des locuteurs de différentes langues à la syllabe, sensibilité qui, pour le moment, n'a été explorée quasiment qu'avec la tâche de détection de fragment (Cutler et al., 1986; Sebastian-Gallés et al., 1992; Otake, Hatano, Cutler, & Mehler, 1993; Bradley et al., 1993; Zwitserlood, Schriefers, Lahiri, & van Donselaar, 1993). Dans le chapitre suivant, nous présentons les résultats obtenus avec le paradigme attentionnel chez des locuteurs de l'espagnol et de l'anglais.

Chapitre IV

Induction Syllabique en Espagnol et en Anglais

Dans l'expérience de Mehler et al. (1981), les sujets détectaient plus rapidement un fragment (p.ex. /ba/ ou /bal/) quand celui-ci correspondait précisément à la structure syllabique du stimulus (p.ex. /ba–lance/ ou /bal–con/). Selon les auteurs, cet effet de « congruence syllabique » était l'indice d'une segmentation et d'une catégorisation en syllabes du signal de parole. Bien que l'expérience n'avait été réalisée qu'avec des sujets français, cette segmentation syllabique était censée être universelle. La syllabe était proposée comme l'« unité perceptive primaire », qui servait à l'acquisition des mots par l'enfant (chacun étant une suite de syllabes), et également à l'accès au lexique mental chez l'adulte (Mehler, 1981; Mehler, Dupoux, & Segui, 1990; Segui, 1984).

Cependant, Cutler, Mehler, Norris et Segui (1983; 1986), qui tentaient de reproduire l'expérience de Mehler et al. (1981) en anglais, ont trouvé que les locuteurs de cette langue n'étaient pas influencés par la structure syllabique des stimuli. Ainsi, ils détectaient /bal/ et /ba/ à la même vitesse dans /balance/ et dans /balcony/. Mieux, cette étude démontre que le contraste entre français et anglais ne réside pas dans une différence entre les stimuli anglais et français, mais dépend de la langue maternelle du locuteur : en échangeant la langue des sujets et celle des stimuli, c'est à dire en faisant écouter des stimuli anglais aux sujets français et vice-versa, il est apparu que les locuteurs français syllabifiaient les stimuli anglais et que les locuteurs anglais ne syllabifiaient pas les stimuli français. Pour rendre compte de ces résultats, Cutler et al. (1986) ont proposé que les sujets possèdent des *stratégies de segmentation* dépendantes de la langue maternelle.[1] Pourtant, les motivations principales pour postuler des unités de segmentation prélexicale

1. L'expression « stratégie de segmentation » est doublement ambiguë. Tout d'abord, il y a une ambiguïté sur le terme de « segmentation » que Cutler et al. (1986) ne dissipent pas clairement. À l'origine, la « segmentation » désignait le découpage du signal de parole en unités prélexicales (cf Mehler et al., 1981; cette conception est clairement exposée dans Mehler, Segui et Dupoux,

étaient les problèmes de l'acquisition du lexique par l'enfant, et de la normalisation perceptive (Mehler et al., 1990). Il était donc paradoxal de proposer que l'unité de segmentation dépendait de la langue. C'est pourquoi Cutler et al. (1986) ont proposé que la stratégie de segmentation n'était pas absolument arbitraire : l'enfant disposerait initialement d'un répertoire fini de stratégies possibles et « sélectionnerait » celle qui est la mieux adaptée à la langue maternelle. Logiquement, Cutler et al. ont recherché quelle(s) caractéristique(s) de la langue pouvaient favoriser ou non l'emploi d'une stratégie de segmentation syllabique.

Dans leur article de 1986, ils notent que la *clarté* des frontières syllabiques

---

1990). Cutler et Norris (Norris & Cutler, 1985; Cutler & Norris, 1988), eux, désignent par « segmentation » la localisation des frontières de mots dans le signal. Ils sont agnostiques vis à vis du problème de la catégorisation prélexicale. Leur interprétation des résultats des expériences de détection de fragment est que les Français utiliseraient les frontières de syllabes comme des indices de frontière de mots possibles. Les Anglais, eux, utiliseraient une stratégie de segmentation *lexicale* différente, en ne postulant des frontières de mots que devant les syllabes « fortes », c'est à dire celles contenant une voyelle pleine. Cette « stratégie de segmentation métrique » est démontrée dans des expériences où les sujets anglais écoutent de la parole murmurée et commettent des erreurs de segmentation lexicale aux endroits prédits par la théorie (Cutler & Butterfield, 1992). Celle-ci prédit aussi le comportement des sujets dans une tâche de « word spotting », où les sujets doivent reconnaître le mot qui débute un stimulus multisyllabique (Cutler & Norris, 1988). De plus, les statistiques lexicales montrent que cette stratégie doit être efficace, car une grande majorité des mots anglais débutent par une syllabe forte (Cutler & Carter, 1987).

L'ensemble de données convergentes soutenant l'hypothèse d'une stratégie de segmentation métrique chez les Anglais est impressionnant. Pourtant, la proposition de Norris et Cutler (1985) selon laquelle les résultats obtenus en détection de fragment montrent que les français appliqueraient, eux, une stratégie de segmentation *lexicale* syllabique, postulant des frontières de mot à chaque frontière syllabique, nous posent deux problèmes : tout d'abord, ils supposent que les tâches de détection de fragment et de « word spotting » reflètent toutes deux une stratégie de segmentation *lexicale*. Pourtant, à ma connaissance, les tentatives de répliquer un résultat à la « papale » en manipulant le statut « fort » ou « faible » de la seconde syllabe, se sont avérées infructueuses. Il n'est donc pas du tout évident que les deux tâches mettent en jeu les mêmes niveaux de traitement. Deuxièmement, la proposition selon laquelle les français postuleraient des frontières de mots aux frontières syllabiques est troublante si l'on considère qu'en français (largement plus qu'en anglais il me semble), les frontières syllabiques correspondent très mal aux frontières de mots à cause de la fréquence des enchaînements et de la liaison. Toutefois, si l'on propose que les frontières à détecter ne sont pas celles des mots mais celles de groupes phonologiques plus larges comme le groupe clitique (ainsi que le propose Anne Christophe dans sa thèse ; 1993), alors les frontières syllabiques sont de bons indices, mais cela vaut pour l'Anglais comme pour le Français.

Une seconde ambiguïté de l'expression « stratégie de segmentation » provient de l'emploi du terme « stratégie ». En psychologie expérimentale, une stratégie désigne généralement une attitude que le sujet est libre d'adopter ou non pour résoudre la tâche à laquelle il est soumis. Or la « stratégie » à laquelle Cutler et al. (1986) font référence est supposée être un mécanisme de base du système de traitement de la parole, dont on ne s'attend pas à ce qu'il soit sous le contrôle stratégique du sujet. Par exemple, il n'est pas évident que les sujets bilingues, susceptibles de posséder plusieurs stratégies de segmentation (Cutler, Mehler, Norris, & Segui, 1989; Kearns, 1994; Bradley et al., 1993), soient capables de passer de l'une à l'autre à volonté.

diffèrent entre les langues. Si la syllabification du français est intuitivement claire (pour les Français), il n'en va pas de même pour l'anglais : les sujets anglais ne savent pas bien où placer la frontière syllabique dans /balance/. En fait, si on leur demande quelles sont, respectivement, la première et la seconde syllabe de ce mot, beaucoup répondront /bal/ et... /lance/. Souvent, ils hésitent à attribuer le /l/ aux deux syllabes et préfèrent ne pas répondre.[2] Les analyses linguistiques, reflétant les intuitions contradictoires des sujets, décrivent le /l/ de "balance" comme *ambisyllabique*. Les principes de base de la syllabification (Principe de l'Attaque Maximale[3] et Règles de sonorité[4]) s'accordent à segmenter une suite CVCV entre la voyelle et la consonne (i.e. CV–CV). Toutefois, en anglais, les consonnes situées entre deux syllabes dont la première est accentuée et la seconde est inaccentuée[5], se conduisent *phonétiquement* comme si elles appartenaient au coda de la première syllabe plutôt qu'à l'attaque de la seconde. Ainsi, "balance" serait syl-

2. En français aussi, il y a des cas où la syllabification n'est pas évidente. Le cas du /s/ est notoire : /costume/ doit-il être syllabé en /co-stume/ ou en /cos-tume/ ? Il n'est pas rare si l'on demande au même sujet quelle est la première syllabe qu'il réponde /cos/, mais que lorsqu'on lui demande d'inverser les deux syllabes, il réponde /stumco/. D'autres groupes consonantiques ont une syllabification ambiguë : /stag-nant/ ou /sta-gnant/, /a-tlantique/ ou /at-lantique/ ? Il existe d'ailleurs différentes théories de la syllabification en français (voir p.108, ainsi que les notes suivantes sur les principes de syllabification). Ceci dit, à l'appui de Cutler et al. (1986), les ambiguïtés de syllabification semblent largement plus répandues en anglais qu'en français. Lorsque les Anglais hésitent dans le cas de /palace/, les Français, eux, syllabifient sans hésitation les suites CVCV (en CV–CV).

3. Le principe de l'Attaque Maximale stipule qu'entre deux voyelles séparées par des consonnes, la frontière syllabique est placée de façon à maximiser le nombre de consonnes dans l'attaque (le début) de la seconde syllabe, ces consonnes devant toutefois former un groupe « légal ». Les groupes « légaux » sont ceux qui peuvent apparaître en début de mot (Pulgram, 1970). Ainsi, /pr/ est « légal » (cf, p.ex., /prix/), mais /ct/ ne l'est pas : il n'y a pas de mot français commençant par /ct/. Cela prédit, par exemple, les syllabifications de /trac-teur/ et de /ca-price/. Comme toutes les consonnes peuvent débuter un mot en français (excepté /ng/ comme dans /bingo/), une chaîne VCV est toujours syllabifiée en V-CV par le principe de l'Attaque maximale.

4. Les phonèmes peuvent être placés sur une échelle de sonorité correspondant à peu près à l'ouverture du passage laissé à l'air pour les prononcer ($\simeq$ intensité sonore perçue). Voici une échelle grossière (notons qu'il en existe de plus fines ; cf Goldsmith, 1990) : occlusives (ptkbdg) $<$ fricatives (fs$\int$vzj) $<$ nasales (nm) $<$ liquides (lr) $<$ voyelles. La règle de sonorité stipule (a) que les phonèmes placés en début de syllabe doivent avoir des sonorités croissantes (p.ex. /pr/ est dans ce cas, mais pas /rp/), et (b) que la frontière syllabique se trouve placée juste avant le minimum de sonorité. Cela prédit, par exemple, /car-ton/, et /ra-cler/. Cette règle ne fait pas de prédiction claire pour les cas comme /captif/ ou /p/ et /t/ sont de sonorité égale. Finalement, le principe de sonorité entre en contradiction avec le principe de l'attaque maximale dans des cas comme /gm/ ou /st/ : la sonorité prédit /fra-gment/ et /cos-tume/, l'attaque maximale : /frag-ment/ et /co-stume/ (/gm/ ne peut débuter un mot français). Ces cas sont précisément ceux où l'intuition peut vaciller.

5. En fait, il y a d'autres facteurs que l'accent qui semblent jouer un rôle, notamment le fait que la première voyelle soit longue ou courte, que la seconde voyelle soit réduite ou non...etc. Les opinions des linguistes diffèrent sur ces points.

labifié, *phonétiquement* : /bal-ance/, et *phonologiquement* : /ba-lance/. Notons que les linguistes diffèrent dans leur traitement de l'ambisyllabicité : pour certains, il faut décrire la consonne ambisyllabique comme appartenant *simultanément* aux deux syllabes (Kahn, 1976), alors que pour d'autres, elle appartient à la seconde syllabe, à un certain niveau de représentation, puis à la première syllabe, dans des niveaux plus proches des représentations superficielles (Selkirk, 1982).

Selon la théorie de Cutler et al. (1986), le fait que les frontières syllabiques soient ambiguës en anglais, décourage l'acquisition d'une stratégie de segmentation syllabique chez les enfants qui apprennent cette langue. Par contre, ils prédisent que les sujets dont la langue maternelle présente des frontières syllabiques claires doivent appliquer la stratégie de segmentation syllabique. Ceci est le cas de l'espagnol. Pourtant, quand Sebastian-Gallés et al. (1992) ont effectué une expérience de détection de fragments avec des sujets espagnols, ils n'ont pas, à leur grande surprise, observé l'interaction caractéristique de l'« effet syllabique ». En cherchant des raisons à ce résultat, ils ont invoqué une explication déjà avancée par Dupoux dans sa thèse : dans les tâches de détection, quand les sujets répondent très rapidement, ils peuvent exploiter un code sub-syllabique « transitoire » (hypothétiquement des demi-syllabes pour Dupoux, 1989). Selon Sebastian-Gallés et al. (1992), l'identification des voyelles pourrait être facilitée chez les Espagnols par le fait qu'elles sont en plus petit nombre qu'en français (cinq contre quinze), et, par hypothèse, acoustiquement plus « claires ». Ils ont nommé cette propriété (hypothétique) « transparence acoustique des voyelles ». En ralentissant les réponses des sujets dans le but de décourager l'emploi d'une stratégie « acoustique », les auteurs ont prédit qu'ils observeraient à nouveau des réponses « syllabiques », et ce fût effectivement le cas. Nonobstant la transparence acoustique, les Espagnols, qui possèdent une langue avec des frontières syllabiques claires, appliquent donc une stratégie de segmentation syllabique, en accord avec la théorie de Cutler et al. (1986).

D'autres résultats remettent toutefois en cause cette théorie. Tout d'abord l'interaction syllabique a été observée en hollandais, une langue où l'ambisyllabicité est très répandue (Zwitserlood et al., 1993). Puis, Bradley et al. (1993) ont trouvé que les Espagnols n'appliquaient pas la segmentation syllabique quand ils écoutaient des stimuli anglais (contrairement aux Français dans l'étude de Cutler et al.; 1986). Pire, selon cette dernière étude, des sujets espagnols ayant appris l'anglais assez tardivement ne syllabifieraient plus même l'espagnol ! Pour être juste, il faut noter que la théorie selon laquelle les sujets utilisent une stratégie de segmentation dépendante de leur langue maternelle (fondée sur l'unité rythmique de celle-ci dans la version récente de la théorie ; cf Cutler 1993) a aussi remporté des succès : elle a correctement prédit le comportement de sujets Français, Anglais et Japonais, écoutant des stimuli japonais (Otake et al., 1993).

Finalement, les résultats obtenus avec la détection de fragments offrent un tableau complexe (cf Kearns, 1994 pour une revue). Les résultats obtenus en espagnol (Sebastian-Gallés et al., 1992; Bradley et al., 1993) montrent, me semble-t-il, que s'il est juste de considérer la présence de l'interaction comme prouvant que les sujets utilisent la structure syllabique, il ne faut pas déduire de l'absence d'interaction, que les sujets ne « récupèrent » pas la structure syllabique des stimuli. De nombreuses raisons méthodologiques peuvent conduire à l'absence d'effet significatif, même s'il est vrai que les expériences avec les sujets anglais de Cutler et al. (1986) étaient aussi similaires que possible avec celles menées avec les Français, et si *l'absence* d'interaction syllabique a été reproduite par Bradley et al. (1993) et Otake et al. (1993). De plus, étant donné que l'effet syllabique peut disparaître même chez des sujets français quand leurs temps de réaction sont rapides (Cf Dupoux, en préparation), on doit prendre en considération la possibilité que les sujets anglais, comme les Français (Mehler et al., 1981), les Espagnols (Sebastian-Gallés et al., 1992; Bradley et al., 1993), les Portugais (Morais, Content, Cary, Mehler, & Segui, 1989), et surtout les Hollandais dont la langue est plus proche de l'anglais (Zwitserlood et al., 1993), se forment une représentation syllabique des stimuli, mais que, pour des raisons qu'il faudra expliquer, ne sont pas influencées par celle-ci *dans la tâche de détection de fragment*.

Quand Cutler, Mehler, Norris et Segui (1986) affirment que « les Anglais ne sont pas influencés par la structure syllabique des stimuli », ils devraient ajouter « dans les premières étapes de la perception ». En effet, d'une part, de nombreux arguments linguistiques existent, qui justifient le rôle de la syllabe dans les processus phonologiques en anglais comme dans les autres langues, et d'autre part, des mesures psycholinguistiques révèlent l'influence de la structure syllabique sur le comportement des sujets anglais. Ces données psycholinguistiques proviennent essentiellement des travaux de R. Treiman (Treiman, 1983, 1986; Treiman & Danis, 1988a, 1988b; Treiman & Zukowski, 1990; Treiman, Straub, & Lavery, 1994; Fowler, Treiman, & Gross, 1993) où les sujets doivent effectuer des tâches métalinguistiques de manipulation (duplication, transpositions...) de phonèmes. Typiquement, les sujets apprennent plus facilement les opérations qui respectent la structure syllabique plutôt que celles qui la violent. Cutler et al. ne prennent pas en compte de tels résultats, sans doute parce qu'ils considèrent que ces tâches sont tardives, et ne reflètent pas les premières étapes du traitement de la parole (Treiman elle-même admet que les sujets sont influencés, par exemple, par l'orthographe).

Quoiqu'il en soit, une théorie ambitieuse, selon laquelle les enfants acquièrent une procédure de segmentation de la parole qui dépend de la langue maternelle, doit être fondée sur d'autres paradigmes que celui de la détection de fragments. C'est pourquoi, nous avons décidé de réaliser des expériences de détection de phonème attentionelle, biaisée syllabiquement, avec des sujets de langues maternelles espagnol et anglais.

TROIS RÉPLICATIONS EN ESPAGNOL

Rappelons que les locuteurs espagnols sont supposés être influencés par la structure syllabique des stimuli ; cette prédiction se fonde sur le fait que les frontières syllabiques sont claires en espagnol. Pourtant, avec la tâche de détection de segment, l'interaction caractéristique de la segmentation syllabique n'a été observée qu'à des temps de réaction relativement lents (cf supra). C'est pourquoi nous avons voulu utiliser le paradigme de détection attentionnelle dans cette langue.

N. Sebastian-Gallés et T. Felguera, de l'université de Barcelone, ont conduit, avec notre collaboration, trois expériences de détection de phonème attentionnelle avec biais syllabique. La première était semblable en tous points à l'expérience 4 : les mots français étaient simplement remplacés par des mots espagnols (cf annexe p.145). Les sujets étaient des étudiants de l'université de Barcelone, tous de langue maternelle espagnole (certains étaient bilingues espagnol-catalan). Les résultats sont présentés sur la figure IV.1 :

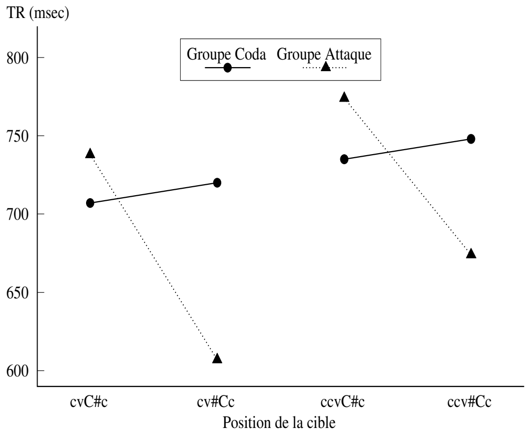


FIG. IV.1 – *Espagnol — Détection avec biais syllabique I*

Dans une anova sur les temps de réaction[6], où étaient déclarés les facteurs : Syllabe (cible en coda ou en attaque), Position (3ème ou 4ème phonème) et Induction (groupe recevant une liste biaisée soit vers le coda, soit vers l'attaque), l'interaction Syllabe $\times$ Induction est hautement significative ($F1(1,48) = 25.5; F2(1,28) = 25.9$) ; il y a également un effet principal dû au facteur Syllabe, les sujets étant plus rapides pour détecter les cibles en attaque plutôt que les cibles en coda ($F1(1,48) = 16.0; F2(1,28) = 4.2$).

6. Nous n'évoquerons pas les anovas sur les erreurs qui ne révèlent aucun effet statistiquement significatif.

Hormis l'effet principal de Syllabe, ces résultats sont similaires à ceux de l'expérience française : les Espagnols détectent plus rapidement un phonème cible placé dans une position syllabique attendue que dans une position non-attendue. Comme les temps de réaction étaient assez lents, une seconde expérience a été réalisée dans le but d'accélérer les réponses des sujets. Pour cela, les instructions insistaient sur l'importance de la vitesse de réponse et la tâche de décision (oui/non) était remplacée par une tâche de détection (go/no-go). Les résultats de cette seconde expérience sont dessinés sur la figure IV.2.

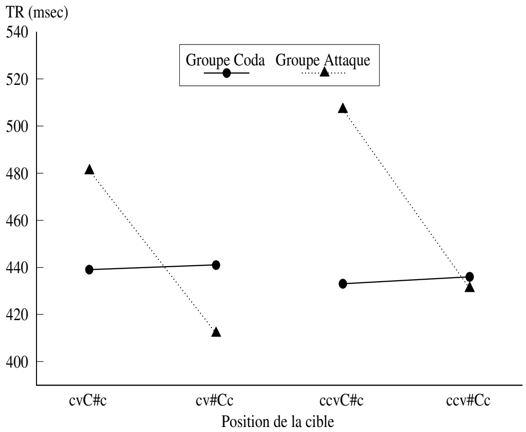


FIG. IV.2 – *Espagnol — Détection avec biais syllabique II*

Ces résultats sont qualitativement identiques à ceux de l'expérience précédente. L'analyse de variance conduit exactement aux mêmes conclusions que la première expérience (Interaction Syllabe × Induction : $F1(1, 34) = 14.3, F2(1, 28) = 8.9$; Syllabe : $F1(1, 34) = 12.5, F2(1, 28) = 4.2$). Bien que les temps de réaction soient assez rapides, on ne peut être assuré de leur statut prélexical. Par conséquent, une troisième expérience a été réalisée où, cette fois, les stimuli étaient des non-mots (cf annexe p.145) pour conforter l'hypothèse que ces résultats n'ont pas une origine post-lexicale. La figure IV.3 montrent les temps de réaction obtenus.

Statistiquement, seule l'interaction Syllabe × Induction est significative ($F1(1, 24) = 22.8; F2(1, 28) = 21.7$), ce qui révèle encore l'efficacité de l'induction attentionnelle syllabique. Il n'y a, par contre, plus d'effet de position syllabique ce qui suggère que celui observé dans les deux expériences précédentes pouvait provenir de caractéristiques accidentelles du matériel. Une autre possibilité est que, quand les stimuli étaient des mots, les réponses des sujets étaient influencées par une rétroaction lexicale, et que *cette influence soit plus importante pour les phonèmes de la seconde syllabe que pour les phonèmes de la première syl-*

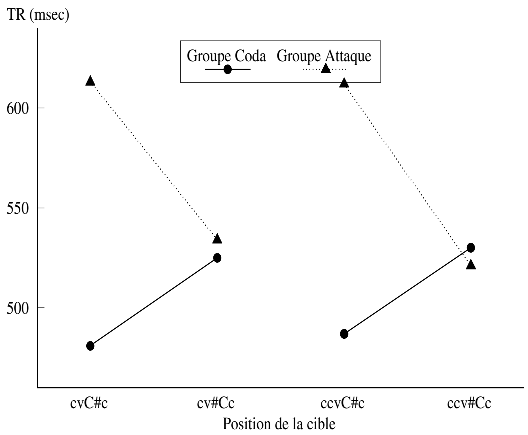


FIG. IV.3 – *Espagnol — Détection avec biais syllabique — Non-mots*

labe. Si cette interprétation était correcte, cela pourrait être considéré comme une belle démonstration du modèle d'accés au lexique par la première syllabe (Cole & Jakimik, 1980; Dupoux, 1989) : les détections de phonèmes placés dans la première syllabe seraient prélexicales, et celles des phonèmes placés dans les syllabes suivantes seraient postlexicales. Cela demeure trés spéculatif étant donné que le matériel n'était pas contrôlé pour d'autres facteurs (fréquence, acoustique...etc) et rappelons qu'en français, nous n'avons pas observé de différence entre les codas et les attaques.[7]

En conclusion, les résultats des trois expériences sont clairs : les sujets espagnols peuvent focaliser leur attention sur une position syllabique précise. Cela conforte l'hypothèse selon laquelle la syllabe joue un rôle dans la perception de la parole dans cette langue. De plus, la troisième expérience rend moins plausible l'hypothèse que la structure syllabique ne puisse être récupérée qu'après l'identification du mot.

Sebastian-Gallés et al. (1992) avaient invoqué la « transparence acoustique » des voyelles espagnoles pour expliquer l'absence d'interaction syllabique dans la tâche de détection de segment quand les réponses étaient rapides. Dans la tâche de

7. Cf l'expérience 6a, p.130 de la thèse de Anne Christophe (1993), qui trouve, en français, une tendance à ce que les phonèmes en attaque de seconde syllabes soient plus faciles à détecter que les phonème en coda de première syllabe. Il est difficile de savoir si cela provient d'un effet de facilitation lexicale, ou si, plus prosaïquement, les phonèmes en coda sont moins bien articulés (ou sont moins « prototypiques »), que les phonèmes en attaque. On pourrait réaliser une expérience avec des stimulis synthétiques où l'on égalise les caractéristiques des phonèmes en coda et des phonèmes en onset, mais alors il n'est pas évident que ces stimuli élicitent un traitement « naturel ».

détection attentionnelle, il n'y a pas d'artefact possible du à la transparence acoustique, et l'on peut penser que cela explique que les résultats soient « plus stables » que ceux obtenus avec la détection de segment. Toutefois une autre interprétation peut être invoquée : ce serait parce que le sujet est plus avancé dans le stimulus, au moment où il effectue la détection de phonème généralisé (par rapport à la détection de fragment initial), que l'information syllabique est plus claire et peut influencer la réponse. Quoiqu'il en soit, il est possible que la tâche de détection attentionnelle fournisse au sujet plus de chances d'utiliser une représentation syllabique. Le cas le plus crucial est bien sûr celui de l'anglais, pour lequel les sujets n'ont jamais semblé sensibles à la structure syllabique dans la tâche de détection de segment.

EXPÉRIENCE 7: INDUCTION SYLLABIQUE EN ANGLAIS

L'expérience suivante est une adaptation à l'anglais de la détection de phonème biaisée syllabiquement (exp. 4). La sélection du matériel pour réaliser une telle expérience est très délicate. Si l'on sélectionne aléatoirement des mots CV–CC et des mots CVC–C dans un dictionnaire anglais (en se fondant sur le principe de l'attaque maximale pour justifier la syllabification), il apparaît qu'il y a une forte corrélation entre la structure syllabique et le caractère tendu ou lâche de la première voyelle [8] : la plupart des voyelles sont tendues dans la structure CV–CC et lâches dans la structure CVC–C (au point qu'il n'y a tout simplement pas de mots anglais CVC–C avec V tendue où le coda serait une occlusive). Pour éviter que les sujets ne s'habituent à la nature de la première voyelle plutôt qu'à la structure syllabique, nous avons limité la sélection aux mots avec première voyelle lâche. [9] Pour éviter également que le pattern accentuel ne distingue les deux catégories « coda » et « attaque », nous avons imposé à tous les stimuli d'avoir l'accent principal en première syllabe, ainsi qu'une seconde syllabe non accentuée. Ces restrictions sélectionnent finalement un matériel semblable à celui utilisé en détection de fragment par Cutler et al. (1986) (paires de type « balance / balcony »).

DESCRIPTION

Matériel : Les stimuli étaient tous des mots américains (sélectionnés dans le *Webster's Pocket*). On a enregistré le phonologue américain M. Hammond qui avait pour consigne de

8. Rappelons que les voyelles lâches de l'anglais sont : I (bit), /E/ (bet), /U/ (good), /@/ (bat), /^/ (but). La notation que nous utilisons est celle de l'alphabet phonétique informatique.

9. Une conséquence infortunée de ce choix est qu'il ne reste alors pas suffisamment de mots CV–CC avec voyelle lâche pour effectuer l'expérience. Cela nous a contraint à inclure des mots CV-CV dans les inducteurs de type Onset.

les lire avec un accent « standard ». Tous ces mots étaient bi- ou tri-syllabiques, avec l'accent primaire sur la première syllabe et la seconde syllabe non-accentuée. La voyelle de la première syllabe était systématiquement une voyelle lâche. La syllabification était dictée par le principe de l'attaque maximale. Les mots tests et les inducteurs de type « coda » avaient une structure CVC–CV ; les mots tests de type « attaque » avaient une structure CV–CCV, les mots inducteurs de type « attaque » avaient une structure CV–CV (car il n'y avait pas suffisamment de mots de structure CV–CCV à la fois pour les mots tests et pour les inducteurs). Comme il n'y a pas suffisamment de mots de structure adéquate commençant par un groupe CC, nous n'avons pas distingué les positions 3ème et 4ème phonème comme nous l'avions fait dans les expériences françaises et espagnoles. Pour éliminer la possibilité que les sujets s'habituent à détecter certains types d'allophones, les phonèmes cibles étaient différents dans les inducteurs (liquides et nasales) et dans les mots tests (occlusives, cf tab.IV.1). Toutefois, pour tester si cette manipulation risquait de faire disparaître l'induction, nous avons inclu dans les deux listes 10 items tests « bis » ayant exactement les mêmes caractéristiques que les inducteurs (cf tab.IV.2).

attaque		coda	
buttress	`b'^-tr|s`	signal	`s'Ig-nL`
petrol	`p'E-trL`	lecture	`l'Ek-CX`
triplet	`tr'I-pl|t`	spectacle	`sp'Ek-tI-kL`
leprosy	`l'E-pr|-si`	pigment	`p'Ig-mxnt`
public	`p'^-blIk`	chapter	`C'@p-tX`
fabric	`f'@-brIk`	subject	`s'^b-JIkt`
diplomat	`d'I-plx-m`@T`	practical	`pr'@k-tI-kL`
speculate	`sp'E-kyU-l`et`	stigmatize	`st'Ig-mx-t`Yz`
supplement	`s'^-plx-mxnt`	tractable	`tr'@k-tx-bL`
metric	`m'E-trIk`	cutlet	`k'^t-lxt`
tablet	`t'@-bl|t`	medley	`m'Ed-li`
sacrifice	`s'@-kr|-f`Ys`	nectar	`n'Ek-tX`
negligent	`n'E-glI-J|nt`	capsule	`k'@p-sL`

TAB. IV.1 – *Mots tests anglais (avec leur représentation phonétique)*

Procédure : La réalisation technique de l'expérience était identique à celle de l'expérience 4bis française : l'expérience se déroulait sous le contrôle d'un ordinateur qui présentait les cibles, restituait les stimuli, et enregistrait les temps de réaction. Les sujets étaient testés individuellement. Les instructions détaillant la tâche de détection de phonème généralisée sont fournies en annexe (p.147). Les sujets devaient effectuer une décision oui/non à chaque essai. Avant de commencer l'expérience proprement dite, le

attaque		coda	
balance	`b'@-l|ns`	timbre	`t'@m-bX`
challenge	`C'@-l|nJ`	symbol	`s'Im-bL`
tunnel	`t'^-nL`	principle	`pr'In-s|-pL`
honey	`h'^-ni`	balcony	`b'@l-k|-ni`
symmetry	`s'I-mx-tri`	seldom	`s'El-dxm`

TAB. IV.2 – *Mots tests bis anglais (avec leur représentation phonétique)*

sujet effectuait un entrainement d'une dizaine d'essais, durant lequel il recevait un retour (« feed-back ») sur ses erreurs (cible manquée ou fausse alarme). Pendant le bloc expérimental, il n'y avait plus d'information sur les erreurs.

Sujets : 42 sujets, étudiants ou membres du personnel de l'université Rutgers (New Jersey) et de l'université de Pennsylvanie à Philadelphie, d'âges compris entre 18 et 36 ans, ont participé volontairement à l'expérience qui durait une quinzaine de minutes. Quatre sujets supplémentaires qui avaient été testés ont été rejetés pour avoir fait plus de 15 % de fausses alarmes. Les sujets étaient assignés aléatoirement à l'un des deux groupes (induction « coda » ou « attaque »).

RÉSULTATS

Aprés avoir éliminé les erreurs (3.4 % des données) et les temps de réaction situés en dehors de l'intervalle 100..1500 (0.5 % des données), on a obtenu les temps de réaction moyens et les taux d'erreurs sur les mots tests présentés dans la table IV.3, en fonction du groupe d'induction et de la position syllabique de la cible.

	Poisition de la cible		
Groupe	Coda	Attaque	Différence
Coda	462 ms (2.9 %)	441 ms (4.0 %)	+21 ms (-1.1 %)
Attaque	480 ms (4.4 %)	481 ms (3.3 %)	-1 ms (+1.1 %)

TAB. IV.3 – *Temps de réaction et Taux d'erreurs — Induction attentionnelle, sujets américains*

L'inspection de la table IV.3 révèle que l'interaction sur les temps de réaction va dans le sens opposée à celui prédit par l'induction syllabique : le groupe coda est plus rapide en position « attaque » qu'en position « coda ». Cependant, les anovas sur ces données avec les deux facteurs déclarés Syllabe (position de la cible) et Groupe (induction coda ou attaque) ne révèlent aucun effet statistiquement significatif si ce n'est celui de Groupe dans l'analyse par items, ce qui est sans grande signification. L'interaction sur les erreurs va dans le sens prédit, mais n'est pas significative. Finalement, pour examiner si c'est le fait d'avoir des phonèmes différents dans les mots tests et dans les mots inducteurs qui est responsable de l'absence d'induction, nous avons conduit des anovas identiques sur les 10 items tests bis, qui n'ont revélé, là encore, aucun effet significatif.

```
Analyses par sujets : PLAN  S21<G2>*S2
G = Groupe d'induction
S = Position de la cible (S1=Coda et S2=Attaque)
Temps de réaction
S          F(1,40)=  0.73   MSE=2127.15   p=0.3980
G          F(1,40)=  0.54   MSE=35196.7   p=0.4667
G.S        F(1,40)=  1.47   MSE=2127.15   p=0.2325
S/G1       F(1,20)=  1.72   MSE=2630.15   p=0.2046
S/G2       F(1,20)=  0.08   MSE=1624.14   p=0.2198
Erreurs
S          F(1,40)=  0.00   MSE=0.3143   p=0.0000
G          F(1,40)=  0.07   MSE=0.6476   p=0.2073
G.S        F(1,40)=  1.36   MSE=0.3143   p=0.2504
S/G1       F(1,20)=  0.52   MSE=0.4143   p=0.4792
S/G2       F(1,20)=  1.00   MSE=0.2143   p=0.3293
Analyse par items : PLAN  S13<S2>*G2
Temps de réaction
S          F(1,24)=  0.14   MSE=6685.7   p=0.2884
G          F(1,24)=  6.59   MSE=1795.2   p=0.0169
G.S        F(1,24)=  1.08   MSE=1795.2   p=0.3091
S/G1       F(1,24)=  0.60   MSE=4694.85   p=0.4461
S/G2       F(1,24)=  0.02   MSE=3786.05   p=0.1113
Erreurs
S          F(1,24)=  0.00   MSE=1.0096   p=0.0000
G          F(1,24)=  0.22   MSE=0.3429   p=0.3567
G.S        F(1,24)=  2.02   MSE=0.3429   p=0.1681
S/G1       F(1,24)=  0.50   MSE=0.6987   p=0.4863
S/G2       F(1,24)=  0.53   MSE=0.6538   p=0.4737
```

DISCUSSION

L'interaction caractéristique de la focalisation sur une position syllabique n'apparait pas chez les locuteurs anglais. Deux possibilités s'offrent pour expliquer ce résultat : (a) les Anglais, conformément à la proposition de Cutler et al. (1986), ne sont pas influencés par la structure syllabique des stimuli, ou bien (b) la structure syllabique de nos stimuli n'est pas celle qu'on pensait.

Notre matériel était similaire à celui de Cutler et al. (1986) qui n'avaient pas observé d'effet de congruence syllabique dans la tâche de détection de fragments. Comme on l'a évoqué plus haut, la syllabification des stimuli, dont la première syllabe est accentuée et contient une voyelle lâche, est ambiguë. En réalisant cette expérience, notre pari était que le principe de l'attaque maximale l'emporterait, *au niveau de représentation utilisé par les sujets pour effectuer la tâche*, sur les autres facteurs influençant la syllabification,[10] et nous avons considéré que le /l/ de /balance/ appartenait à l'attaque de la seconde syllabe. Si ce /l/ appartient en fait à la première syllabe (ou aux deux simultanément) alors il n'est pas surprenant que l'induction n'ait pas eu lieu car les deux groupes, s'ils focalisaient leur attention sur une position syllabique, la focalisaient sur la position coda de la première syllabe. A l'appui de cette interprétation, on peut citer les intuitions de syllabification des sujets anglais de Treiman et Danis (1988b) : quand la première syllabe était accentuée et contenait une voyelle courte (=lâche), les sujets préféraient le découpage VC–V dans 66 % des cas.

Tout cela demeure spéculatif. Toutefois, une pierre de touche de l'argumentation de Cutler et al. (1986) est que les Français, eux, sont sensibles à la structure syllabique, non seulement des stimuli français mais aussi des stimuli anglais. Nous avons donc décidé de suivre leur exemple et de faire passer l'expérience de détection de phonème attentionnelle américaine à des sujets français.

EXPÉRIENCE 8: INDUCTION AVEC STIMULI AMÉRICAINS ET SUJETS FRANÇAIS

DESCRIPTION

Sujets : 30 sujets de langue maternelle française, étudiants en lettres, en allemand ou en anglais, de première ou seconde année de l'université de Metz, ont participé à cette expérience. Tous avaient étudié l'anglais au lycée, et leur niveau moyen était moyen (on a refusé trois sujets qui avaient effectué de longs séjours en Angleterre). Quatre sujets ont été rejetés pour avoir fait trop d'erreurs.

10. Pourquoi prendre ce pari ? Nous étions influencés par l'analyse de l'ambisyllabicité de Selkirk (1982) : celle-ci suppose qu'au niveau phonétique superficiel, le phonème appartient à la première syllabe, et qu'à un niveau phonologique plus abstrait, elle appartient à la seconde syllabe. Notre « pari » était que les sujets employaient une représentation phonologique plutôt que phonétique pour effectuer la tâche de détection de phonème.

Procédure : La procédure était identique à celle employée avec les sujets américains (même matériel, même programme). Seules les instructions ont été ré-écrites en français (annexe p.148).

RÉSULTATS

Les erreurs (7.3 % des données) ont été remplacées, ainsi que les temps en dehors de l'intervalle 100..1500 msec (2.1 % des données). La table IV.4 fournit les moyennes.

	Cible		
Groupe	Coda	Attaque	Différence
Coda	549 ms (8.8 %)	525 ms (10.9 %)	+24 ms (-2.1 %)
Attaque	616 ms (5.5 %)	570 ms (7.7 %)	+46 ms (-2.2 %)

TAB. IV.4 – *Temps de réaction et taux d'erreurs — Induction attentionnelle, stimuli américains et sujets français*

Les anovas ont été conduites avec les facteurs Groupe d'induction (intra-sujet) et Syllabe (position de la cible : coda ou attaque ; intra-sujet). L'effet principal dû à Syllabe est significatif (« attaque » 35 msec plus rapide que « coda »), mais uniquement dans l'analyse par sujets. Cet effet est significatif seulement pour le groupe attaque, mais il n'y ait pas d'interaction significative Groupe × Syllabe. Il y a également un effet de Groupe significatif dans les analyses par items apparemment dû à une compensation (« trade-off ») entre erreurs et temps de réaction. Des anovas sur les temps de détection sur les items test « bis » ne révèlent aucun effet significatif.

```
Analyse par sujets: PLAN  S14<G2>*S2
G = Groupe d'induction (G1=Coda et G2=Attaque)
S = Position de la cible (S1=Coda et S2=Attaque)
Temps de réaction
S              F(1,26)=  5.01    MSE=3413.45    p=0.0340
G              F(1,26)=  1.34    MSE=32803.2    p=0.2576
G.S            F(1,26)=  0.49    MSE=3413.45    p=0.4901
S/G1           F(1,13)=  0.83    MSE=4837.73    p=0.3789
S/G2           F(1,13)=  7.42    MSE=1989.16    p=0.0174
Erreurs
S              F(1,26)=  1.36    MSE=0.8407    p=0.2541
G              F(1,26)=  2.38    MSE=1.0824    p=0.1350
G.S            F(1,26)=  0.00    MSE=0.8407    p=0.0000
S/G1           F(1,13)=  0.45    MSE=1.2637    p=0.4859
```

```
S/G2           F(1,13)=  1.37    MSE=0.4176    p=0.2628
Analyse par items: PLAN  S13<S2>*G2
Temps de réaction
S              F(1,24)=  1.85    MSE=8588.52    p=0.1864
G              F(1,24)= 11.23    MSE=3642.37    p=0.0027
G.S            F(1,24)=  0.43    MSE=3642.37    p=0.4818
S/G1           F(1,24)=  0.56    MSE=6646.66    p=0.4615
S/G2           F(1,24)=  2.46    MSE=5584.23    p=0.1299
Erreurs
S              F(1,24)=  0.15    MSE=8.4391    p=0.2981
G              F(1,24)=  6.50    MSE=0.4263    p=0.0176
G.S            F(1,24)=  0.00    MSE=0.4263    p=0.0000
S/G1           F(1,24)=  0.13    MSE=4.6474    p=0.2784
S/G2           F(1,24)=  0.15    MSE=4.2179    p=0.2981
```

L'inspection des tables de données des deux expériences (Tab.IV.3 et Tab.IV.4) suggère que les Français ne se comportent pas exactement comme les Américains : seul le groupe français induit en attaque est plus rapide pour les cibles placées dans cette position. Pour comparer les deux expériences, nous avons réuni les deux ensembles de données et effectué des Anovas en déclarant un facteur supplémentaire « Langue des sujets (L) ». Celui-ci produit un effet significatif (les américains sont plus rapides que les français), mais aucune interaction. On ne peut donc pas conclure que les français et les américains ont un comportement différent. Voici ces anovas par sujets :

```
Analyse par sujets: PLAN  S<G2*L2>*S2
G = Groupe d'induction (G1=Coda et G2=Attaque)
S = Position de la cible (S1=Coda et S2=Attaque)
L = Langue (américain ou français)
Temps de réaction
S              F(1,66)=  4.87    MSE=2633.87    p=0.0308
G              F(1,66)=  1.68    MSE=34253.8    p=0.1994
G.S            F(1,66)=  0.11    MSE=2633.87    p=0.2588
L              F(1,66)=  9.48    MSE=34253.8    p=0.0030
L.S            F(1,66)=  2.22    MSE=2633.87    p=0.1410
L.G            F(1,66)=  0.16    MSE=34253.8    p=0.3096
L.G.S          F(1,66)=  1.71    MSE=2633.87    p=0.1955
S/G1           F(1,33)=  2.41    MSE=3499.81    p=0.1301
S/G2           F(1,33)=  4.26    MSE=1767.94    p=0.0470
Erreurs
S              F(1,66)=  0.88    MSE=0.5216    p=0.3516
G              F(1,66)=  0.87    MSE=0.8189    p=0.3544
G.S            F(1,66)=  0.49    MSE=0.5216    p=0.4864
```

L	F(1,66)= 14.54	MSE=0.8189	p=0.0003
L.S	F(1,66)= 1.31	MSE=0.5216	p=0.2565
L.G IE	F(1,66)= 2.33	MSE=0.8189	p=0.1317
L.G.S IE	F(1,66)= 0.33	MSE=0.5216	p=0.4324
S/G1	F(1,33)= 1.03	MSE=0.7489	p=0.3175
S/G2	F(1,33)= 0.29	MSE=0.2944	p=0.4062

DISCUSSION

Le résultat principal de cette expérience est que les sujets français ne sont pas influencés par le biais des listes, comme l'atteste l'absence d'intéraction Groupe $\times$ Syllabe. Ce résultat est identique à celui des sujets américains : les Français étaient simplement un peu plus lents et commettaient plus d'erreurs, ce qui n'est pas surprenant étant donné que les phonèmes anglais ne leur étaient pas familiers.

Une conclusion possible est que les sujets français, ainsi que les américains, percevaient la cible comme appartenant au coda de la première syllabe dans tous les cas. Mais alors, le résultat de Cutler et al. (1986), où les français « syllabifiaient » des stimuli anglais similaires aux nôtres devient difficilement compréhensible. Pourquoi, nos sujets français se fonderaient-ils sur une syllabification « acoustico-phonétique », et ceux de Cutler et al. sur une syllabification « phonologique » respectant les règles de syllabification spécifiques du français ?

À la recherche d'une explication, nous avons examiné les mots tests employés par Cutler et al. (1986). Il nous est apparu que presque tous étaient très similaires à des mots français. Les voici : *balance, balcony, calorie, calculate, galaxie, galvanize, malady, malcontent, palace, palpitate, salad, salvage, talon, talcum*. Les mots tests de l'expérience attentionnelle (cf p.72 ; voir également les mots inducteurs) étaient également similaires à des mots français, mais dans une moindre mesure, et après avoir fini l'expérience, les sujets affirmaient n'avoir identifié pratiquement aucun mot. Nous nous demandons si les sujets français écoutant l'anglais dans l'étude de Cutler et al. pourraient avoir utilisé des représentations syllabiques post-lexicales. L'idéal serait d'avoir les résultats d'une tâche de détection de fragment sur des pseudo-mots anglais par des sujets français.

Une autre interprétation est que les sujets utilisent des représentations différentes selon la tâche. Dans la détection de fragment, pour les Français, la syllabification serait dictée par le principe de l'attaque maximale (p.ex. « ba–lance ») ; alors que dans la détection de phonèmes attentionnelle, la syllabification serait plutôt sensible à l'acoustique-phonétique du signal (i.e. « bal–ance »). Cette idée est cependant difficile à réconcilier avec les données disponibles qui suggèrent que la détection de fragments est trés influencée par l'acoustique du signal (cf l'utilisation de la « transparence acoustique » selon Sebastian-Gallés et al. 1992 ;

ainsi que les études de « cross-splicing » par Rietveld et Frauenfelder, citées par Frauenfelder, 1992).

Finalement, il se pourrait que les sujets français n'aient pas utilisé de représentation syllabique pour effectuer la tâche de détection attentionnelle. La contradiction entre les indices acoustiques et la syllabification « théorique » les a peut-être empêchés de se forger des attentes syllabiques.

En conclusion, le résultat obtenu en français avec le paradigme de détection attentionnelle a été répliqué en espagnol. Par contre, nous n'avons pas observé d'induction syllabique en anglais. Ce résultat peut être considéré comme une confirmation de l'hypothèse de Cutler et al. (1986) selon laquelle les locuteurs anglais n'utilisent pas la structure syllabique des stimuli. Toutefois, l'absence d'effet syllabique quand les sujets français écoutent les stimuli anglais, laisse ouverte la possibilité que la structure syllabique des stimuli ne soit pas telle qu'on le supposait. Il est clairement nécessaire de tester d'autres types de mots en anglais, particulièrement ceux où les intuitions des sujets semblent plus fermes (c'est à dire des mots avec des voyelles tendues (polar/polka), plutôt qu'avec des voyelles lâches (balance/balcony)).

Chapitre V

Accès à un code infra-syllabique

Doit-on attendre la fin d'une syllabe pour accéder à l'information qu'elle contient ? Autrement dit, la syllabe est-elle une étape limitante (un « bottleneck ») dans la perception la parole ? En particulier, les phonèmes sont-ils « récupérés » seulement *après* l'identification de la syllabe à laquelle ils appartiennent ? À l'appui de cette conception, Segui et al. (1990) citent des corrélations entre la durée ou la complexité d'une syllabe et les temps de latences pour détecter un phonème qui la débute. Cependant, l'idée que le système perceptif doive attendre la fin de chaque syllabe pour transmettre l'information vers les niveaux supérieurs a été critiquée. D'une part, Marlsen-Wilson (1984) a présenté des données qui indiquent que les sujets peuvent décider qu'un stimulus est un non-mot avec un temps de décision constant à partir du phonème critique, (c'est à dire celui à partir duquel le stimulus ne peut plus être un mot). Le point important est que les temps de décision ne sont pas affectés par la position syllabique du phonème critique (début ou fin de syllabe). D'autre part, des sujets peuvent apparemment détecter un phonème avant d'atteindre la fin de la syllabe qui le contient (Norris & Cutler, 1988; Dupoux, 1989). Toutefois, ces résultats demeurent compatibles avec un modèle en cascade, où un premier niveau de détecteurs de syllabes fournit en continue des « degrés d'activation » ; la détection de phonème pourrait alors s'effectuer par « conspiration » syllabique (Dupoux, 1993). Une autre possibilité est que les sujets puissent répondre sur la base d'un code demi-syllabique (Dupoux, 1993) : les expériences de Dupoux et de Norris et Cutler utilisaient uniquement des syllabes CVC, et montrent que les sujets peuvent « ignorer » le coda de syllabe. Dans ce chapitre, nous nous demandons si de l'information partielle sur la consonne initiale d'une sylabe CV peut être identifiée avant que la syllabe CV entière soit reconnue.

Pour étudier ce problème, nous nous sommes inspirés d'un article intitulé « Discrete versus Continuous Stage Models of Human Information Processing: In Search of Partial Output » (Miller, 1982). L'auteur propose un paradigme expérimental qui permet de tester si le sujet peut amorcer une réponse motrice sur la base d'une analyse partielle d'un stimulus. Ce paradigme se fonde sur le fait que deux réponses effectuées par le même bras peuvent être « préparées » plus efficacement que deux réponses effectuées par des bras différents. Plus concrètement, si l'on connaît à l'avance le bras de réponse, alors on est plus rapide pour répondre (il y a un « amorçage » de la réponse motrice). Jeff Miller a eu l'idée de regarder si les sujets sont plus rapides quand une information partielle contenue dans un stimulus prédit le bras de réponse plutôt que quand elle ne le prédit pas.

Les sujets devaient classifier quatre stimuli visuels (par exemple 'BE', 'BO', 'ME', 'MO') en utilisant l'index et le majeur de chaque main (un doigt est assigné à chaque stimulus). Dans les blocs expérimentaux, les stimuli qui partagent une information partielle sont attribués à la même main (p.ex.: 'BE' et 'BO', qui partagent la même consonne initiale, sont assignés à la main droite ; 'ME' et 'MO', à la main gauche). Dans les blocs contrôles, les stimuli assignés à chaque main sont complètement différents et ne partagent aucune information partielle (par ex.: 'BE', 'MO' sont assignés à la main droite, 'ME', 'BO' à la main gauche). Si l'information partielle est extraite du stimulus avant que celui-ci soit reconnu entièrement, alors on s'attend éventuellement à observer des temps de réaction plus rapides dans les blocs expérimentaux que dans les blocs contrôles.

Miller (1982) a effectivement observé un effet de préparation du bras pour différents ensembles de stimuli, et en particulier pour les stimuli visuels BE, BO, ME, MO (voir également Jeff Miller , 1983). Il considère cela comme une preuve que l'identité de la première lettre est disponible pour faciliter la réponse avant que le stimulus entier soit identifié. Évidemment, ce résultat pourrait ne pas paraître trop surprenant car les lettres forment des codes nettement distincts, et de plus, sont séparées spatialement. On pourrait également penser que tout type de similarité entre les stimuli assignés à une main peut être utilisé pour faciliter la réponse. C'est précisément parce que ce n'est pas le cas, c'est à dire parce que la similarité ne provoque pas toujours de facilitation, que la méthode de Jeff Miller nous a intéressée.

Dans une de ses expériences (Miller, 1982, exp. 4), les sujets devaient classifier les lettres U, V, M, N. La question était de savoir si la forme globale de la lettre pouvait amorcer la réponse : U et V sont semblables entre-elles, et assez différentes de M et N, qui sont elles-mêmes assez similaires entre-elles. De fait, dans une tâche de comparaison, les temps de réaction sont nettement plus élevés pour discriminer deux lettres de formes globales similaires (p.ex. U et V) plutôt que deux lettres dissimilaires (p.ex. M et U). L'interprétation classique de ce résultat est que, quand les lettres sont globalement différentes, le sujet peut répondre avant d'avoir identifié précisément les lettres, sur la base de la forme globale. Cela suppose que la forme globale est identifiée avant l'identité de la lettre. Si tel était le cas, dans le paradigme de Jeff Miller, on s'attend alors à ce que les sujets soient plus rapides quand des lettres similaires plutôt que dissimilaires sont attribuées à la même main : la forme globale pouvant prédire le bras de réponse. Or cela n'est pas le cas. Jeff Miller interprète l'absence de facilitation en proposant que la forme globale d'une lettre n'est pas accessible au système de réponse avant que la lettre elle-même soit complètement identifiée. [1]

Pour rendre compte du fait qu'on observe une préparation de réponse pour des stimuli comme (BE, BO) vs (ME, MO) mais pas pour (U, V) vs (M, N), Jeff Miller propose « qu'un trait partiel d'un stimulus visuel ne peut servir à amorcer la réponse que s'il peut activer un code discret par lequel l'information peut être transmise » (Miller, 1992, p.293). [2] La notion de « code discret » est fondamentale, voilà plus précisément ce qu'entend Jeff Miller par « code » :

> « It appears that the discrete versus continuous distinction reduces to a debate about the size of the unit of information transmission (i.e. the "grain size" of information transmission). Continuous models state or imply that the grain size is effectively zero, so that any available information immediately begin to prime responses (cf., McClelland, 1979). Discrete models, on the other hand, must claim that information is transmitted discretely with respect to some information-rich internal "codes," perhaps of the type often studied within information processing psychology (e.g. Posner & Taylor, 1969). In these models, the grain size is considerably larger than zero, though it does not necessarily encompass all of the information in the whole stimulus.
>
> « The notion of "code" required here is closely related to that of Posner (1978), who said that "by a *code* I mean a format by which the information is represented" (p.27). The idea is that the system may represent a stimulus in terms of several internal codes (e.g. letter identity and size). Each code could be processed discretely in the sense that no partial information about the code is ever made available to later processes. Yet response preparation could begin as soon as any code was completely activated, without waiting for full recognition of the other relevant codes. This model is refered to as the *asynchronous discrete coding* (ADC) model, since each code is transmitted discretely but different codes may be transmitted at different times » (Miller, 1982, p.292).

1. Il faut trouver une explication alternative au résultat de la tâche de discrimination. On peut proposer que dans la tâche de discrimination, le sujet se focalise sur la forme globale pour répondre rapidement. Chez Miller la forme globale ne détermine pas parfaitement la réponse (il faut encore décider du doigt), et le sujet préfère peut-être alors utiliser un code alphabétique.

2. « Preliminary information about visual characteristics seems to be useful for response preparation only if it can be used to activate a discrete code by which the information can be transmitted »

Jeff Miller a employé exclusivement des stimuli visuels. Mais il nous a semblé que sa méthode était bien adaptée pour déterminer si des codes partiels, en particulier un code phonémique, peuvent être activés quand les sujets doivent reconnaître une syllabe présentée dans la modalité auditive. Dans un modèle de syllabaire où aucune information partielle n'est accessible au système de décision avant que la syllabe ne soit identifiée, on ne devrait pas observer d'effet de préparation de réponse. Si, par contre, le système de décision accède à l'information perceptive de façon continue, ou par « paquets » plus petits que la syllabe elle-même, alors on peut s'attendre à observer un effet de préparation de la réponse. Dans l'expérience qui suit, nous avons demandé à des sujets de classifier quatre syllabes qui partageaient ou non la première consonne (FA, FU, SA, SU). Si celle-ci pouvait être identifiée avant la syllabe, alors on prédit un effet de préparation de réponse.

EXPÉRIENCE 9: FAFU – SASU

Cette expérience est similaire à la « BE BO ME MO » de Jeff Miller décrite plus haut, sauf que les stimuli étaient auditifs (FA FU SA SU). Pour des raisons pratiques (possibilité de tester de nombreux sujets), nous l'avons réalisée en Espagne, dans le laboratoire de N. Sebastian-Gallès à l'Université de Barcelone. [3]

DESCRIPTION

Matériel : Une locutrice espagnole a enregistré deux exemplaires de chacune des syllabes FA, FU, SA et SU. Nous avons ensuite enregistré sur une bande magnétique une liste quasi-aléatoire de 200 stimuli avec un intervalle entre deux stimuli qui était de 5 sec en début de bande, 4.5 sec à partir du 15ième essai, 4 sec à partir du 30ième essai et 3.5 secondes à partir du 60ième essai (le rythme s'accélérerait). La liste était «quasi-aléatoire» car nous avons fait en sorte que deux syllabes successives soit toujours différentes [4]. Sur la deuxième piste de la bande, des clics inaudibles ont été placés au début de chaque stimulus. La bande magnétique a été dupliquée en 5 exemplaires.

---

3. Je tiens à remercier Nuria Sebastian-Gallès de m'avoir accueilli et permis d'utiliser son matériel et « ses » sujets. Je remercie aussi Albert Costa qui m'a aidé dans la réalisation pratique de l'expérience (traduction des instructions, accueil des sujets...)

4. Nous avons beaucoup hésité avant d'imposer cette contrainte, car elle signifie que d'un essai à l'autre, la réponse change de main dans 2/3 des cas (voir procédure). Cela introduit donc un biais qui pourrait être utilisé par les sujets pour prédire le bras dans l'essai suivant. Dans un «pilote» de cette expérience où nous avons employé une «vraie» liste aléatoire, nous avons jugé que les réponses sur les essais où la même syllabe était répétée étaient de nature totalement différente des autres (effet de répétition). C'est pourquoi nous avons opté pour la liste contrainte.

Procédure : Le sujet était assis en face d'un PC (IBM PS/2) relié à un lecteur de cassette (TASCAM Porta One) et écoutait les stimuli à travers un casque. Après qu'il ait lu les instructions affichées sur l'écran de l'ordinateur, un bloc d'essais débutait. A chaque essai (démarré par le clic placé sur la piste 2 de la bande magnétique), le sujet entendait l'une des quatre syllabes FA, FU, SA, ou SU et devait appuyer sur l'une des quatre touches Z, X, N ou M pour classer la syllabe entendue. Le clavier espagnol est semblable au clavier américain : ces touches sont sur la dernière rangée, à gauche pour Z et X, et à droite pour N et M. Le sujet pressait respectivement, Z et X avec le majeur et l'index de la main gauche, N et M avec l'index et le majeur de la main droite. L'appariement des syllabes aux touches dépendait de la condition expérimentale du bloc (voir plus bas). L'ordinateur enregistrait la touche appuyée et le temps de réponse. Les sujets avaient deux secondes pour répondre.

Il y avait deux phases à chaque bloc d'essais : une phase d'apprentissage et une phase de test. Durant la phase d'apprentissage, le sujet voyait affiché *en permanence* sur la première ligne de l'écran, l'assignement des syllabes aux touches. Par exemple :

```
Z    X                                   N    M
FA   FU                                  SA   SU
```

De plus, à chaque essai, le sujet recevait une information sur sa réponse : selon que celle-ci était correcte ou pas, 'Bien' ou 'Mal' s'affichait à l'écran. En cas de mauvaise réponse, la touche correcte (sur laquelle le sujet aurait du appuyer) clignotait à l'écran. Quand la différence entre le nombre de réponses correctes et le nombre de réponses erronées atteignait vingt-cinq, l'ordinateur basculait dans la phase de test. Dans cette phase, l'écran restait vide, sauf quand le sujet commettait une erreur, auquel cas la touche correcte clignotait pendant une seconde. La phase de test, contrairement à la phase d'apprentissage, avait une longueur fixe de 100 essais. Le bloc finissait avec cette phase. Tous les sujets ont atteint le critère d'apprentissage avant le centième essai, ce qui fait qu'aucun n'a atteint la fin de la liste de 200 stimuli durant la phase de test.

Nous avons employé un dessin intra-sujet dans lequel chaque sujet passait deux blocs : un bloc expérimental (avec, par exemple, FA et FU assignés à une main et SA et SU à une autre) et un bloc contrôle (avec, par exemple, FA et SU assignés à une main et SA et FU à l'autre). La même bande magnétique était employée pour tous les blocs, seul changeait l'assignement des touches. Les sujets se reposaient environ 5 minutes entre les deux blocs et l'expérience totale durait environ 25 minutes. L'ordre des blocs était contrebalancé, de telle façon que la moitié des sujets débutait par le bloc contrôle et l'autre moitié par le bloc expérimental. Finalement, cinq appariements syllabes–touches différents ont été employés pour faire varier les mains et les doigts associés à une syllabe particulière. Les sujets étaient testés par groupe de cinq et pour des raisons matérielles il était plus facile de n'avoir que cinq appariements. Les positions des syllabes n'étaient donc pas parfaitement contrebalancées. Les cinq appariements utilisés sont affichés dans la table V.1.

Sujets : 30 étudiants de l'université de Barcelone, de langue maternelle espagnole et, pour la plupart, bilingues espagnol/catalan, ont participé à cette expérience en échange de points nécessaires à la validation de leur cursus. Ils étaient testés par groupes de quatre ou

	Coté de Réponse (Main)				
	Gauche			Droit	
Bloc	Majeur	Index	—	Index	Majeur
Expérimental	fa	fu	—	sa	su
Contrôle	fu	sa	—	su	fa
Expérimental	fu	fa	—	su	sa
Contrôle	fa	su	—	sa	fu
Expérimental	sa	su	—	fa	fu
Contrôle	su	fa	—	fu	sa
Expérimental	su	sa	—	fu	fa
Contrôle	sa	fu	—	fa	su
Expérimental	fu	fa	—	sa	su
Contrôle	fa	su	—	fu	sa

TAB. V.1 – *Assignements stimulus–doigt de réponse*

cinq, chacun dans un box isolé. Deux sujets supplémentaires ont été éliminés pour avoir fait plus de 20 % d'erreurs.

RÉSULTATS

Nous avons effectué deux Analyses de Variance (a) sur les temps de décision moyens et (b) sur les taux d'erreurs (appuis sur une mauvaise touche). Deux facteurs étaient déclarés : la Condition : bloc expérimental ou bloc contrôle (intra-sujets), et l'Ordre de passation des blocs : contrôle puis expérimental ou l'inverse (facteur entre-sujets). La table V.2 fournit les résultats moyens (sans distinguer l'« Ordre »).

	Contrôle fa su – fu sa	Expérimental fa fu –sa su	Différence
TR Moyen	807 ms	754 ms	53 ms
Erreurs	9.6 %	6.9 %	2.7 %

Pour les temps de réaction, chaque cellule est la moyenne d'environ 3000 données (100 RT × 30 sujets moins les erreurs). L'écart-type des temps de réaction est de 105 msec entre sujets, et la moyenne des écart-types intra-bloc et intra-sujet est de 261 msec.

TAB. V.2 – *Résultats FaFu SaSu*

Voici les résultats des Anovas :

```
PLAN  S15<O2>*C2
C = Condition (Bloc Expérimental vs Contrôle)
O = Ordre (1er bloc Expérimental ou Contrôle)
Temps de réaction
C              F(1,28)= 10.48    MSE=3979.78    p=0.0031
O              F(1,28)=  0.55    MSE=22568.6    p=0.4645
O.C            F(1,28)=  0.92    MSE=3979.78    p=0.3457
Erreurs
C              F(1,28)= 11.27    MSE=9.2262     p=0.0023
O              F(1,28)=  1.71    MSE=23.383     p=0.2016
O.C            F(1,28)=  0.02    MSE=9.2262     p=0.1115
```

Dans l'analyse des erreurs comme dans l'analyse des temps de réaction, seul le facteur Condition produit un effet significatif. Il n'interagit pas avec le facteur Ordre. En sus de ces Anovas, nous avons examiné le nombre d'essais de la phase d'entraînement pour déterminer si l'appariement était plus difficile à apprendre dans les blocs expérimentaux que contrôles : en fait les nombres d'essais moyens sont de 30.8 et 30.7 respectivement et cette différence est non significative dans un t-test.

DISCUSSION

Les résultats sont clairs : quand la consonne initiale prédit le bras de réponse, les sujets sont plus rapides et commettent moins d'erreurs que lorsque les syllabes assignées à la même main ne partage aucun phonème. Il semble donc qu'on doive conclure que les sujets ont eu accès à un code sub-syllabique pour amorcer leur réponse.

Si l'on examine une représentation de l'acoustique des stimuli, on observe deux segments successifs : un premier qui contient un « bruit » correspondant à la consonne fricative, et un second, périodique, qui spécifie la voyelle. Le bruit fricatif est très similaire à l'intérieur des couples (fa,fu) et (sa, su).[5] L'interprétation « à la Jeff Miller » est que les sujets ont pu utiliser ce premier segment acoustique pour préparer le bras de réponse, en activant un code sub-syllabique. Notons que ce code sub-syllabique n'est pas nécessairement un phonème, ce pourrait tout aussi bien être un trait phonétique.

5. Il contient aussi de la coarticulation due à la voyelle suivante. Avec un éditeur de son, qui permet d'écouter n'importe quel segment de signal, il apparaît que la voyelle peut être « devinée » quelques dizaines de msec avant la partie voisée.

Une remarque, cependant, vient tempérer cette interprétation : la tâche est assez difficile puisque les réponses sont effectuées assez largement après la fin des stimulus (dont la durée moyenne était de 410 msec pour un temps de réaction moyen de 800 msec)[6]. On peut donc se demander si l'interprétation selon laquelle la différence de temps de réaction reflète un traitement perceptif précoce de la consonne, est correcte. Il semble possible que la syllabe soit d'abord identifiée globalement, et que le code phonémique ne soit récupéré qu'après coup. L'effet serait alors entièrement dû à l'étape décisionnelle de « branchement » de la réponse : le code phonémique « faciliterait » la réponse plutôt qu'il ne la « préparerait ». Nous référerons à cette deuxième hypothèse comme « hypothèse décisionnelle ».[7]

Différencier une activation d'un code phonémique « pré-syllabique » et « post-syllabique » est très délicat. Mais si l'effet de préparation a pour origine une identification « en temps réel » de la première consonne, alors il devrait être influencé par la distribution temporelle de l'information acoustique dans le signal. En particulier, on prédit qu'il ne devrait pas y avoir d'effet d'amorçage quand cette information vient seulement quand la syllabe peut déjà être identifiée. Une telle situation peut être créee en assignant à chaque main des syllabes partageant *la même voyelle*. En effet quand le sujet atteint l'information vocalique, la syllabe, aussi bien que la voyelle peut être identifiée ; il n'y alors pas de raison d'observer un effet de préparation (car il est « trop tard » pour se préparer). Par contre, dans l'hypothèse d'une facilitation « post-syllabique », la distribution temporelle de la consonne ou de la voyelle importe peu : on s'attend à ce que le code spécifiant la voyelle puisse « faciliter » la réponse tout autant que le code consonantique le faisait dans cette expérience.

EXPÉRIENCE 10: FASA – FUSU

On a réalisé une expérience similaire à la précédente, mais cette fois-ci, dans le bloc expérimental, les syllabes partageaient la voyelle et non la consonne (fa,sa – fu,su). Si l'effet de préparation est dû à une activation en « temps réel » du code phonémique (hypothèse « perceptive »), alors la distribution temporelle de l'information dans le signal détermine l'importance de la préparation, et les réponses ne devraient pas être amorçées par l'information vocalique (cf discussion précédente). Si, au contraire, l'interprétation « décisionnelle » est correcte, alors

6. Le sujet le plus rapide a un temps de réaction moyen de 510 msec. A titre de comparaison, les temps les plus rapides dans la classification de syllabe (expérience 1) étaient de l'ordre de 250 msec.

7. Proctor et Reeve (1986) ont critiqué l'interprétation de « l'effet Miller » en terme de « préparation » motrice. Mais voir, par exemple, Miller et Dexter (1988) et Miller, Riehle, et Requin (1992) pour des réponses et des développements intéressants que nous ne pouvons détailler ici.

la voyelle devrait « faciliter » la réponse tout autant que la consonne dans l'expérience précédente.

DESCRIPTION

La même bande magnétique que précédemment a été employée. La procédure était identique à celle de l'expérience précedente. Seul l'assignement des touches à été modifié, en échangeant fa et su. Ainsi, dans les blocs expérimentaux, les syllabes associées à la même main avaient la même voyelle (p.ex. FA SA – FU SU). Les sujets, au nombre de 30, n'avaient pas participé à l'expérience précédente. Cinq ont été éliminés et remplacés pour avoir fait plus de 20 % d'erreurs.[8]

RÉSULTATS

Les données, résumées dans la table V.3, ont été analysées comme dans l'expérience précédente.

	Contrôle fa su – fu sa	Expérimental fa sa – fu su	Différence
TR Moyen	842 ms	771 ms	71 ms
Erreurs	10.1 %	6.1 %	4 %

Pour les temps de réaction, chaque cellule est la moyenne d'environ 3000 données (100 RT × 30 sujets moins les erreurs). L'écart-type des temps de réaction est de 110 msec entre sujets, et de 277 msec intra-sujet et intra-condition.

TAB. V.3 – *Résultats fAsA fUsU*

```
PLAN  S15<O2>*C2
C = Condition (Bloc Expérimental vs Contrôle)
O = Ordre (1er bloc Expérimental ou Contrôle)
Temps de Réaction
C              F(1,28)= 17.04    MSE=4362.36    p=0.0003
O              F(1,28)=  1.22    MSE=24340.5    p=0.2788
O.C            F(1,28)=  1.02    MSE=4362.36    p=0.3212
Erreurs
C              F(1,28)= 22.44    MSE=10.3405    p=0.0001
O              F(1,28)=  3.12    MSE=27.6786    p=0.0882
```

8. Chez tous les sujets remplacés, dans cette expérience comme dans la précédente, le taux d'erreur le plus élevé était dans le bloc contrôle. Dans les taux donnés dans les tableaux, la différence entre les erreurs dans les deux conditions est donc plutôt sous-estimée.

```
O.C            F(1,28)=  0.23    MSE=10.3405    p=0.3648
```

L'effet de Condition est significatif sur les erreurs ainsi que sur les temps de réaction. Le facteur d'Ordre n'interagit pas avec la Condition. D'autre part, les nombres d'essais d'entraînement étaient de 32.4 dans les blocs expérimentaux et de 29.4 dans les blocs contrôles, mais cette différence n'était pas significative ($t(29) = 1.6; p = .12$).

DISCUSSION

Quand la voyelle prédit le bras de réponse, les sujets sont plus rapides et commettent moins d'erreurs. La facilitation est même supérieure à celle observée quand la consonne prédisait le bras de réponse (cette différence n'est toutefois pas significative)[9]. Selon la discussion de l'expérience précédente, cela suggère que l'amorçage n'est pas dû à la prise en compte *en continue* de l'information disponible dans le signal mais plutôt à l'étape décisionnelle qui, après que le stimulus ait été reconnu, effectue le branchement stimulus-réponse.

On peut se convaincre du caractère tardif de l'effet en traçant les distributions des temps de réaction bruts en fonction du type de bloc (expérimental ou contrôle). Sur les figures V.1 et V.2, on a affiché, à droite : les distributions cumulées de probabilité de réponses, et à gauche : la distance entre la distribution contrôle et la distribution expérimentale (selon une méthode de visualisation due à E. Dupoux). Le trait plein représente la distribution des réponses dans les blocs expérimentaux, le trait pointillé correspond aux blocs contrôles.

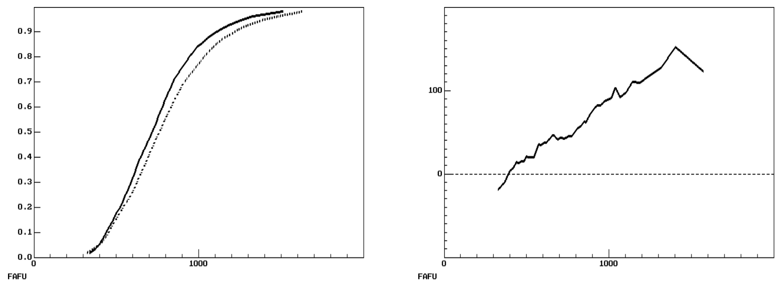
Comme on peut le constater la taille de l'effet *augmente* avec le temps de décision[10] Il devient dès lors paradoxal de considérer qu'un tel effet reflète une « préparation ».

Nature des erreurs : Nous avons décidé d'analyser les erreurs plus en détail, dans cette expérience, ainsi que dans la précédente.[11] Plus particulièrement, nous voulions déterminer si les erreurs sont plutôt de nature motrice (le sujet confondant

9. Pour évaluer cela, nous avons réalisé des Anovas en fusionnant les deux expériences et en déclarant un facteur supplémentaire "Expérience". Ce facteur ne produit d'interaction dans aucune analyse. (L'effet du facteur Condition demeure évidemment significatif : F(1,56)=27 pour les temps de réaction, et F(1,56)=33 pour les erreurs). Il n'y a donc aucune différence significative entre cette expérience et la précédente, du moins sur l'analyse des réponses correctes.

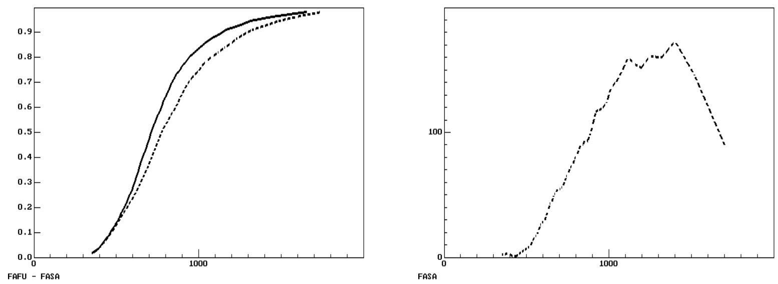
10. Ceci peut être contrasté avec l'effet en classification de syllabe, où le ralentissement dû à la variabilité apparaissait dès les temps de réaction les plus rapides.

11. Jeff Miller (1982) ne fournit aucune analyse des erreurs.



Distributions Cumulées Taille de l'effet

FIG. V.1 – *Distribution des TR. Exp. FaFu*



Distributions Cumulées Taille de l'effet

En abscisse : Temps de réaction; Dans les distributions cumulées, le trait pointillé correspond aux blocs contrôles; le trait plein aux blocs expérimentaux.

FIG. V.2 – *Distribution des TR. Exp. FaSa*

les doigts de la même main) ou plutôt de nature perceptive (il confondrait deux stimuli similaires). Dans la table V.4, nous indiquons le nombre d'erreurs en fonction de la similarité entre la réponse et le stimulus présenté, dans les deux expériences.

Dans la condition où les stimuli attribués à la même main partagent la consonne (expérience FaFu, condition expérimentale), les erreurs proviennent essentiellement du stimulus qui partage la consonne (colonne 'C'. p.ex., en entendant /fu/, le sujet «appuie» sur /fa/). Dans la condition où la voyelle signale la main (exp. fAsA, cond. expér.), les erreurs proviennent essentiellement du stimulus qui partage la voyelle (colonne 'V'. par ex. en entendant /fa/, le sujet appuie sur /sa/).[12] Par conséquent, dans les blocs Tests des deux expériences, la confusion est plus importante entre les stimuli qui sont assignés à la même main. Cela est-il un effet « de main » ou bien un effet de similarité ? Les conditions contrôles nous ap-

12. Chacun de ces effet est significatif. L'interaction « Expérience × Similarité » restreinte aux bloc expérimentaux l'est également.

Expérience	Condition						
		Expérimentale			Contrôle		
	Similarité[a]	∅	V	C	∅	V	C
FaFu		0.5	1.6	$\boxed{4.2}$	1.7	1.5	$\boxed{5.6}$
fAsA		0.2	$\boxed{3.0}$	1.7	0.8	1.6	$\boxed{5.7}$

[a] ∅ = aucun phonème partagé (ex: fa,su) ; V = voyelle partagée (ex: fa, sa) ; C = consonne partagée (ex: fa,fu).

[b] On a entouré les taux d'erreurs maximaux dans chaque condition.

TAB. V.4 – *Confusions (taux erreurs) en fonction de la Similarité entre le stimulus présenté et celui détecté, et de la Condition Expérimentale.*

portent la réponse : quand les stimuli attribués à la même main ne partagent aucun phonème, il y a très peu d'erreurs sur la même main (colonne ∅). Les erreurs ne proviennent donc vraisemblablement pas d'une confusion entre les doigts de la même main.

Un fait remarquable qui ressort de cette analyse est que, dans les situations contrôles, les stimuli confondus sont essentiellement ceux qui partagent la même consonne plutôt que ceux qui partagent la voyelle. Il y a donc, contrairement à ce que pourrait laisser penser la seule analyse des réponses correctes, une asymétrie entre les consonnes et les voyelles. Cela suggère que, dans les blocs contrôles, les erreurs puissent être des anticipations de réponse fondées sur la consonne, ce qui serait en contradiction avec notre interprétation décisionnelle et ferait pencher la balance en faveur de l'accès à de l'information partielle. Or, si l'on examine les temps de réaction sur les erreurs, on s'aperçoit qu'ils sont systématiquement *plus lents* que les temps de bonne réponse (quelle que soit la condition, Expérimentale ou Contrôle, et la similarité). Les erreurs ne sont donc pas des réponses particulièrement rapides. L'hypothèse qu'elles proviennent d'anticipation sur la base d'une information partielle ne semble pas confirmée.

En résumé, les erreurs ne sont pas dues à une confusion entre les doigts d'une main. Elles ne sont pas distribuées uniformément car il y a nettement moins d'erreurs entre des stimuli diffèrant à la fois par la voyelle et par la consonne, qu'entre ceux partageant un de ces deux phonèmes. Les confusions dépendent donc de la similarité entre les stimuli. Toutefois, cette similarité dépend de la condition expérimentale : en fonction de celle-ci l'attention semble être focalisée sur la consonne ou sur la voyelle. Dans le bloc expérimental de l'expérience « FaFu », les sujets confondent plus les stimuli partageant la consonne, par exemple /fa/ et /fu/ plutôt que /sa/ et /su/. Dans le bloc expérimental de « FaSa », c'est le contraire. Par défaut, dans les blocs contrôles, l'attention semble focalisée sur la consonne.

DISCUSSION GÉNÉRALE DES EXPÉRIENCES 9 ET 10

Nous avions choisi le paradigme de Jeff Miller car il semblait pouvoir révéler l'activation de codes transitoires issus de l'analyse partielle d'un stimulus. Plus précisement, dans notre cas, la question était de savoir si un code phonémique pouvait être identifié avant que la syllabe entière soit reconnue. Or, nous avons trouvé que l'effet d'amorçage moteur, indice selon Jeff Miller de l'activation précoce d'un code partiel, était également présent si l'information signalant ce code permettait également d'achever l'identification de la syllabe. Notre conclusion est que l'effet d'amorçage est en fait une facilitation décisionelle, mais ne reflète pas le décours temporel (« time-course ») des activations des codes phonèmiques et syllabiques.

Même si l'on rélègue l'effet de facilitation aux étapes décisionnelles, il faut néanmoins souligner qu'il implique l'existence d'un code infra-syllabique (structurellement sinon temporellement) ou, du moins, d'une relation de similarité entre syllabes qui fait que /fa/ et /fu/ sont plus proches entre-elles que /fa/ et /su/. À vrai dire, cela pourait ne paraître trop surprenant. En effet, il existe un code « phonémique » tout trouvé qui est le code *orthographique*. Il y a peu de doute que nos sujets, étudiants de psychologie de première année, pouvaient parfaitement se représenter les quatre syllabes en utilisant les *lettres* : F, S, A , U. Il est plausible qu'après avoir perçu la syllabe, les sujets pouvaient employer une représentation *orthographique* de celle-ci, pour effectuer leur réponse. L'intérêt, pour eux, serait que l'« algorithme » de branchement stimulus–doigt est plus facile à apprendre, retenir et/ou exécuter à partir du code orthographique qu'à partir d'un code syllabique. Bien sûr, on ne peut distinguer cette explication, d'une autre en terme de représentation *phonémique*.

Le rôle possible du code orthographique dans les tâches de psycholinguistique ne doit pas être sous-estimé : ainsi, dans une tâche de jugement de similarité, des enfants de 5 ans (pré-scolaires), jugent plus similaires des syllabes globalement semblables (/vis/ et /bez/) [13] que des syllabes qui ne partagent que le premier phonème (/bez/ et /bug/) ; par contre, des adultes et des enfants sachant écrire, trouvent plus similaires les syllabes qui partagent le même phonème initial (Treiman & Breaux, 1982). [14] Ce résultat peut être comparé à ce qu'a révélé notre analyse des erreurs : dans les blocs contrôles, les syllabes les plus confondues possédaient précisement la même consonne initiale.

Il serait passionnant de savoir si des sujets ne connaisant pas l'écriture alphabétique (enfants, illettrés, ou bien utilisateurs d'un système non alphabétique)

13. Chaque phonème ne diffèrent de son vis-à-vis que par un trait distinctif.

14. Pour d'autres résultats intéressants obtenus avec la tâche de jugement de similarité, cf Walley, Smith, et Jusczyk (1986) et Chodorow et Manning (1983).

montrent le même effet de facilitation. Nous n'avons malheureusement pas accès à de telles populations. Pour éliminer le recours au code orthographique, une autre possibilité est d'utiliser une relation de similarité qui ne transparaisse pas dans celui-ci ; c'est à dire d'utiliser des phonèmes similaires mais qui ne sont pas distingués nettement dans le système orthographique. Si l'on observe une facilitation, alors on saura qu'elle ne provient pas du code orthographique.

EXPÉRIENCE 11: PA TA – BA DA

Nous désirons exploiter une relation de similarité qui ne soit pas reflétée par un code orthographique, et même, si possible, par aucun code *conscient*. Nous avons décidé d'employer les trait phonétiques de voisement et de place, en utilisant des syllabes se différenciant par des plosives initiales: /pa/, /ta/, /ba/ et /da/. Nous comparons les cas où les syllabes assignées à la même main partagent (a) le trait de voisement (pa ta — ba da) (b) le trait de place d'articulation (pa ba — ta da) (c) aucun trait (ba ta — pa da). Il nous semble que la plupart des sujets n'ont pas conscience des traits distinctifs de voisement et de place. Par conséquent, si la facilitation observée dans les expériences précédentes dépend de l'existence d'un code conscient (en particulier orthographique)[15], on ne s'attend pas à l'observer ici.

DESCRIPTION

Matériel : Un locuteur espagnol a enregistré un exemplaire de chaque syllabe /pa/, /ta/, /ba/, /da/. De façon similaire à l'expérience 9, on a construit une liste quasi-aléatoire formée de 200 stimuli que l'on a enregistrés sur une bande magnétique.

Procédure : L'équipement et la procédure employés étaient identiques à ceux des expériences précédentes. Il a suffit de modifier les instructions en remplaçant /fa/, /fu/, /sa/, /su/ par /pa/, /ta/, /ba/ et /da/ respectivement. Il y avait 4 conditions (cf tab. V.5) définies deux facteurs : (a) le trait (dit « principal ») assigné à une main dans les blocs expérimentaux : Voisement ou Place d'articulation et (b) la congruence de l'index (le trait « secondaire » définit le type de doigt (index vs majeur) ou non).[16]

Chaque sujet passait dans l'une de ces conditions. La moitié d'entre eux débutaient par le bloc contrôle et l'autre moitié par le bloc expérimental. Il y avait donc 8 groupes de sujets.

---

15. Par code conscient, nous entendons « un code explicitement manipulable ».

16. Ce facteur, non déclaré dans les expériences précédentes, sert à tester l'hypothèse d'une facilitation de la réponse quand un trait définit le doigt (index vs majeur).

Trait	Index	Expérimental	Contrôle
Voisement	congr.	pa ta — da ba	ta ba — pa da
	non congr.	pa ta — ba da	ta ba — da pa
Place	congr.	pa ba — da ta	ta ba — da pa
	non congr.	pa ba — ta da	ba ta — da pa

TAB. V.5 – *Assignements*

Sujets : Soixante-quatre sujets espagnols de l'université de Barcelone ont été testés, dans des conditions identiques à celles des expériences précédentes. Ces sujets étaient des étudiants de première année de psychologie qui n'avaient pas suivi de cours de phonétique. On a éliminé et remplacé 13 sujets supplémentaires qui avaient commis plus de 30 % d'erreurs dans un bloc.

RÉSULTATS

		Contrôle	Expérimental	Différence
Voisement	TR Moyen	737 ms	707 ms	30 ms
	Erreurs	15.4 %	12.6 %	2.8 %
Place	TR Moyen	725 ms	712 ms	13 ms
	Erreurs	15.1 %	11.8 %	3.3 %

Pour les temps de réaction, chaque cellule est la moyenne d'environ 3200 données (100 RT × 32 sujets moins les erreurs). L'écart-type des temps de réaction est de 59 msec entre sujets, et de 235 msec intra-sujet et intra-bloc.

TAB. V.6 – *Erreurs et Temps de Réaction en classification de pa, ta, ba, da*

Les temps de réaction moyens et les taux d'erreurs par sujets (cf tab.V.6) ont été soumis à des ANOVAs à quatre facteurs : un facteur intra-sujet : Condition (Expérimental vs Contrôle) et trois facteurs entre-sujets : Trait (Voisement vs Place), Index (Congruent ou Non-Congruent), Ordre (Nature du premier bloc : Expérimental ou Contrôle). Ces Anova étant trop longues pour figurer ici, elles sont détaillées en annexe (cf p.148). En résumé, il y a seulement deux effets significatifs : (a) la Condition, les sujets étant 22 msec plus rapides et faisant moins d'erreurs dans les blocs tests que dans les blocs contrôles (TR : $F(1,56)=6.48$, $p=.01$; erreurs : $F(1,56)=7.61$, $p=.008$) et (b) une interaction Condition × Ordre (TR : $F(1,56)=8.77$, $p=.005$; mais sur les erreurs $F(1,56)=1$). Rapellons que le facteur Ordre désigne le type du premier bloc (Expérimental ou Contrôle) ; cette interaction reflète en fait une accélération de 25 msec entre le premier et le second bloc dont le résultat est que la différence Expérimental - Contrôle qui est de 48 msec chez les sujets qui commencent par le bloc expérimental, devient nulle (-3.8 msec) chez ceux qui commencent par le bloc contrôle. Dans l'analyse canonique, les facteurs Trait et Index, ne produisent aucun effet ou interaction significative. On a tout de même conduit des analyses restreintes à chaque type de trait. Pour la Place d'articulation, l'effet de Condition était significatif sur les erreurs mais pas sur les temps de réaction (TR: 13 ms, $F(1,28)=1.37$, $p=.25$; erreurs: 3.3 %, $F(1,28)=5.19$, $p=.03$). Pour le voisement, c'était l'inverse : (TR: 30 ms, $F(1,28)=5.58$, $p=.03$; erreurs: 2.8 %, $F(1,28)=2.75$, $p=.11$).

Discussion

Globalement, les sujets sont plus lents et commettent plus d'erreurs quand ils ne peuvent pas prédire la main de réponse à partir d'un trait phonétique de la première consonne. L'effet de condition apparaît sur les erreurs pour la place d'articulation et sur les temps de réaction pour le voisement, mais l'absence d'interaction ne permet pas d'affirmer qu'il y a une différence entre les deux types de trait.

Ces résultats suggèrent que l'effet de facilitation observé avec cette tâche ne doit pas son existence à un code conscient. Toutefois, il faut noter que l'effet est nettement plus faible dans cette expérience que dans les deux précédentes. On ne peut donc exclure la possibilité qu'un code conscient, comme l'orthographe, *accentue* l'effet de facilitation.

Discussion Générale

Dans les expériences présentées dans ce chapitre, les sujets devaient classer quatre syllabes avec quatre doigts : l'index et le majeur des mains gauche et droite. Si la syllabe était « l'unité de perception primaire », on aurait pu s'attendre à ce que les sujets effectuent un appariement direct entre le niveau syllabique et les doigts de réponse. Or, l'expérience 9 (« fafu ») a révélé que les réponses étaient plus faciles quand les stimuli assignés au même bras débutaient par la même consonne, plutôt que quand ils ne partageaient aucun phonème. Les sujets n'ont donc pas fondé leurs réponses sur un niveau purement syllabique, qui produirait un code quand une syllabe est identifiée. Ceci est d'autant plus remarquable que la tâche ne requérait explicitement que la manipulation de syllabes.

Si l'on accepte l'interprétation de Jeff Miller (1982), le résultat de l'expérience 9 montrerait que la consonne est identifiée avant la syllabe, et servirait à *préparer* la réponse motrice. Toutefois, étant donné que les temps de décision dépassent largement la durée des syllabes, ce résultat est également compatible avec une interprétation selon laquelle le code phonémique est récupéré après l'identification de la syllabe.

L'expérience 10 (« fasa ») suggère que cette seconde explication est probablement la plus correcte : on observe la même facilitation quand les syllabes attribuées à la même main partagent la voyelle. Or quand l'information vocalique est atteinte, la syllabe est reconnue, et donc il faut conclure que la voyelle « facilite » la réponse plutôt qu'elle ne la « prépare ». Cette interprétation est renforcée par une analyse temporelle qui revèle que l'effet est présent essentiellement aux temps de réaction lents.

Nous nous sommes ensuite demandés si le code facilitateur pouvait être le code orthographique. L'expérience 11 (« pata »), montre que les sujets peuvent exploiter une similarité en termes de traits phonétiques de voisement et de place d'articulation pour faciliter leur réponse. Cela montre qu'un code non-orthographique, et même non-conscient (i.e. non manipulable explicitement), peut faciliter la réponse.

Dans les chapitres précédents, nous avions trouvé des effets syllabiques dans des tâches où les sujets devaient manipuler des phonèmes, ici nous observons des effets « sub-syllabiques » dans une tâche où les sujets n'ont *a priori* qu'à manipuler des syllabes. Cette remarque prendra de l'importance dans le chapitre de conclusion. Dans le prochain et dernier chapitre expérimental, nous abandonnons les paradigmes de détection/classification de stimuli linguistiques (phonème ou syllabe) pour étudier l'influence des frontières syllabiques sur la détection de click.

Chapitre VI

Détection de clicks et Frontière Syllabique

Dans les expériences des chapitres précédents, les sujets devaient détecter des syllabes ou des phonèmes à l'intérieur de stimuli auditifs. Ces tâches mobilisaient leurs capacités *métalinguistiques* (ou plus précisément « métaphonologiques »). Ainsi, des sujets analphabètes auraient probablement les plus grandes difficultés à effectuer ces tâches (Morais et al., 1979; Read, Yun-Fei, Hong-Yin, & Bao-Qing, 1986).[1] Cela ne signifie pas nécessairement que les effets observés n'ont rien à voir avec les processus engagés dans la perception de la parole : ainsi, les temps de réaction sont étroitement liés aux caractéristiques acoustiques des stimuli. Toutefois, il est clairement désirable de compléter ces tâches par d'autres qui font moins appel à l'« introspection » des sujets (Morais & Kolinsky, 1994). Dans ce chapitre, nous allons utiliser la technique de détection de clicks (Abrams & Bever, 1969; Holmes & Forster, 1970)[2] pour tester l'hypothèse selon laquelle la syllabe est une unité perceptive. Avant de présenter ce paradigme, nous allons détailler plusieurs autres techniques qui évitent également le recours à la « métaphonologie ».

Un premier exemple est le paradigme d'alternance. Dans le but d'étudier le temps de « déplacement » de l'attention, Cherry et Taylor (1954) ont présenté à des sujets un message linguistique en alternance dans les deux oreilles. Ils mesuraient l'intelligibilité en fonction de la fréquence des changements d'oreille. Leur idée était que l'intelligibilité devait être minimale précisément quand l'alternance se produit à une fréquence telle que les changements d'oreille soient en opposition de phase avec le déplacement de l'attention. Ils ont effectivement observé un tel minimum de performance pour une fréquence d'alternance aux environs de 4 Hz. Cependant, Huggins (1964) a réinterprété ce résultat en remarquant que cette

1. Bien que la détection/classification de syllabes serait certainement plus facile que la détection de phonèmes (Morais et al., 1989).

2. Un click est un petit bruit surimposé sur le message linguistique. Les sujets doivent détecter des clicks placés à différents endroits du message, et les variations du temps de latence sont interprétées comme reflétant le traitement linguistique. Une description plus détaillée est fournie plus bas.

fréquence correspondait typiquement au débit *syllabique*. Selon lui, la perte d'intelligibilité serait due à l'interruption (maximale à 4 Hz) des unités perceptives que seraient les syllabes. A l'appui de cette hypothèse, Huggins (1964) montre que lorsqu'on accélère le débit syllabique de 20 %, on observe un déplacement concomitant de 20 % du minimum d'intelligibilité. Cela prouve que l'alternance affecte bien le traitement d'informations contenues dans le signal, mais ce résultat est-il vraiment du à l'interruption d'unités syllabiques ? Samuel (1991) a effectué un test plus direct de cette hypothèse en comparant deux situations : dans l'une, le changement d'oreille respecte les frontières syllabiques (donc les syllabes ne sont pas interrompues) ; dans l'autre le changement d'oreille a lieu au milieu des syllabes. Ses résultats sont extrêmement clairs : il n'y a aucune différence entre les deux situations. De plus, Samuel (1991) trouve également un effet similaire à celui d'Huggins avec des mélodies musicales (et une tâche de comparaison AX). Il interprète ces résultats comme provoqués par des processus de séparation des sources sonores (Bregman, 1978).

Un deuxième paradigme est celui du masquage rétrograde : quand on présente deux stimuli auditifs à la suite l'un de l'autre, il est difficile de faire une tâche acoustique fine sur le premier si le second débute à moins de 250 msec de la fin du premier. L'interprétation est que le second remplace le premier dans une mémoire échoïque. Remarquant que la durée de 250 msec est celle d'une syllabe typique, Massaro (1972,1974) a proposé, sur la base de ces résultats, que la syllabe était l'unité de perception de la parole. Notons toutefois que l'argument est très indirect. Le fait que la fenêtre d'analyse du signal soit de la taille de la syllabe n'implique certainement pas que le signal soit *segmenté* en syllabes. Il n'y a aucune preuve du rôle des frontières syllabiques : la mémoire échoïque pourrait tout aussi bien contenir des anti-syllabes (signal entre voyelles successives). Ou bien, on pourrait imaginer des détecteurs de phonèmes indépendants utilisant une fenêtre d'intégration des indices acoustiques aussi large que 250 msec.

Finalement, Kolinsky, Morais, et Cluytens (1994) présentent une tentative plus récente qui évite l'emploi d'une tâche métaphonologique.[3] Ils utilisent le phénomène de migration illusoire entre des sous-unités de mots présentés dichotiquement. Par exemple, si l'on présente au sujet les deux stimuli /kojou/ et /biton/, dans l'oreille droite et dans l'oreille gauche respectivement, il peut avoir l'illusion d'entendre /bijou/ et /coton/. Dans ce cas, il a « échangé » les deux syllabes /ko/ et /bi/. Quelquefois, la migration peut s'interpréter comme un échange du premier

3. Dans cette tâche les sujets doivent détecter des mots, il s'agit donc d'une tâche méta*linguistique*, qui n'est pourtant pas nécessairement méta*phonologique*. À la différence de la détection de phonème ou de syllabe, les cibles à détecter ont un sens, et les sujets pourraient « utiliser » le sens plutôt que la forme pour répondre. Quoiqu'il en soit, cette tâche présente le grand intérêt de pouvoir être effectuée par des analphabètes.

phonème, ou encore d'un trait phonétique attaché au premier phonème. Kolinsky et al. (1994) observent que la syllabe « migre » relativement plus souvent que des unités plus petites chez des locuteurs français. Chez des locuteurs portugais, par contre, il y a une prépondérance des migrations du phonème initial, ce qui va à l'encontre des résultats observés avec la détection de syllabe (Morais et al., 1989). Le problème de ce paradigme, où la réponse doit être un mot, est que l'origine des migrations (c'est à dire le niveau de traitement qui les provoque) est difficile à cerner, les illusions ne provenant pas nécessairement des premières étapes de traitement (voir, p.ex., l'effet « McGurk »).

L'expérience qui suit évite l'emploi d'une tâche métalinguistique. Nous avons décidé d'utiliser la détection de click, c'est à dire d'un petit bruit, surimposé sur un message linguistique. Dans les premières études, les sujets devaient simplement *localiser* le click, c'est à dire décrire où il se trouvait dans le message après avoir entendu celui-ci jusqu'à la fin. Ainsi, Ladefoged et Broadbent (1960) ont observé que les sujets faisaient des erreurs de localisation assez systématiques : ils rapportaient typiquement que le click était placé une ou deux syllabes avant sa position réelle. Un peu plus tard, Fodor et Bever (1965) ont trouvé que la position subjective du click était systématiquement attirée par les frontières syntaxiques de clause. Leur interprétation est que la clause est une « unité perceptive » qui « résiste à l'interruption par le click ». Cette technique est quelque peu tombée en désuétude principalement parce que c'est une tâche de mémoire plutôt qu'une tâche perceptive, et qu'elle est susceptible de nombreux biais de réponses. Cependant, une variante simple consiste à demander aux sujets de *détecter* le click dès qu'ils l'entendent (Abrams & Bever, 1969; Holmes & Forster, 1970, 1972). L'idée est la suivante : la détection de click et la compréhension du message devant être effectuées simultanément, toute augmentation de la charge de traitement due aux processus de compréhension est reflétée par un ralentissement de la détection de click.[4] Par exemple, Holmes et Forster (1970) ont trouvé que les clicks placés sur une frontière syntaxique étaient détectés plus rapidement que des clicks placés au milieu d'un syntagme (i.e. en cours de traitement). Plus récemment, Stoffer (1985) a employé la même technique pour étudier la perception des phrases *musicales*. Il a observé que le click était détecté plus rapidement aux frontières définies par la structure hiérarchique musicale. L'interprétation est que l'auditeur organise en temps réel la phrase musicale en unités cohérentes, et que son attention se trouve libérée aux frontières de celles-ci, facilitant la détection de click. Dans l'expérience qui suit, nous examinons l'influence des frontières syllabiques sur le temps de détection du click.

4. Une autre interprétation est possible qui ne fait pas appel à la notion de « charge de traitement » : si une unité forme une « gestalt », on peut imaginer qu'un click soit plus facile à détecter entre deux « gestalts » plutôt qu'au milieu d'une.

EXPÉRIENCE 12: DÉTECTION DE CLICKS

L'hypothèse « syllabique » est que la détection de click pourrait être affectée quand celui-ci se trouve en coïncidence avec une frontière syllabique (p.ex. entre le /a/ et le /l/ de /ba-lance/, plutôt qu'entre le /a/ et le /l/ de /bal-con/). Si les unités perceptives « résistent à l'interruption », alors on s'attend à ce qu'il soit plus facile de détecter un click placé entre deux syllabes qu'un click placé au milieu d'une syllabe. Pour contrôler le mieux possible l'environnement phonétique, on a placé le click entre la voyelle et la consonne de stimuli de type « pa–lace » et « pal–mier » ('!' désignant le click, on compare la détection de '!' dans /pa!lace/ et dans /pa!lmier/). On a également placé un click après le segment 'CVC' des mêmes mots. Si la variabilité due aux changements d'environnement phonétique n'est pas trop importante, on pourrait observer que la détection de '!' est plus rapide dans /pal!mier/ que dans /pal!ace/).

DESCRIPTION

Matériel : 144 mots français, tous bisyllabiques, ont été sélectionnés. Parmi eux, il y avait 60 mots tests, appariés deux à deux pour le segment initial CVC et dont l'un commençait par une syllabe ouverte et l'autre par une syllabe fermée. Ils formaient des paires de type /pa.lace/, /pal.mier/. Pour tester un éventail de types de frontières syllabiques, vingt mots tests avaient une occlusive pour consonne post-vocalique, vingt avaient une liquide et vingt autres un /s/ (cf tab.VI.1).

Occlusives (/k/,/p/)		Liquides (/l/,/r/)		Fricative (/s/)	
CVC.	CV.C	CVC.	CV.C	CVC.	CV.C
lapsus	lapin	tartine	tarif	cascade	cassette
capture	capuche	carton	carotte	pistache	piscine
capsule	capot	tourment	touriste	moustache	mousson
soupçon	soupir	furtif	fureur	suspect	sucette
rupture	rupestre	virgule	virus	constat	concert
facteur	fakir	balcon	ballet	cristal	crisser
vaccin	vacances	palmier	palace	frisquet	frisson
tactile	taquin	valser	valise	plastic	placide
victime	vicaire	filmer	filet	plastron	placer
jonction	jonquille	culture	culotte	transfert	transept

TAB. VI.1 – *Mots tests pour la détection de click*

On a enregistré, avec une fréquence d'échantillonnage de 16 kHz, un locuteur masculin lisant ces stimuli au rythme d'un toutes les deux secondes. Leur durée moyenne était de 643 msec ($\sigma = 84$). Puis, en guise de click, on a synthétisé un ton pur à 1400 Hz qui durait 30 msec. L'amplitude maximale du click était de 4000 (max=16384), ce qui correspond à l'amplitude maximale typique atteinte par les voyelles dans notre enregistrement. Le click était par conséquent nettement audible.

A l'aide d'un éditeur de son, on a estimé les positions de la fin de la première voyelle et de la fin de la consonne post-vocalique pour chacun des 60 mots tests. Notons ici que ces mesures révèlent un comportement des liquides (/r/, /l/) différent de celui des autres consonnes (occlusives (/k/, /p/) et fricative /s/) : les liquides sont plus courtes quand elles se trouvent en attaque de seconde syllabe que quand elles se trouvent en coda de première syllabe, alors que c'est le cas inverse pour les autres consonnes (cf tab.VI.1).

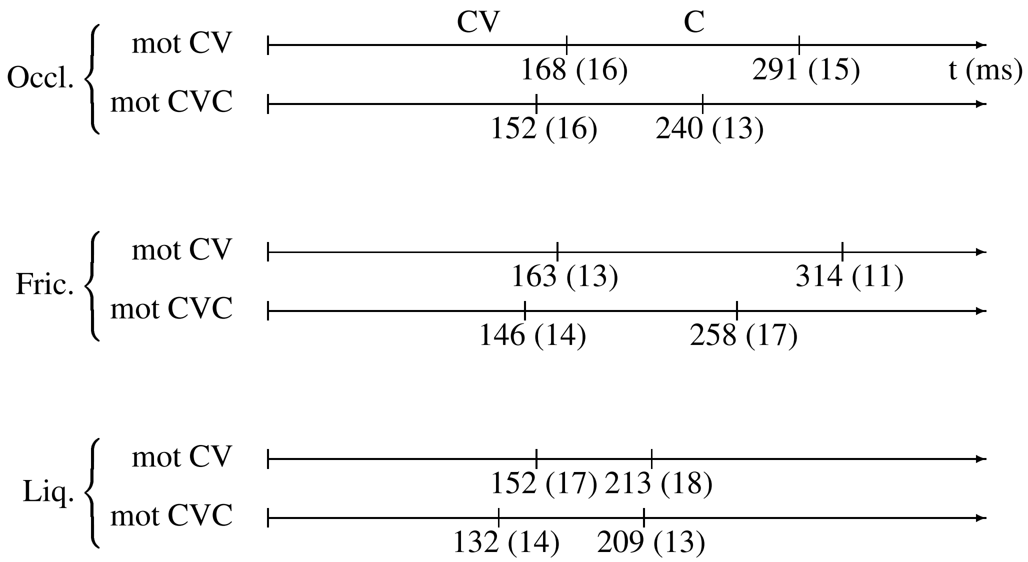


FIG. VI.1 – *Longueurs moyennes des segments CV et CVC dans des mots commençant par une syllabe CV, ou bien par une syllabe CVC (entre parenthèses : écart-types de la moyenne)*

On a construit une liste expérimentale dans laquelle tous les mots apparaissent deux fois : une fois dans la première moitié de la liste, et une fois dans la seconde. Cette liste est une suite d'essais. Chaque essai est constitué d'un mot et, éventuellement, d'un click pouvant apparaître avec différents décalages par rapport au début du mot. 176 essais possèdent un click et 112 essais n'en possèdent pas. Chaque mot test apparaît une fois avec le click juste après la voyelle, et une autre fois avec le click juste après la consonne post-vocalique (ceci étant contrebalancé entre les deux moitiés de la liste). Comme les clicks apparaissent toujours vers le milieu du mot dans les essais « tests », les clicks des distracteurs ont été placés préférentiellement soit vers la fin soit vers le début du stimulus. Les distracteurs des première et deuxième moitiés de l'expérience ne sont pas exactement les mêmes (il y a seulement un recouvrement partiel). Ainsi, les mots accompagnés d'un click ne sont pas exactement les mêmes dans les deux moitiés de la liste. Enfin, il n'y a jamais plus de trois réponses identiques (click ou non-click) sur 3 essais successifs. Une deuxième liste expérimentale a été obtenue en échangeant les deux moitiés de la première.

Procédure : Chaque sujet était testé individuellement. L'expérience se déroulait entièrement sous le contrôle d'un PC. Le sujet était averti qu'il allait entendre des mots dans l'oreille droite et, parfois, un click dans l'oreille gauche. Dès qu'il entendait un click, il devait appuyer le plus rapidement possible sur le bouton de réponse. Le sujet avait deux secondes pour répondre puis l'essai suivant débutait (l'intervalle entre chaque essai était d'environ 2.5 secondes). De plus, le sujet devait effectuer une tâche secondaire de reconnaissance des mots : tous les quinze essais, trois mots étaient successivement présentés (visuellement), et pour chacun d'entre eux le sujet devait indiquer (au clavier) s'il l'avait entendu dans les quinze essais précédents. L'expérience totale durait environ un quart d'heure.

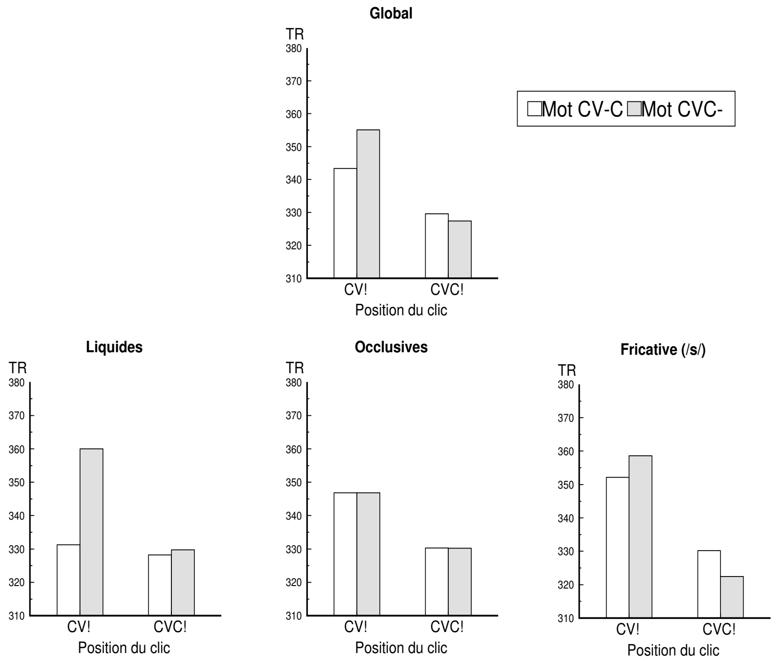
Sujets : Quarante sujets français, étudiants de diverses universités parisiennes ont participé volontairement à cette expérience. Vingt sujets étaient attribués à chaque liste expérimentale.

RÉSULTATS

Le taux global de fausse alarmes était de 2.6 % et le taux de manqués était de 0.5 %. Après avoir écrêté à 100 et 1000 msec (éliminant ainsi 0.6 % des données), on a effectué des analyses de variance, par sujets et par items, sur les moyennes des temps de réaction. Trois facteurs étaient déclarés : Mot (CV.C ou CVC.), Position du click (CV! ou CVC!), et Type de consonne postvocalique (Liquide, Occlusive ou Fricative).

Le seul effet significatif dans les deux analyses (par items et par sujets) est dû au facteur Position du click : le click était détecté 20 msec plus rapidement après un segment CVC qu'après un segment CV. L'interaction entre Mot et Position du click (7 ms), et la différence de 12 msec entre les mots CVC– et CV–C quand le click apparaissait après la voyelle, sont toutes deux significatives dans l'analyse par sujets, mais pas dans l'analyse par items. En analysant ces deux effets (l'interaction Mot$\times$Position du click, et l'effet de Mot pour les clicks postvocaliques) pour chaque type de consonne, il apparaît qu'ils ne sont significatifs par items et par sujets, que pour les Liquides (Interaction=13 msec, effet de Mot=29 msec), mais pas pour les occlusives, ni pour les fricatives.

```
Analyse par sujets: PLAN  S40*C2*M2*T3
C = Position du  click (C1=CVC! et C2=CV!)
M = Type de mot (M1=CV-C et M2=CVC-)
T = Type de consonne Post-vocalique:
T1 = Liquide
```



Le temps moyen de détection est de 338 msec. L'écart-type entre sujets est de 106 msec et celui entre items est de 7.3 msec (vous lisez bien '7.3'). Intra-sujet, l'écart-type moyen est de 83 msec.

FIG. VI.2 – *Temps de réaction en détection de click.*

T2 = Occlusive
T3 = Fricative

T	F(2,78)= 0.59	MSE=889.838	p=0.4432
C	F(1,39)= 38.50	MSE=1343.72	p=0.0000
C.T	F(2,78)= 1.87	MSE=1103.29	p=0.1610
M	F(1,39)= 4.74	MSE=585.286	p=0.0356
M.T	F(2,78)= 4.28	MSE=749.633	p=0.0172
M.C	F(1,39)= 9.49	MSE=608.445	p=0.0038
M.C.T	F(2,78)= 2.35	MSE=779.131	p=0.1021
M/C1	F(1,39)= 0.49	MSE=553.556	p=0.4881
M/C2	F(1,39)= 12.94	MSE=640.175	p=0.0009
M.C/T1	F(1,39)= 8.92	MSE=825.341	p=0.0049
M/C1/T1	F(1,39)= 0.12	MSE=427.002	p=0.2691
M/C2/T1	F(1,39)= 15.85	MSE=1040.26	p=0.0003
M.C/T2	F(1,39)= 0.00	MSE=761.19	p=1.0000
M/C1/T2	F(1,39)= 0.00	MSE=831.833	p=1.0000
M/C2/T2	F(1,39)= 0.00	MSE=569.789	p=1.0000

```
M.C/T3         F(1,39)=  3.58    MSE=580.176    p=0.0659
M/C1/T3        F(1,39)=  2.59    MSE=477.076    p=0.1156
M/C2/T3        F(1,39)=  0.95    MSE=905.297    p=0.3357
M/T1           F(1,39)= 14.29    MSE=641.922    p=0.0005
M/T2           F(1,39)=  0.00    MSE=640.432    p=1.0000
M/T3           F(1,39)=  0.02    MSE=802.196    p=0.1117
Analyse par items: PLAN  S10<T3>*C2*M2
T              F(2,27)=  0.30   MSE=494.743    p=0.2567
C              F(1,27)= 30.75   MSE=406.715    p=0.0000
C.T            F(2,27)=  1.36   MSE=406.715    p=0.2737
M              F(1,27)=  0.92   MSE=602.535    p=0.3460
M.T            F(2,27)=  0.98   MSE=602.535    p=0.3883
M.C            F(1,27)=  2.60   MSE=574.763    p=0.1185
M.C.T          F(2,27)=  0.72   MSE=574.763    p=0.4959
M/C1           F(1,27)=  0.22   MSE=515.621    p=0.3572
M/C2           F(1,27)=  2.92   MSE=661.677    p=0.0990
M.C/T1         F(1,9)=  5.52    MSE=306.213    p=0.0433
M/C1/T1        F(1,9)=  0.00    MSE=335.737    p=1.0000
M/C2/T1        F(1,9)= 19.55    MSE=175.161    p=0.0017
M.C/T2         F(1,9)=  0.00    MSE=809.742    p=1.0000
M/C1/T2        F(1,9)=  0.00    MSE=441.008    p=1.0000
M/C2/T2        F(1,9)=  0.00    MSE=886.575    p=1.0000
M.C/T3         F(1,9)=  1.04    MSE=608.332    p=0.3345
M/C1/T3        F(1,9)=  0.43    MSE=770.118    p=0.4716
M/C2/T3        F(1,9)=  0.33    MSE=923.294    p=0.4203
M/T1           F(1,9)=  8.48    MSE=204.684    p=0.0173
M/T2           F(1,9)=  0.00    MSE=517.84     p=1.0000
M/T3           F(1,9)=  0.00    MSE=1085.08    p=1.0000
```

Discussion

L'effet le plus robuste observé dans cette expérience est que le click est plus facile à détecter après le segment CVC qu'après le segment CV. On peut invoquer deux causes possibles : (a) la réponse s'accélérerait quand le click s'éloigne du début du stimulus ; (b) il y aurait un masquage énergétique moins important après la consonne qu'après la voyelle. L'hypothèse (a) est raisonnable : dans les tâches de détection, il est typique que le temps de réaction décroisse quand la cible est éloignée du début de l'essai (Luce, 1986). Cela peut s'interpréter comme une baisse du critère de décision du sujet. Pour tester l'hypothèse (a), nous avons effectué une régression linéaire entre la position temporelle du click et les temps de réaction moyens sur les mots tests. La corrélation est hautement significative (corr= $-.32$; F(1,118)=13.7 ; $p = .003$). Cela n'est pas surprenant puisque il y a une corrélation importante entre position temporelle et position phonémique (CV! vs CVC!). Toutefois, dans une régression multiple où l'on rajoute la variable catégorielle position phonémique (« après V » ou « après C »), l'effet de position temporelle disparaît totalement. De plus, des régressions restreintes à CV! et CVC! ne révèlent aucune corrélation significative (corr=-.09 et corr=-.05 respectivement). Ces résultats rendent peu probable l'hypothèse d'une décroissance du temps de réaction avec la position temporelle de la cible. Quant à la seconde hypothèse (masquage énergétique), si elle ne peut être écartée sur la base de nos données, on voit difficilement comment elle peut expliquer la différence entre les mots CV.C et CVC. avec consonne post-vocalique liquide.

La théorie syllabique prédisait que le click serait plus facile à détecter quand il coïnciderait avec la frontière syllabique. Cet effet devait se traduire : (1) par un avantage des mots CV–C sur les mots CVC– quand le click était après la voyelle (CV!) et (2) éventuellement par une interaction entre la structure des mots et la position du click. Nos résultats suggèrent que ce n'est pas le cas, *sauf quand la consonne post-vocalique est une liquide*. Se pourrait-il que la frontière syllabique soit influencée par le type de consonne postvocalique ? Pour sélectionner le matériel, nous avions utilisé une syllabification « intuitive ». Toutefois, après coup, nous avons appris qu'il existait plusieurs théories sur la syllabification des groupes de consonnes. Certaines se fondent sur des critères acoustico-phonétiques, d'autres sont d'inspiration plus phonologique. La table VI.2 présente six propositions (d'après Laeufer, 1992).

		Grammont	Delattre		Pulgram Malmberg	Noske	Levin
			apt.	force			
OL	*caprice*	-pr	-pr	-pr	-pr	-pr	-pr
	atlas	-tl	-tl	-tl	t-l	t-l	t-l
ON	*technique*	-kn	-kn	-kn	k-n	k-n	k-n
OF	*adverbe*	-dv	-dv	-dv	d-v	d-v	d-v
OO	*structure*	-kt/k-t	k-t	k-t	k-t	k-t	k-t
FL	*casserole*	-sr	-sr	-sr	s-r	s-r	-sr
	disloque	-sl	-sl	s-l	s-l	s-l	s-l
	influent	-fl	-fl	-fl	-fl	f-l	-fl
FN	*transmis*	-sm	s-m	s-m	s-m	s-m	s-m
FF	*blasphème*	-sf/s-f	s-f	s-f	s-f	s-f	s-f
FO	*diphtongue*	f-t	f-t	f-t	f-t	f-t	f-t
NL	*minerai*	-nr	-nr	-nr	n-r	n-r	n-r
NN	*calomnie*	-mn/m-n	-mn	-mn	m-n	m-n	m-n
NF	*hameçon*	m-s	-ms	-ms	m-s	m-s	m-s
NO	*samedi*	m-d	m-d	-md	m-d	m-d	m-d
LL	*galerie*	-lr/l-r	-lr	-lr	l-r	l-r	l-r
	berlue	-rl/r-l	r-l	r-l	r-l	r-l	r-l
LN	*calmant*	l-m	l-m	-lm	l-m	l-m	l-m
LF	*répulsif*	l-s	l-s	-ls	l-s	l-s	l-s
LO	*culbute*	l-b	l-b	-lb	l-b	l-b	l-b

O = occlusives ; F = fricatives ; N = nasales ; L = liquides

TAB. VI.2 – *Syllabifications de mots Français*

Le cas des groupes de consonnes qui débutent par une occlusive est d'un intérêt particulier pour nous. Pratiquement toutes les théories s'accordent à syllabifier les groupes occlusive–occlusive (OO) entre les deux consonnes : /capture/ est syllabifié en /cap-ture/. Par contre, les théories sont en désaccord sur le statut des groupes occlusive-fricative (OF) : /capsule/ est-il /ca-psule/ ou /cap-sule/ ? Nos stimuli à consonne post-vocalique occlusive contenaient quatre mots de type OF et six mots de type OO. Nous avons donc effectué des analyses des temps de détection séparées pour ces deux groupes de stimuli. Celles-ci n'ont pas révélé de différence entre les stimuli OO et les stimuli OF. Il ne semble donc pas qu'une théorie alternative de la syllabification puisse expliquer l'absence d'effet sur les occlusives.

Se pourrait-il alors que les effets observés aient été provoqués par un artefact et n'aient pas de lien direct avec la structure syllabique des stimuli ? On ne peut être que frappé par le fait, qu'à la fois dans les mesures de durées et dans les temps de réaction, les liquides se comportent différemment des obstruantes (occlusives et /s/). Mais, à l'examen, il n'y a pas de relation directe entre les temps de réaction et la position temporelle du click.

Il est remarquable que dans les études originales de détection de fragments (Mehler, 1981; Cutler et al., 1986), tous les stimuli avaient une consonne post-vocalique liquide (*palace, palmier, balance, balcon*...etc). Cette observation a conduit Frauenfelder et Rietveld (cités dans Frauenfelder (1992)) à réaliser, en hollandais, une expérience où ils variaient le type de consonne post-vocalique. En fait, ils ont observé l'interaction syllabique quand la consonne était une liquide (p.ex. *polis–poolse*), mais pas quand c'était une occlusive (p.ex. *poker–pookte*) ; les nasales (p.ex. *steno–steentje*) présentaient une tendance intermédiaire. Or les liquides sont réalisées par des allophones différents selon leur position syllabique (coda ou attaque), alors qu'il n'y a qu'un allophone pour les occlusives (en hollandais). L'interprétation de Frauenfelder et Rietveld est que, dans la tâche de détection de fragment, les sujets se forment une image de la cible qui contient des détails *allophoniques* dépendants de la structure syllabique. Leurs réponses seraient alors ralenties quand le signal contient un allophone différent de celui qu'ils attendent.

Il y a donc un parallèle entre leur résultat et le nôtre : dans les deux cas, l'« effet syllabique » provient des liquides. Toutefois, l'analogie s'arrête là : dans la détection de click, les sujets n'ont pas à se former une image acoustique ou phonétique d'une cible linguistique, et donc les sujets ne s'attendent pas a priori à un certain type d'allophone. De plus, deux remarques méthodologiques doivent être signalées : leur expérience était réalisée en hollandais alors que la nôtre utilisait le français ; et une autre étude hollandaise ne trouve pas d'effet du type de consonne post-vocalique (Zwitserlood et al., 1993).

En français, il n'y a pas d'étude publiée de détection de fragment avec des consonnes qui ne soient pas des liquides. Toutefois, au laboratoire de Sciences Cognitives et Psycholinguistique, E. Dupoux (en préparation) a comblé cette lacune en utilisant *exactement le même matériel* (même enregistrement, mêmes stimuli) que celui de notre étude de détection de clicks. Les résultats qu'il a obtenu sont présentés table VI.3.

Liquides

	cv.c	cvc.
cv	671	748
cvc	681	685

36 ms
(p=.005)

Occlusives

	cv.c	cvc.
cv	691	728
cvc	736	713

30 ms
(p=.004)

Fricatives

	cv.c	cvc.
cv	702	743
cvc	730	745

13 ms
(p=.09)

TAB. VI.3 – *Temps de réaction dans une tâche de détection de fragments (en ligne, le type de fragment : CV ou CVC ; en colonne : le type de mot ; et sous le tableau : significativité par sujets de l'interaction)*

L'interaction « syllabique » est significative aussi bien sur les occlusives que sur les liquides (elle n'est que marginale sur les stimuli avec un /s/ en position post-vocalique, dont la syllabification est d'ailleurs intuitivement moins claire).[5] Cela résout la question sur l'ambiguïté de la syllabification des stimuli avec occlusives, et par là-même, suggère que l'effet observé sur les liquides en détection de click pourrait bien n'être pas dû à la structure syllabique des stimuli.[6]

Finalement, cette expérience nous enseigne que la détection de click n'est pas très sensible à la présence de frontières syllabiques. Cela peut avoir deux causes : soit les frontières de syllabes ne sont pas des points particuliers pour les processus psychologiques, soit la méthode n'y est pas sensible. En tout cas, pour poursuivre avec cette technique, il faudrait d'abord déterminer si les effets de masquage acoustique sont négligeables ou non, en contrôlant précisément l'acoustique (ce qui implique d'utiliser des stimuli synthétiques. Mais alors, on court le risque de supprimer les indices de syllabicité contenus dans le signal.). Quoiqu'il

5. Les temps assez lents sont dus au fait que les sujets étaient « ralentis » par une tâche secondaire.

6. Ceci dit, on pourrait argumenter que la structure syllabique soit plus « marquée » dans le signal pour les liquides que pour les occlusives (conformément à la proposition de Rietveld et Frauenfelder) et qu'elle n'influence les temps de détection de click que dans le premier cas. Mais ceci n'est qu'une hypothèse post-hoc, et c'est seulement en obtenant plus de données sur le décours temporel de la syllabification qu'on pourrait décider si elle est consistante.

en soit, la comparaison des temps de réaction entre des stimuli différents se heurtera toujours au risque que l'effet éventuellement observé soit dû à des facteurs non contrôlés.

Il nous semble cependant que la technique de détection de click pourrait être utilisée dans un esprit très différent de celui qui a motivé l'expérience précédente. Nos études d'induction attentionnelle en détection de phonème (chap.III) suggèrent de l'utiliser, non pas pour mesurer une « charge de traitement », mais plutôt pour essayer d'induire des attentes quant à la position du click chez les sujets. Par exemple, si l'on habitue les sujets à détecter des clicks placés à des frontières syllabiques, il se pourrait qu'ils soient ensuite plus rapides pour détecter un click arrivant à nouveau dans une telle position plutôt qu'au milieu d'une syllabe (on comparerait ces sujets à d'autres pour qui les clicks apparaîtraient le plus souvent en milieu de syllabe). Cela prouverait que la frontière syllabique (a) est psychologiquement réelle et (b) est identifiée dans le temps très court que prend la détection de click. Utilisée de cette façon la détection de click pourrait être un instrument extrêmement puissant : on pourrait corréler le click avec quasiment n'importe quel type d'événement et observer si les sujets peuvent alors « prédire » la position du click. Pour ne prendre qu'un exemple : on pourrait placer le click le plus souvent en correspondance avec les syllabes accentuées, et voir si des locuteurs d'une langue utilisant l'accent pour distinguer les mots (p.ex. les Espagnols, par opposition aux Français) peuvent utiliser cette régularité.

Chapitre VII

Conclusion

Felix qui potuit rerum cognoscere causas
Goscinny et Uderzo, Astérix en Corse, 1973, page 22 ;
(voir aussi : Virgile, Géorgiques II).

La situation

Pendant longtemps, dans l'esprit des chercheurs, la perception de la parole était confondue avec la reconnaissance des phones (ou des traits) tant il apparaissait évident que la tâche du système perceptif était de fournir une représentation phonétique du signal de parole, celle-ci devant ensuite servir à reconnaître les mots. La question principale était, en conséquence, « comment le cerveau parvient-il à identifier les phones (ou les traits) dans le signal acoustique ? ». Deux conceptions partageaient le domaine : selon la première les phones étaient des objets articulatoires reconnus par un décodage complexe (Liberman et al., 1967) ; selon la seconde, les traits étaient caractérisés par des propriétés acoustiques, certes abstraites mais néanmoins « lisibles » dans le signal (Stevens & Blumstein, 1981; Blumstein & Stevens, 1985).[1] Cependant, la théorie motrice n'a jamais détaillé *comment* les connaissances articulatoires pouvaient permettre de reconstruire les phones. Quant aux invariants caractérisant les traits phonétiques, leur existence demeure hypothétique.

Graduellement, on a commencé à s'interroger sur la nécessité de construire une représentation phonétique pour reconnaître les mots (p.ex. Warren & Ackroff, 1976; Rubin, Turvey, & Gelder, 1976; Klatt, 1979). Le *lexique* est passé au centre des préoccupations des chercheurs.[2] Certains ont proposé d'abandonner totale-

1. Voir J. L. Miller, 1990 pour un point de départ d'une revue de l'opposition entre ces deux théories.

2. Pour Frauenfelder (1992), la distinction entre la « perception des sons de parole » et la

ment l'hypothèse d'un décodage phonétique et ont suggéré que les mots étaient reconnus à partir de représentations proches de l'acoustique du signal (Klatt, 1979; Pisoni, 1992). Une conception alternative, popularisée par Elman et McClelland (1986) et par Marlsen-Wilson (1987; voir aussi Warren & Marslen-Wilson, 1988, 1988), est que le système perceptif fournit, en continu et en parallèle, des degrés de confirmation de la présence de traits acoustico-phonétiques dans le signal, et que c'est au niveau du lexique que s'effectue la « catégorisation » du signal. Sous l'influence de ces modèles, la perception de la parole est maintenant couramment considérée comme un processus continu d'activation et de compétition entre des détecteurs de traits ou de phonèmes ; l'idée que le système perceptif construit une représentation prélexicale catégorisée du signal de parole n'est pas à la mode.

Pourtant, Mehler, Dupoux et Segui (1990) ont fait remarquer que si la représentation perceptive est tellement fluide et fluctuante, il est difficile de comprendre comment le bébé fait pour acquérir son lexique. S'il est envisageable que les adultes peuvent déterminer si deux objets sonores sont le même mot en les comparant à un patron (« pattern ») préalablement mémorisé, l'enfant, lui, ne peut pas appliquer une telle stratégie. Un préalable pour pouvoir apprendre un lexique, semble-t-il, est que l'enfant soit capable de déterminer si deux items sonores sont linguistiquement identiques ou non. En conséquence, raisonnent Mehler et al. (1990), notre système perceptif doit extraire du signal de parole un code relativement « stable ».

Cet argument ne favorise pas en-soi un code plutôt qu'un autre. Cependant, pour Mehler et al. (1990), un code syllabique offre de nombreux attraits. D'une part, on admet souvent que la syllabe doit être plus facile à récupérer dans le signal que le phonème ou le trait (Liberman & Studdert-Kennedy, 1978) : par exemple, la normalisation au débit semble s'effectuer en grande partie intra-syllabiquement (Miller, 1987). D'autre part, en postulant que les mots sont des syllabes, l'enfant aurait résolu en grande partie le problème de la segmentation lexicale car les débuts de syllabes sont souvent des débuts de mots[3]. Mehler et al. présentent des arguments expérimentaux pour la syllabe tel que le fait que les enfants discri-

---

« reconnaissance des mots » doit être questionnée. Le déplacement du centre d'intérêt du premier vers le second problème est très visible quand on compare les recueils d'articles du début des années quatre-vingt (Cole, 1980; Myers, Laver, & Anderson, 1981; Perkell & Klatt, 1986) à ceux plus récents (Frauenfelder & Tyler, 1987; Marlsen-Wilson, 1989; Altmann, 1990).

3. Comme nous l'avons signalé page 64, les frontières de syllabes sont loin de toujours correspondre avec les frontières de mots ; et ceci, en français, principalement à cause des clitiques (p.ex. « je.la.ra.che ») et de la liaison (« pe.ti.tours »). Cependant, en première approximation, et pour la construction du lexique, postuler des frontières syllabiques aux frontières de mots est une bonne heuristique. D'ailleurs il est extrêmement commun que les enfants français commettent des erreurs « révélatrices » en segmentant « un avion » ou « un éléphant » en « un navion » et « un néléphant ».

minent des suites de phonèmes qui sont des syllabes ('pat' vs 'tap') mais pas des suites de phonèmes qui n'en sont pas ('tsp' vs 'pst').

Si Mehler et al. (1990) s'étaient contentés d'affirmer que les frontières de syllabes servaient d'indices de frontières de mots (ainsi que Norris et Cutler (1985) le proposent), leur proposition n'aurait sans doute pas soulevé beaucoup de problèmes. Cependant, ils sont allés beaucoup plus loin, en proposant que la syllabe était l'« unité de perception » de la parole et, également, l'« unité d'accès au lexique » chez l'adulte (voir aussi Mehler, 1981; Segui, 1984).

A première vue, cette proposition ne cadre pas avec les conceptions contemporaines sur la reconnaissance des mots. Si la nature a « horreur du vide », on peut dire que les psychologues contemporains ont « horreur de la discontinuité », et résistent à l'emploi des termes comme « segmentation » ou « catégorisation », et ceci, dans le domaine de la parole, sous les influences de Klatt (1980), Marlsen-Wilson (1987, 1989), Elman et McClelland (1984, 1986). L'image dominante de la perception de la parole est celle de processus qui fonctionnent en continu et en parallèle.

Le terme de segmentation employé par Mehler et al. (1981), suggère qu'il est nécessaire d'attendre la fin d'une syllabe pour identifier son premier phonème. Traditionnellement, la « segmentation » est un processus supposé précéder la catégorisation : 1. on découpe le signal, 2. on le catégorise. Or, c'est un fait que les sujets peuvent effectuer une détection avant d'atteindre la fin de la syllabe (Norris & Cutler, 1988). Une observation similaire en français avait conduit Dupoux (1989) à proposer une première étape de catégorisation en demi-syllabes. La thèse que la syllabe est l'unité d'accès au lexique semble également problématique puisqu'il n'y a pas d'évidence de discontinuité qui serait due aux frontières syllabiques dans l'accès au lexique (voir Marlsen-Wilson, 1984 et Frauenfelder & Henstra, 1988, mais aussi Dupoux & Mehler, 1990). En fait, dans leurs écrits, Mehler et ses collègues n'affirment jamais que les niveaux de traitement supérieurs doivent attendre la fin de la syllabe, ou même l'identification d'une unique syllabe, pour s'effectuer.[4] Mehler (communication personnelle) envisage fort bien l'idée d'un modèle en cascade du type de ceux décrits par McClelland (1979) ou Norris (1986). Ainsi, des détecteurs de syllabes pourraient avoir une sortie continue et un phonème pourrait être « détecté » précocement par une « conspiration » des syllabes commençant par celui-ci (Dupoux, 1993).

Discriminer un modèle en cascade où un niveau syllabique est suivi d'un niveau phonémique, et un modèle où les phonèmes sont identifiés directement, est une tâche délicate. Comment justifier expérimentalement l'existence d'un niveau

4. Mais, ils présentent comme arguments en faveur de la syllabe les faits que les temps de détection de phonème initial sont corrélés avec la durée de la syllabe, et que la présence d'une consonne en coda ralentit le temps de réaction (Segui et al., 1990).

syllabique ? La pierre de touche des données de Mehler et ses collègues est l'effet de congruence syllabique, c'est à dire le fait que les sujets détectent plus facilement des fragments qui correspondent exactement à la structure syllabique des syllabes. Toutefois, comme l'ont montré Cutler, Mehler, Norris et Segui (1986), ce résultat dépend de la langue et particulièrement, n'apparaît pas chez les locuteurs anglais. Leur conception actuelle est que les sujets n'utilisent pas universellement la syllabe, mais plutôt l'unité rythmique de leur langue. Cette théorie explique par exemple le comportement des Japonais qui détectent plus facilement les fragments respectant la structure moraïque de leur langue (Otake et al., 1993). Pour l'anglais la proposition est que ceux-ci utilisent une unité « accentuelle », mais il n'y a pas de preuve positive de l'existence de celle-ci, obtenue avec la tâche de détection de fragment. A. Cutler a rassemblé de nombreux éléments en faveur d'une utilisation du rythme pour la segmentation *lexicale*, mais elle-même n'interprète pas ces résultats en faveur d'une catégorisation dans une unité de taille syllabique (Norris & Cutler, 1985; Cutler, 1993).

Nos recherches étaient motivées par le découverte de nouveaux paradigmes expérimentaux destinés à tester le rôle de la syllabe dans la perception de la parole. Au cours de nos expériences, nous avons fait trois observations qui semblent *a priori* en contradiction avec des prédictions de l'hypothèse selon laquelle la syllabe est l'« unité primaire de perception » (particulièrement dans une version qui postule l'existence de « détecteurs de syllabes ») :

1° Dans le paradigme de Garner (exp. 1 et 2), les sujets n'arrivaient pas à focaliser leur attention sur la première syllabe de stimuli bisyllabiques. S'il existait des « détecteurs de syllabes », on aurait pu s'attendre à ce que la décision puisse se fonder sur le premier détecteur qui dépasse un seuil ; il n'y aurait alors pas dû y avoir d'interférence entre deux syllabes adjacentes.

2° Dans la tâche de classification de Jeff Miller (exp. 9, 10, 11), les sujets n'ont pas effectué un appariement « direct » entre les syllabes et les réponses : on observe des effets de similarité, en termes de phonèmes et même de traits phonétiques. Cela met en cause l'idée que chaque syllabe évoque un code « atomique ».

3° Dans les expériences de détection de phonème attentionnelle (chap.III et IV), les sujets détectaient aussi rapidement un phonème placé en début de seconde syllabe qu'un phonème placé en fin de première syllabe. Or, on aurait pu s'attendre à ce qu'en fin de syllabe, celle-ci étant presque identifiée, le code phonémique devienne immédiatement disponible ; en début de syllabe par contre, il aurait fallu attendre plus de temps pour que la syllabe soit identifiée et que le code phonémique soit récupéré.

Ces résultats « non-syllabiques », s'ajoutent aux observations de Norris et

Cutler (1988) et Marlsen-Wilson (1984), qui suggèrent que la syllabe n'est pas une étape limitante (un « bottleneck ») de transmission de l'information. Cependant, nous avons également obtenu de clairs effets syllabiques :

1° Dans la tâche de classification de phonème « à la Garner » (exp.3), les sujets étaient plus gênés par une variation intra-syllabique que par une variation extra-syllabique. Cela montre que, pour fonder leur réponse, les sujets n'utilisent pas une représentation qui serait une simple suite de phonèmes. Les phonèmes appartenant à la même syllabe sont plus « liés » entre-eux que des phonèmes n'appartenant pas à la même syllabe.

2° Dans les expériences de détection de phonème biaisée attentionnellement (chapitre III), les sujets détectaient plus facilement un phonème dans une position syllabique prévisible que dans une position syllabique non prévisible. Par contre, ils n'étaient pas sensibles à la position phonémique séquentielle. Cela suggère encore que les sujets sont sensibles aux relations syllabiques entre phonèmes.

Ces découvertes « syllabiques » s'ajoutent à l'effet de congruence en détection de fragments (Mehler et al., 1981), à l'effet de complexité syllabique en détection de phonème initial (Segui et al., 1990), et à l'interaction entre longueur syllabique et effets lexicaux (Dupoux & Mehler, 1990). Dans la section suivante, nous proposons un modèle qui a l'ambition de résoudre la contradiction entre ces résultats « syllabiques », et les résultats précédents « anti-syllabiques ».

Un modèle

La première liste de résultats expérimentaux (« anti-syllabiques ») révèle la naïveté du modèle implicite de la décision mise en jeu dans chaque tâche. On supposait une relation quasi directe entre le traitement (la sortie des détecteurs) et les réponses du sujet. Pourtant, comme nous l'avons suggéré dans les discussions des chapitres expérimentaux, la décision joue un rôle primordial dans la réponse. Ainsi, il nous semble qu'une large partie de l'interférence Garner peut être attribuée à un effet de *distraction* (cf également Shand, 1976). Dans le paradigme de classification à quatre doigts (inspiré de Jeff Miller, 1982), nous avons argumenté que les effets de similarité n'étaient pas nécessairement dus au fait que le phonème ou les traits étaient identifiés en temps réels, puisque la facilitation apparaissait également quand l'information était tardive.

L'importance des processus décisionnels dans les tâches employées en psychologie est reconnue par de nombreux auteurs (Forster, 1979; Pylyshyn, 1984). La conception la plus répandue est que parmi l'ensemble des représentations calculées « automatiquement » par le système de traitement, le sujet focalise son attention sur l'une d'entres elles (ou à la rigueur partage son attention entre quelques-unes) pour effectuer sa réponse. Le choix d'une représentation plutôt qu'une autre est sous le contrôle du sujet, mais typiquement on suppose que celui-ci choisit la mieux adaptée à la tâche (McNeill & Lindig, 1973). Cette conception est sous-jacente dans des affirmations telles que « dans une tâche de détection de phonème, le sujet peut choisir d'utiliser une représentation phonétique prélexicale ou une représentation phonologique post-lexicale » (Foss & Blank, 1980; Cutler et al., 1987). Dans cette optique, les manipulations attentionnelles consistent essentiellement à inciter le sujet à utiliser une représentation plutôt qu'une autre (Eimas et al., 1990).

Une caractéristique des modèles de traitement de la parole habituels est que les représentations sont rarement supposées être plus complexes que des simples suites de symboles. Ainsi, dans les modèles qui accordent un rôle à la syllabe, il y a d'une part un niveau syllabique et d'autre part un niveau phonémique (p.ex. Dupoux, 1989). Dans ce cadre, on interprète l'effet de congruence syllabique (Mehler et al., 1981) en supposant que les sujets fondaient leur réponses sur le niveau syllabique dans les cas de congruence (« match » : p.ex. /pa/ dans /palace/) et sur le niveau phonémique dans les cas de non-congruence (« mismatch » : (/pa/ dans palmier/) (pour une explication plus détaillée, voir Dupoux & Mehler, 1992).[5]

Pourtant, cette conception selon laquelle les stimuli possèdent à la fois une représentation syllabique et une représentation phonémique, à des niveaux de traitement différents, nous semble étrange. Nos expériences utilisant le paradigme de détection de phonème attentionnelle peuvent difficilement s'interpréter dans un tel cadre. Rappellons que le sujet pouvait focaliser son attention sur un phonème précis à l'intérieur de la structure syllabique des stimuli. Par contre, il ne pouvait pas focaliser son attention sur le « nième » phonème. Il ne semble donc pas avoir de représentation où les phonèmes forment une simple chaîne. Par contre, ces résultats sont plus facilement interprétables si l'on considère que le sujet utilise une représentation phonémique, non pas séquentielle, mais *possédant une structure syllabique*. Plutôt que de multiplier les niveaux de représentations, nous proposons donc *d'enrichir* l'une d'entre-elles. Cela entraîne un changement de perspective : au lieu de manipuler l'attention entre les niveaux comme Eimas et al. (1990), nous avons fait varier l'attention à l'intérieur de la représentation d'un niveau.

Voici la clé du paradoxe que nous proposons : les « effets syllabiques » décrits plus hauts s'interprètent très bien en supposant que notre système perceptif fournit une représentation du signal *structurée syllabiquement*. Ils n'impliquent pas (mais

5. Mehler et al. (1981) semblent déduire de ce résultat que la syllabe est reconnue *avant* le phonème. Toutefois, si l'on prend au sérieux un modèle de décision comme celui que propose Forster (1979), ce résultat peut s'expliquer par le fait que le niveau qui gouverne la réponse est celui sur lequel il y a le moins d'unités à comparer.

ne rejettent pas non plus) l'existence de détecteurs de syllabes, ni même d'une représentation purement syllabique où la parole serait représentée par une chaîne comme « syll43.syll567... ».

En fait, il me semble qu'il faudrait a priori distinguer soigneusement deux rôles que pourrait jouer la syllabe : (1) celui d'unité de contact avec le signal de parole et (2) celui d'unité de représentation de celle-ci. Ainsi, les arguments expérimentaux cités en faveur de la syllabe, peuvent être séparés entre ceux qui concernent le signal, et ceux qui concernent le format représentationnel de la parole. Le plus classique des arguments pour la syllabe comme unité de décodage est qu'elle permettrait, mieux que le phonème, de capturer la variabilité du signal. L'argument est surtout négatif puisqu'il repose sur les nombreux effets de contexte sur la perception des phonèmes. Pour reprendre l'exemple de Mann et Repp (1980), un segment est identifié comme /s/ devant /a/ et /ch/ devant /u/. Pourtant il ne me semble pas que les effets de contextes doivent nécessairement être résolus en élargissant l'unité de « template-matching ». Par exemple, dans la perception des couleurs, il y a également des effets de contexte très impressionnants (Land, 1977), ceux-ci ne sont pas résolus en postulant un analyseur de couleur « à très large champ récepteur ». Aucune donnée ne s'oppose à la conception que la syllabe joue le rôle d'unité de contact avec le signal[6], mais pour le moment son statut n'est pas meilleur que, par exemple, celui des diphones (Klatt, 1980) ou des demi-syllabes (Fujimira, 1976; Samuel, 1989). Cependant, on peut légitimement se demander si le décodage de la parole se fait simplement par un « pattern-matching » d'unités présentes successivement dans le signal. Par exemple, quand un locuteur ralentit son débit de parole, il ne segmente pas celle-ci en unités séparées par des pauses : la parole demeure continue. Cette continuité semble donc essentielle, et cela s'oppose, du moins dans mon esprit, à l'idée d'un banc de détecteurs d'« unités de perception ». Comme alternative au « template-matching », plusieurs auteurs ont proposé l'extraction en parallèle et de façon asynchrone des traits phonétiques (Cole, Stern, & Lasry, 1986; Lahiri & Marslen-Wilson, 1991; Stevens, 1986) ; mais cette approche se heurte à diverses difficultés dont la principale est que les traits proposés jusqu'ici ne sont pas indépendants (cf la discussion qui suit l'article de Cole et al.). Une démonstration assez convaincante de l'existence d'un syllabaire pour la perception serait l'existence d'un effet de fréquence syllabique, comme celui que Levelt et Wheeldon (1994) ont trouvé en production. Toutefois, il faudrait décorréler la fréquence des syllabes de celles des diphones et des demi-syllabes qu'elles contiennent, tâche pratiquement impossible à réaliser.

Nous suspectons, à l'instar de Liberman et Mattingly (1985), que les premières étapes du traitement de la parole sont réalisées par un « module », et que dans la

---

6. En admettant un modèle en cascade où les syllabes peuvent faire monter vers les niveaux suppérieurs le résultat d'une analyse partielle

plupart des tâches (notamment toutes celles de détection), les sujets n'ont accès qu'à la représentation de sortie de ce module, et encore seulement de façon dérivée. Dans ce cadre, une question fondamentale porte sur le format de cette représentation. Nos résultats, ainsi que ceux de Mehler et al. (1981), incitent à penser que les représentations que se forme le sujet en écoutant la parole et en mémorisant le fragment cible possèdent toutes deux une structure syllabique, et qu'un « mismatch » de structure ralentit la réponse. Cela nous conduit à l'idée que les sujets se forment une représentation phonologique structurée des stimuli linguistiques, idée qui a été proposée, sur la base d'arguments théoriques, par Church (1987) et Frazier (1987). Ainsi, Church (1987) propose un « parseur phonologique » qui construit une représentation phonémique *syllabifiée* a partir des informations allophoniques contenues dans le signal ; l'exigence de devoir fournir une suite syllabifiée permet de contraindre fortement les interprétations phonétiques du signal : la syllabe joue comme un « code correcteur ». Chez Church, les unités de base sont les phonèmes, mais, la représentation de sortie pourrait « afficher » des traits phonétiques (ainsi que le proposent, p.ex., Lahiri & Jongman, 1990 et Lahiri & Marslen-Wilson, 1991), et les insérer dans une structure syllabique de sortie. Le système de reconnaissance lexicale examinerait cette représentation, et celle-ci serait également à l'origine de celles dérivées pour effectuer les tâches de détection. La figure VII.1 présente une esquisse du modèle proposé.

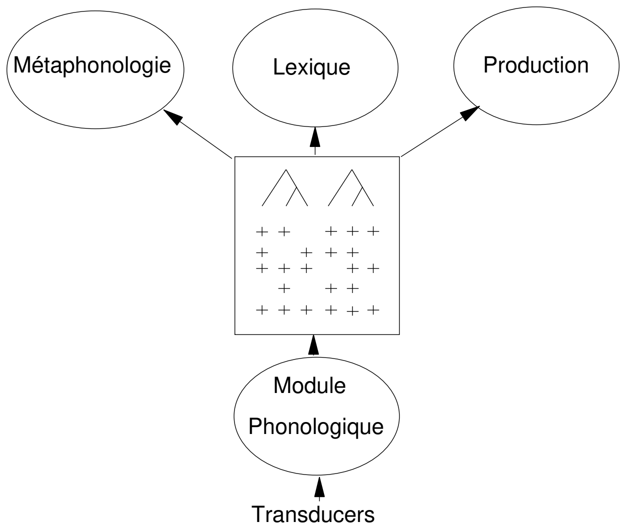


FIG. VII.1 – *Un Modèle*

Les schémas classiques de traitement de l'information, par la multiplication des niveaux, donnent une image éclatée des traitements acoustiques, phonétiques et syllabiques. Par exemple, Dupoux (1989) suppose qu'il existe un code syllabique et que celui-ci peut-être transformé, subséquemment, en code phonémique. Au contraire, nous proposons qu'il existe une représentation unique, phonologique, qui s'élabore progressivement dans le temps (voir plus bas), et qui est le résultat de la « traduction » du signal accomplie par un « module phonologique ». Cette représentation sert de point d'entrée à tous les processus qui ont besoin d'inspecter la parole, en particulier : le Lexique (qui permet de reconnaître les mots), la Production (qui permet de répéter la parole), et la Métaphonologie. Par « Métaphonologie », on désigne les processus et les représentations utilisées par la conscience phonologique, ainsi que par les tâches métaphonologiques, dont font partie les tâches de détection de phonèmes ou de syllabes.

Dans ce modèle, la question fondamentale est celle du format de la représentation de sortie. Sur le dessin, nous avons fait figurer des traits placés dans une structure syllabique. Cependant, beaucoup d'autres représentations étaient *a priori* concevables :

1° une suite de phones : [$p^h ato$]

2° une matrice de traits distinctifs (ceux-ci pouvant éventuellement être récupérés de façon asynchrone) :

$$\begin{bmatrix} + & + & - & - \\ - & + & - & + \\ - & - & - & + \end{bmatrix}$$

3° des versions « probabilistes » des représentations précédentes, par exemple pour les phonèmes :

$$\begin{bmatrix} p(.5) & a(.8) & t(.5) \\ t(.3) & o(.1) & d(.3) \\ s(.2) & u(.1) & p(.2) \end{bmatrix}$$

4° une suite de phones (ou de traits) sous-déterminée : [$b?do$] ou '?' désignent un phonème non reconnu.

5° une suite de symboles pour chaque syllabe : [$\alpha\beta\gamma$]

6° une représentation phonémique avec une structure syllabique : [$(ba)(l\tilde{a}s)$].

Plutôt que ces représentations, nous favorisons une représentation en traits phonétiques placés dans une structure syllabique. L'introduction de la syllabe est justifiée par les résultats de cette thèse ainsi que par ceux décrit par Segui et al. (1990). Les syllabes ne sont toutefois pas des « atomes », ne serait-ce que parce qu'elles doivent être « analysées » pour accéder au lexique quand une frontière de mots tombe à l'intérieur d'une syllabe. La motivation pour les traits provient essentiellement des effets de similarité entre syllabes, comme celui que nous avons observé dans l'expérience 11, mais également des arguments psychologiques (p.ex. Shepard, 1972) et linguistiques « classiques » (p.ex. Jakobson & Waugh, 1980). Quand on perçoit un phonème ou une syllabe ambigu, c'est simplement parce que tous les traits n'ont pas été reconnus. Précisons bien que cette représentation n'exclut pas les phonèmes : les traits ne sont pas attachés directement à la syllabe, mais à des positions à l'intérieur de la structure que fournit celle-ci (pour des arguments psychologiques, cf Chodorow & Manning, 1983). Finalement, notre schéma ne fait pas apparaître les informations accentuelles et sur l'intonation bien que nous pensions qu'elles sont transmises par le module phonologique.

Du point de vue temporel, pour expliquer l'apparente absence de discontinuité syllabique dans la montée de l'information, on peut supposer que le « remplissage » de la syllabe se fait graduellement, chaque trait étant introduit quand il a reçu suffisamment de confirmation, et qu'il est *cohérent* avec le reste de la structure (par exemple, un trait typique d'une voyelle ne sera « activé » qu'une fois par syllabe). Cette dernière remarque fait toute la différence avec les modèles qui supposent que chaque trait est extrait indépendamment (p.ex. Lahiri & Marslen-Wilson, 1991). La voyelle pourrait être identifiée avant la ou les consonnes de l'attaque (Remington, 1977). Il n'y a pas d'objection de principe à ce que la construction d'une seconde syllabe commence avant que celle de la précédente soit achevée (ce qui pourrait expliquer les effets d'interférence observés dans les expériences 1 et 2).

Dupoux (1989) avait dû introduire un niveau de représentation sub-syllabique pour expliquer que les sujets sont parfois sensibles à la syllabe (aux temps de réaction lents, typiquement), parfois non sensibles à elle (aux temps rapides typiquement). Dans notre modèle, la syllabe se construit progressivement : elle n'est pas encore achevée quand les sujets répondent rapidement, mais elle l'est quand ils répondent « lentement » ; les sujets n'inspectent toujours qu'une seule et même représentation mais à des moments différents. Les deux modèles différent dans leurs prédictions empiriques. Dans celui de Dupoux, il serait imaginable que les sujets puissent focaliser leur attention sur le niveau demi-syllabique et répondre à partir de ce niveau *quelque soit le temps de réponse*. Dans le modèle que nous proposons, les sujets ne peuvent ignorer la structure syllabique une fois qu'elle est récupérée. Les propres analyses de Dupoux (1989) suggèrent que c'est le deuxième cas qui est vrai, ce dont Dupoux rendait compte en supposant que le code sub-syllabique était « transitoire ».

À la différence d'un modèle où des détecteurs de syllabes fournissent en continu et en parallèle des degrés d'activation, nous proposons que la sortie du module perceptif est une représentation symbolique « stable », c'est à dire non probabiliste.[7] La principale motivation de cette hypothèse est l'argument, avancé par Mehler et al. (1990), de l'acquisition du lexique par le bébé. L'intérêt des frontières syllabiques est de fournir un indice de frontière de mot. La syllabe elle-même est représentée par un code complexe, plutôt que par un symbole atomique. Cela résout un « paradoxe » du modèle SARAH (Mehler et al., 1990). Dans celui-ci, les détecteurs de syllabes semblent supposés innés[8]. Cela suggérait que l'enfant devait posséder un analyseur (et un symbole) pour chaque syllabe *potentielle* d'une langue humaine. Cette conclusion est difficile à accepter car ce nombre dépasserait nettement la taille du lexique d'un adulte ; et alors, on se demande pourquoi les langues humaines n'utiliseraient pas toutes des syllabes complexes, avec une syllabe pour chaque mot...

Une caractéristique importante de notre modèle est que le module phonétique est, précisement, un module. Par conséquent ses représentations internes ne sont pas accessibles. Comme, de plus, nous n'autorisons pas une sortie « probabiliste », (« graded activation »), cela représente une contrainte assez forte sur l'information à laquelle les sujets ont accès. Cependant, cette proposition semble s'opposer à une conception qui se répand parmi les chercheurs selon laquelle les sujets ont accès à une connaissance assez détaillée des caractéristiques acoustiques des phones de leur langue. Ainsi, dans plusieurs études, J. L. Miller trouve que les sujets sont capables de juger de la prototypicalité d'un stimulus phonétique (Miller, 1994). C'est à dire qu'ils sont capables de dire si un stimulus est un bon exemplaire d'une syllabe donnée, par exemple /pa/, et même de donner une « note de prototypicalité ». Cette capacité requiert-elle l'accès conscient à des informations acoustiques détaillées ? Cela n'est pas nécessaire : on pourrait expliquer cet effet en supposant que *le temps* de traitement du module phonologique, s'accroît quand le stimulus diminue en qualité. Le sujet aurait ainsi une mesure de l'« effort » fourni par le module. D'ailleurs l'effet de prototypicalité doit, presque à coup sûr, se retrouver dans les temps d'identification de la syllabe.

FINALEMENT...

Durant notre « quête de l'unité de perception de la parole », nous avons été poursuivis par le paradoxe suivant : quand notre théorie du moment privilégiait le phonème (ou le trait), nos expériences montraient l'importance de la syllabe ;

7. Cette distinction est l'analogue entre l'opposition dans les modèles de production, entre celui de Shattuck-Hufnagel (1979) et celui de Dell (1986).

8. Cependant, dans le même article (p.242), les auteurs proposent que le bébé « compile un banc d'analyseurs syllabiques ».

quand au contraire nous privilégions la syllabe, c'était le phonème (ou le trait) qui sortait vainqueur du verdict expérimental. Comme on le sait, les paradoxes proviennent la plupart du temps d'une hypothèse tacite. Dans notre cas, c'était la supposition que ces unités existaient à des niveaux de représentation *distincts*, et que l'un de ces niveaux devait, dans le traitement de l'information, nécessairement se trouver avant les autres. À la place, nous proposons d'adopter comme hypothèse de travail pour les recherches futures, l'idée qu'il existe un format unique, privilégié, de la parole, qui sous-tend la perception, la production et la performance des sujets dans les tâches psycholinguistiques. Ceci est une hypothèse extrêmement forte qui, nous l'espérons, provoquera des falsifications constructives.

Bibliographie

Note: Le signe '▷' signale les réfénces intimement liées à notre problèmatique. Celles précédées du signe '◇' signalent des articles permettant une vision global du domaine.

ABBS, J. H. (1986). « Invariance and Variability in Speech Production: A Distinction between Linguistic Intent and its Neuromotor Implementation ». In J. S. Perkell, & D. H. Klatt (1986), pp. 202–219.

ABRAMS, K., & BEVER, T. G. (1969). « Syntactic structures modifies attention during speech perception and recognition ». *Quaterly Journal of Experimental Psychology*, *21*, 280–290.

ALTMANN, G. T. M. (Ed.). (1990). *Cognitive models of speech processing: psycholinguistic and computational perspectives.* Cambridge, MA: MIT Press.

BERTELSON, P. (Ed.). (1986). *The onset of literacy*. Cambridge, MA: MIT Press.

BERTELSON, P., & DE GELDER, B. (1991). « The emergence of phonological awareness: comparative approaches ». In I. G. Mattingly, & M. Studdert-Kennedy (1991), pp. 393–412.

BERTELSON, P., DE GELDER, B., TFOUNI, L. V., & MORAIS, J. (1989). « Metaphonological abilities of adult illiterates: New evidence of Heterogeneity ». *European Journal of Cognitive Psychology*, *1*(3), 239–250.

BEST, C. T., MCROBERTS, G. W., & NOMATHEMBA, M. S. (1988). « Examination of Perceptual Reorganization for Nonnative Speech Contrasts: Zulu Click Discrimination by English-Speaking Adults and Infants ». *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 345–360.

BIJELJAC-BABIC, R., BERTONCINI, J., & MEHLER, J. (1993). « How do four-day-old infants categorize multisyllabic utterances? ». *Developmental Psychology*, *29*, 711–721.

BLUMSTEIN, S. E., & STEVENS, K. N. (1985). « On some issues in the pursuit of acoustic invariance in speech: A reply to Lisker ». *Journal of the Acoustical Society of America*, *77*(3), 1203–1204.

BRADLEY, D. C., SÁNCHEZ-CASAS, R. M., & GARCÍA-ALBEA, J. E. (1993). « The status of the syllable in the perception of Spanish and English ». *Language and Cognitive Processes*, *8*(2), 197–233.

BREGMAN, A. S. (1978). « The formation of auditory streams ». In J. Requin (Ed.), *Attention and Performance VII*. Hillsdale, N. J.: Erlbaum.

BREGMAN, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

BROWMAN, C. P., & GOLDSTEIN, L. (1990). « Representation and Reality: Physical Systems and Phonological Structure ». *Journal of Phonetics*, *18*, 411–424.

CARLSON, R., & GRANSTRÖM, B. (Eds.). (1982). *The Representation of Speech in the Peripheral Auditory System*. Amsterdam, North-Holland: Elsevier.

CHERRY, E. C., & TAYLOR, W. K. (1954). « Some further experiments on the recognition of speech with one and with two ears ». *Journal of the Acoustical Society of America*, *26*, 554–559.

▷CHODOROW, M. S., & MANNING, S. K. (1983). « Syllable similarity: the effects of differences in vowels, consonants, and order ». *Quaterly Journal of Experimental Psychology*, *35A*, 139–154.

CHOMSKY, N. (1955). « Semantic considerations in grammar ». In *Georgetown University Monograph Series on Languages and Linguistics, vol.8*, pp. 141–150. Washington, DC.

▷CHURCH, K. W. (1987). « Phonological parsing and lexical retrieval ». *Cognition*, *25*, 53–70.

COLE, R. A. (Ed.). (1980). *Perception and production of fluent speech*. Hillsdale, N.J.: Erlbaum.

COLE, R. A., & JAKIMIK, J. (1980). « How are syllables used to recognize words ? ». *Journal of the Acoustical Society of America*, *67*, 965–970.

COLE, R. A., RUDNICK, A. I., ZUE, V. W., & REDDY, D. R. (1980). « Speech as patterns on paper ». In R. A. Cole (Ed.), *Perception and production of fluent speech*, pp. 3–50. Hillsdale, N.J.: Erlbaum.

COLE, R. A., STERN, R. M., & LASRY, M. J. (1986). « Performic fine phonetic distinctions: Templates versus Features ». In J. S. Perkell, & D. H. Klatt (1986), pp. 325–341.

CONNINE, C. M. (1987). « Constraints on interactive processes in auditory word recognition: The role of sentence context ». *Journal of Memory and Language*, *26*, 527–538.

CONNINE, C. M., & CLIFTON, C. J. (1987). « Interactive use of lexical information in speech perception ». *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 291–299.

CONTENT, A., MOUSTY, P., & RADEAU, M. (1990). « BRULEX : Une base de données lexicales informatisée pour le français écrit et parlé ». Université Libre de Bruxelles. Manuscript.

COULMAS, F. (1989). *The writing systems of the world*. Cambridge, MA: Blackwell.

CUTLER, A. (1993). « Language-specific processing: does the evidence converge? ». In G. T. M. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing: The second sperlonga meeting*, pp. 115–123. Hove, East Sussex, UK: Erlbaum.

CUTLER, A., & BUTTERFIELD, S. (1992). « Rhythmic cues to speech segmentation: evidence from juncture misperception ». *Journal of Memory and Language*, *31*, 218–236.

CUTLER, A., BUTTERFIELD, S., & WILLIAMS, J. N. (1987). « The perceptual integrity of syllabic onsets ». *Journal of Memory and Language*, *26*, 406–418.

CUTLER, A., & CARTER, D. M. (1987). « The predominance of strong initial syllables in the English vocabulary ». *Computer Speech and Language*, *2*, 133–142.

CUTLER, A., & FODOR, J. A. (1979). « Semantic focus and sentence comprehension ». *Cognition*, *7*, 49–59.

CUTLER, A., MEHLER, J., NORRIS, D., & SEGUI, J. (1983). « A language specific comprehension strategy ». *Nature*, *304*, 159–160.

▷CUTLER, A., MEHLER, J., NORRIS, D., & SEGUI, J. (1986). « The syllable's differing role in the segmentation of French and English ». *Journal of Memory and Language*, *25*, 385–400.

CUTLER, A., MEHLER, J., NORRIS, D., & SEGUI, J. (1987). « Phoneme identification and the lexicon ». *Cognitive Psychology*, *19*, 141–177.

CUTLER, A., MEHLER, J., NORRIS, D., & SEGUI, J. (1989). « Limits on bilinguism ». *Nature*, *320*, 229–230.

CUTLER, A., & NORRIS, D. (1979). « Monitoring sentence comprehension ». In W. E. Cooper & E. T. C. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merill Garrett*, pp. 113–134. Hillsdale, NJ: Erlbaum.

CUTLER, A., & NORRIS, D. (1988). « The role of strong syllables in segmentation for lexical access ». *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 113–121.

CUTLER, A., NORRIS, D., & WILLIAMS, J. N. (1987). « A note on the role of phonological expectations in speech segmentation ». *Journal of Memory and Language*, *26*, 480–487.

DAY, R. S., & WOOD, C. C. (1972). « Mutual interference between two linguistic dimensions of the same stimuli ». *Journal of the Acoustical Society of America*, *52*.

DELGUTTE, B. (1981). *Representation of speech-like sounds in the discharge patterns of auditory nerve fibers*. Ph.D. thesis, MIT, Cambridge, MA. Unpublished Doctoral Dissertation.

DELL, F. (1973). *Les règles et les sons*. Paris: Hermann.

◇DELL, F., & VERGNAUD, J.-R. (1984). « Les développements récents en phonologie : quelques idées centrales ». In F. Dell, D. Hirst, & J.-R. Vergnaud (Eds.), *Forme sonore du langage*, pp. 1–42. Paris: Hermann.

DELL, G. S. (1986). « A spreading activation theory of retrieval in sentence production ». *Psychological Review*, *93*, 283–321.

DIEHL, R. L., KLUENDER, K. R., FOSS, D. J., PARKER, E. M., & GERNSBACHER, M. A. (1987). « Vowels as islands of reliability ». *Journal of Memory and Language*, *26*, 564–573.

DUCHET, J.-L. (1992). *La phonologie* (troisième edition)., Vol. 1875 of *Que-sais-je ?* Paris: P.U.F.

▷DUPOUX, E. (1989). *Identification des mots parlés, Détection de Phonèmes et Unité Prélexicale*. Ph.D. thesis, École des Hautes Études en Sciences Sociales.

DUPOUX, E. (1993). « Prelexical processing: the syllabic hypothesis revisited ». In G. T. M. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing: The second sperlonga meeting*, pp. 81–114. Hove, East Sussex, UK: LEA.

DUPOUX, E. (1994). « The syllabe : A bootleneck in speech processing ? ». manuscript en préparation.

DUPOUX, E., & MEHLER, J. (1990). « Monitoring the lexicon with normal and compressed speech: Frequency effects and the prelexical code ». *Journal of Memory and Language*, *29*, 316–335.

DUPOUX, E., & MEHLER, J. (1992). « Unifying awareness and on line studies of speech: a tentative framework ». In J. Alegria, D. Holender, J. Morais, & M. Radeau (Eds.), *Analytic approaches to human cognition*, pp. 59–76. The Netherlands: Elsevier.

EIMAS, P. D., HORNSTEIN, S. B. M., & PAYTON, P. (1990). « Attention and the role of dual codes in phoneme monitoring ». *Journal of Memory and Language*, *29*, 160–180.

EIMAS, P. D., SIQUELAND, E. R., JUSCZYK, P. W., & VIGORITO, J. (1971). « Speech perception in infants ». *Science*, *171*, 303–306.

EIMAS, P. D., TARTTER, V. C., & MILLER, J. L. (1981). « Dependency relations during the processing of speech ». In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the Study of Speech*. Hillsdale, NJ: Erlbaum.

EIMAS, P. D., TARTTER, V. C., MILLER, J. L., & KEUTHEN, N. J. (1978). « Asymetric dependencies in processing phonetic features ». *Perception & Psychophysics*, *23*, 12–20.

ELMAN, J. L., & MCCLELLAND, J. L. (1984). « Speech Perception as a Cognitive Process: The Interactive Activation Model ». In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice*, Vol. 10, pp. 337–374. New York, NY: Academic Press.

◇ELMAN, J. L., & MCCLELLAND, J. L. (1986). « Exploiting lawful variability in the speech wave ». In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes*, pp. 360–385. Hillsdale, N. J.: Erlbaum.

ELMAN, J. L., & MCCLELLAND, J. L. (1988). « Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes ». *Journal of Memory and Language*, *27*, 143–165.

ERIKSEN, B. A., & ERIKSEN, C. W. (1974). « Effects of noise letters upon identification of a target letter in a nonsearch task ». *Perception & Psychophysics*, *16*, 143–149.

FANT, G., & LINDBLOM, B. (1961). « Studies of minimal speech and sound units ». In *Speech Transmission Lab; Quaterly Progress Report*, Vol. 2, pp. 1–11. Stockholm: Royal Institute of Technology.

FODOR, J. A. (1983). *Modularity of Mind*. Cambridge, MA: MIT Press. trad. française : *La modularité de l'esprit*, Editions de Minuit, Paris, 1986.

FODOR, J. A., & BEVER, T. G. (1965). « The psychological reality of linguistic segments ». *Journal of Verbal Learning and Verbal Behavior*, *4*, 414–420.

FODOR, J. A., & PYLYSHYN, Z. W. (1981). « How direct is visual perception? Some reflections on Gibson's ecological approach ». *Cognition*, *9*, 139–196.

◇FORSTER, K. I. (1979). « Levels of processing and the structure of the language processor ». In W. E. Cooper & E. T. C. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*, pp. 27–85. Hillsdale, NJ: Erlbaum.

FOSS, D. J., & BLANK, M. A. (1980). « Identifying the speech codes ». *Cognitive Psychology*, *12*, 1–31.

FOSS, D. J., & GERNSBACHER, M. A. (1983). « Cracking the dual code: Toward a unitary model of phoneme identification ». *Journal of Verbal Learning and Verbal Behavior*, *22*, 609–632.

FOSS, D. J., & JENKINS, C. J. (1973). « Some effects of context on the comprehension of ambiguous sentences ». *Journal of Verbal Learning and Verbal Behavior*, *12*, 577–589.

FOWLER, C. A. (1986). « An event approach to the study of speech perception from a direct-realist perspective ». *Journal of Phonetics*, *14*, 3–28.

FOWLER, C. A. (1987). « Perceivers as realists, talkers too: commentary on papers by Strange, Diehl et al., and Rakerd and Verbrugge ». *Journal of Memory and Language*, *26*, 574–587.

◇FOWLER, C. A., & ROSENBLUM, L. D. (1991). « The perception of phonetic gestures ». In I. G. Mattingly, & M. Studdert-Kennedy (1991), pp. 33–59.

FOWLER, C. A., RUBIN, P., REMEZ, R. E., & TURVEY, M. T. (1980). « Implication for speech production of a general theory of action ». In B. Butterworth (Ed.), *Language production: Speech and Talk*, Vol. 1. London: Academic Press.

FOWLER, C. A., TREIMAN, R., & GROSS, J. (1993). « The structure of English syllables and polysyllables ». *Journal of Memory and Language*, *32*, 115–140.

FOX, R. A. (1984). « Effects of lexical status on phonetic categorisation ». *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 526–540.

FRAUENFELDER, U. H., & HENSTRA, J. (1988). « Lexical Activation and deactivation of phonological representations in lexical access ». In *Proceedings of the third international morphological meeting. Krems, Austria*, pp. 10–12.

FRAUENFELDER, U. H., & SEGUI, J. (1989). « Phoneme monitoring and lexical processing: Evidence for associative context effects ». *Memory & Cognition*, *17*, 134–140.

FRAUENFELDER, U. H., SEGUI, J., & DIJKSTRA, T. (1990). « Lexical Effects in Phonemic Processing: Facilitatory or Inhibitory ? ». *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 77–91.

◇FRAUENFELDER, U. H., & TYLER, L. K. (Eds.). (1987). *Spoken Word Recognition*. Cambridge, MA: MIT Press.

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

FRAZIER, L. (1987). « Structure in auditory word recognition ». *Cognition*, *25*, 157–187.

FUJIMIRA, O. (1976). « Syllables as concatenated demisyllables and affixes ». *Journal of the Acoustical Society of America*, *59*(S1), S55.

GANONG, W. F. (1980). « Phonetic categorisation in auditory word perception ». *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 110–125.

GARFIELD, J. L. (Ed.). (1987). *Modularity in knowledge representation and natural-language understanding*. Cambridge, MA: MIT Press.

GARNER, W. R. (1974). *The Processing of Information and Structure*. Potomac, Md: Erlbaum.

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

GIBSON, J. J. (1979). *The ecological approach to visual perception*. Boston: Hougthon Mifflin.

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

GOODMAN, J. C., & HUTTENLOCHER, J. (1988). « Do we know how people identify spoken words? ». *Journal of Memory and Language*, *27*, 684–698.

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

HALLE, M., & VERGNAUD, J. R. (1988). *An Essay on Stress*. Cambridge, Mass.: MIT Press.

HOCKETT, C. F. (1955). *A manual of phonology*. Bloomington, IN: Indiana University Press.

HOLMES, V. M., & FORSTER, K. I. (1970). « Detection of extraneous signals during sentence recognition ». *Perception & Psychophysics*, *7*, 297–301.

HOLMES, V. M., & FORSTER, K. I. (1972). « Click localization and syntactic structure ». *Perception & Psychophysics*, *12*, 9–15.

HUGGINS, A. W. (1964). « Distortion of the temporal pattern of speech: Interruption and alternation ». *Journal of the Acoustical Society of America*, *36*, 1055–1064.

JAKOBSON, R., & WAUGH, L. (1980). *La charpente phonique du langage*. Paris: Les Éditions de minuit.

KAHN, D. (1976). *Syllable-based generalizations in English phonology*. Ph.D. thesis, Massachussetts Institute of Technology, Cambridge, MA. Doctoral dissertation.

KAYE, J. (1990). *Phonology: A cognitive view*. Cambridge, MA: MIT Press.

KEARNS, R. (1994). *Prelexical Speech Processing by mono- and bilinguals*. Ph.D. thesis, Cambridge University.

KEWLEY-PORT, D., & LUCE, P. A. (1984). « Time-varying features of initial stop consonants in auditory running spectra: A first report ». *Perception & Psychophysics*, *35*, 353–360.

◇KLATT, D. (1989). « Review of selected models in speech perception ». In W. D.

Marlsen-Wilson (Ed.), *Lexical Representation and Process*, pp. 169–226. Cambridge, Mass: MIT Press.

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

KLATT, D. H. (1980). « Speech perception: A model of acoustic- phonetic analysis and lexical access ». In R. A. Cole (Ed.), *Perception and production of fluent speech*, pp. 243–288. Hillsdale, N.J.: Erlbaum.

KOLINSKY, R., MORAIS, J., & CLUYTENS, M. (1994). « Intermediate representations in spoken word recognition: Evidence from word illusions ». Journal of Memory and Language in press.

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., & MILLER, J. D. (1978). « Speech perception by the chinchilla: Identification functions for synthetic stimuli ». *Journal of the Acoustical Society of America*, *63*, 905–917.

LADEFOGED, P., & BROADBENT, D. E. (1960). « Perception of sequence in auditory events ». *Quaterly Journal of Experimental Psychology*, *13*, 162–170.

LAEUFER, C. (1992). « Syllabification and Resyllabification in French ». In *Theoretical analyses in Romance linguistics*, pp. 18–36, Linguistic Symposium on Romance Languages, Ohio State Univeristy, 1989, Amsterdam: J. Benjamins Pub. Co.

▷LAHIRI, A., & JONGMAN, A. (1990). « Intermediate level of analysis: feature or segments? ». *Journal of Phonetics*, *18*, 435–443.

◇LAHIRI, A., & MARSLEN-WILSON, W. D. (1991). « The mental representation of lexical form: A phonological approach to the recognition lexicon ». *Cognition*, *38*, 245–294.

LAND, E. H. (1977). « The retinex theory of color vision ». *Scientific American*, *239*(6), 108–128.

LENNEBERG, E. H. (1967). *Biological foundations of language*. New York: Wiley.

LEVELT, W. J. M., & WHEELDON, L. (1994). « Do speakers have access to a mental syllabary ». *Cognition*, *50*, 239–269.

LIBERMAN, A. M., COOPER, F. S., HARRIS, K. S., & MACNEILAGE, P. F. (1963). « A Motor Theory of speech perception ». *Journal of the Acoustical Society of America*, *35*, 1114.

◇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., & MATTINGLY, I. G. (1985). « The motor theory of speech perception revised ». *Cognition*, *21*, 1–36.

LIBERMAN, A. M., MATTINGLY, I. G., & TURVEY, M. T. (1972). « Language codes and memory codes ». In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. Washington, DC: Winston and Sons.

LIBERMAN, A. M., & STUDDERT-KENNEDY, M. (1978). « Phonetic perception ». In R. Held, H. Leibowitz, & H.-L. Teuber (Eds.), *Handbook of sensory physiology: Perception*, Vol. VIII, pp. 143–178. Berlin: Springer-Verlag.

LIBERMAN, I. Y., SHANKENWEILER, D., FISHER, F. W., & CARTER, B. (1974). « Explicit syllable and phoneme segmentation in the young child ». *Journal of Experimental Child Psychology*, *18*, 201–212.

LIBERMAN, M. Y., & STREETER, L. A. (1978). « Use of nonsense-syllable mimicry in the study of prosodic phenomena ». *Journal of the Acoustical Society of America*, *63*, 231–233.

LOCKHEAD, G., & POMERANTZ, J. R. (Eds.). (1991). *The Perception of Structure*. Washington, DC: American Psychological Association.

LUCE, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.

MANN, V. A., & REPP, B. H. (1980). « Influence of vocalic context on perception of the /ch/-/s/ distinction ». *Perception & Psychophysics*, *28*, 213–228.

MARLSEN-WILSON, W. (Ed.). (1989). *Lexical Representation and Process*. Cambridge, MA: MIT Press.

▷MARLSEN-WILSON, W. D. (1984). « Function and Process in Spoken Word Recognition ». In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and Performance: Control of language Processes*, pp. 125–150. Hillsdale, N.J.: Erlbaum.

MARLSEN-WILSON, W. D. (1987). « Functional Parallelism in spoken word recognition ». *Cognition*, *25*, 71–102.

MASSARO, D. W. (1972). « Preperceptual images, processing time and perceptual units in auditory perception ». *Psychological Review*, *79*, 124–125.

MASSARO, D. W. (1974). « Perceptual units in Speech Recognition ». *Journal of Experimental Psychology*, *102*, 199–208.

◇MASSARO, D. W. (1987). « Categorical partition: A fuzzy-logical model of categorization behavior ». In S. Harnad (Ed.), *Categorical Perception: The grounwork of cognition*, pp. 254–283. Cambridge, UK: Cambridge University Press.

MASSARO, D. W., & COHEN, M. M. (1991). « Integration versus interactive activation: The joint influence of stimulus and context in perception ». *Cognitive Psychology*, *23*(4), 558–614.

MATTINGLY, I. G., & STUDDERT-KENNEDY, M. (Eds.). (1991). *Modularity and the motor theory of speech perception*. Hillsdale, NJ: Erlbaum.

◇MCCELLAND, J. L. (1991). « Stochastic Interactive proceses and the effect of context in speech perception ». *Cognitive Psychology*, *23*(1), 1–44.

◇MCCLELLAND, J. L. (1979). « On the time relations of mental processes: A framework for analysing processes in cascade ». *Psychological Review*, *86*, 287–330.

MCCLELLAND, J. L., & RUMELHART, D. E. (1986). *Parallel distributed processing: Explorations in the microstructures of cognition. Vol 2. Psychological and Biological Models.* Cambridge, Mass.: MIT Press.

MCNEILL, D., & LINDIG, K. (1973). « The perceptual reality of phonemes, syllabes, words and sentences ». *Journal of Verbal Learning and Verbal Behavior*, *12*, 419–430.

MCQUEEN, J. (1991). « The influence of the lexicon on phonetic categorization: Stimulus quality in word-final ambiguity ». *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 433–443.

◇MEHLER, J. (1981). « The role of syllables in speech processing: Infant and adult data ». *Philosophical Transactions of the Royal Society*, *295*, 333–352.

MEHLER, J., DOMMERGUES, J. Y., FRAUENFELDER, U., & SEGUI, J. (1981). « The syllable's role in speech segmentation ». *Journal of Verbal Learning and Verbal Behavior*, *20*, 298–305.

▷MEHLER, J., DUPOUX, E., & SEGUI, J. (1990). « Constraining Models of Lexical Access: The onset of word recognition ». In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: psycholinguistic and computational perspectives*, chap. 11, pp. 236–262. Cambridge, Mass: MIT Press.

MILLER, G. A. (1951). *Language and Communication*. New York, NY: McGraw-Hill.

MILLER, J. L. (1987). « Mandatory Processing in Speech Perception: A case study ». In J. L. Garfield (1987), pp. 310–322.

MILLER, J. L. (1990). « Speech Perception ». In D. N. Osherson & H. Lasnik (Eds.), *Language: An Invitation to Cognitive Science, vol. 1*, pp. 69–93. Cambridge, MA: MIT Press.

MILLER, J. L. (1994). « On the internal structure of phonetic categories: a progress report ». *Cognition*, *50*, 271–285.

MILLER, J. L., & DEXTER, E. R. (1988). « Effects of speaking rate and lexical status on phonetic perception ». *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 369–378.

MILLER, J. L., & LIBERMAN, A. M. (1979). « Some effects of later-occurring information on the perception of stop consonants and semivowels ». *Perception & Psychophysics*, *25*, 457–465.

▷MILLER, J. (1982). « Discrete versus continuous stage models of human information processing: in search of partial output ». *Journal of Experimental Psychology: Human Perception and Performance*, *8*(2), 273–296.

MILLER, J. (1983). « Can response preparation begin before stimulus recognition finishes? ». *Journal of Experimental Psychology: Human Perception and Performance*, *9*(2), 161–182.

MILLER, J., RIEHLE, A., & REQUIN, J. (1992). « Effects of preliminary perceptual output on neuronal activity of primary motor cortex ». *Journal of Experimental Psychology: Human Perception and Performance*, *18*(4), 1221–1138.

MILLER, J. (1978). « Interactions in processing segmental and suprasegmental features of speech ». *Perception & Psychophysics*, *24*(2), 175–180.

▷MILLS, C. B. (1980). « Effects of context on reaction time to phonemes ». *Journal of Verbal Learning and Verbal Behavior*, *19*, 75–83.

MORAIS, J., BERTELSON, P., CARY, L., & ALEGRIA, J. (1986). « Literacy Training and Speech Segmentation ». *Cognition*, *24*, 45–64.

MORAIS, J., CARY, L., ALEGRIA, J., & BERTELSON, P. (1979). « Does awareness of speech as a sequence of phones arise spontaneously ? ». *Cognition*, *7*, 323–331.

MORAIS, J., CONTENT, A., CARY, L., MEHLER, J., & SEGUI, J. (1989). « Syllabic segmentation and literacy ». *Language and Cognitive Processes*, *4*, 57–67.

MORAIS, J., & KOLINSKY, R. (1994). « Perception and awareness in phonological processing: the case of the phoneme ». *Cognition*, *50*, 287–297.

MYERS, T., LAVER, J., & ANDERSON, J. (Eds.). (1981). *The cognitive representation of speech*. Amsterdam: North Holland.

NORRIS, D. (1992). « Connectionism: A new breed of bottom-up model ? ». In R. G. Reilly, & N. E. Sharkey (1992), pp. 351–371.

NORRIS, D. G. (1986). « Word recognition: Context effects without priming ». *Cognition*, *22*, 93–136.

NORRIS, D. G., & CUTLER, A. (1985). « Juncture detection ». *Linguistics*, *23*, 689–705.

▷NORRIS, D. G., & CUTLER, A. (1988). « The relative accessibility of phonemes and syllables ». *Perception & Psychophysics*, *43*, 541–550.

▷OTAKE, T., HATANO, G., CUTLER, A., & MEHLER, J. (1993). « Mora or syllable? Speech segmentation in Japanese ». *Journal of Memory and Language*, *32*, 258–278.

◇PERKELL, J. S., & KLATT, D. H. (Eds.). (1986). *Invariance and Variability in Speech Processes*. Hillsdale, NJ: Erlbaum.

PETITOT, J. (1985). *Les catastrophes de la parole*. Paris: Maloine.

PISONI, D. B. (1977). « Identification and disrimination of the relative onset time of two-component tones: Implications for voiving perception in stops ». *Journal of the Acoustical Society of America*, *61*, 1352–1361.

PISONI, D. B. (1992). « Talker normalization in speech perception ». In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech Perception, Production and Linguistic Structure*, pp. 143–151. Amsterdam, Holland: IOS Press.

PISONI, D. B., & LUCE, P. A. (1986). « Trading Relations, Acoutistic Cue Integration, and Context Effects in Speech Perception ». Tech. rep. 12, Indiana University. Progress Report.

PISONI, D. B., & LUCE, P. A. (1987). « Acoustic-phonetic representations in word recognition ». *Cognition*, *25*, 21–52.

PISONI, D. B., & TASH, J. (1974). « “Same-different” reaction times to consonants, vowels and syllables ». *Journal of the Acoustical Society of America*, *55*(A), 436.

▷PITT, M. A., & SAMUEL, A. G. (1990). « Attentional Allocation during Speech Perception: How fine is the focus? ». *Journal of Memory and Language*, *29*, 611–632.

PITT, M. A., & SAMUEL, A. G. (1993). « An empirical and meta-analytic evaluation of the phoneme identification task ». *Journal of Experimental Psychology: Human Perception and Performance*, *19*(4), 699–725.

POSNER, M. (1978). *Chronometric Explorations of Mind*. Hillsdale, NJ: Erlbaum.

POSNER, M., & TAYLOR, R. (1969). « Substractive method applied to separation of visual and name components of multiletter array ». *Acta Psychologica*, *30*, 104–114.

POSNER, M. I., & MITCHELL, R. F. (1967). « Chronometric analysis of classification ». *Psychological Review*, *74*, 392–409.

POSNER, M. I., & SNYDER, C. R. R. (1975). « Facilitation and inhibition in the processing of signals ». In P. M. Rabbitt & S. Dornic (Eds.), *Attention and Performance V*. New York: Academic Press.

POTTER, R., KOPP, G., & GREEN, H. (1947). *Visible Speech*. New York: Van Nostrand.

PROCTOR, R. W., & REEVE, T. G. (1986). « A caution regarding use of the hint procedure to determine whether partial stimulus information activates responses ». *Perception & Psychophysics*, *40*(2), 110–118.

PULGRAM, E. (1970). *Syllable, Word, Nexus, Cursus*. The Hague: Mouton.

⋄PYLYSHYN, Z. W. (1984). *Computation and Cognition: Toward a foundation for Cognitive Science*. Cambridge, MA: MIT Press.

READ, C., YUN-FEI, Z., HONG-YIN, N., & BAO-QING, D. (1986). « The ability to manipulate speech sounds depends on knowing alphabetic writing ». *Cognition*, *24*, 31–45.

REBER, A. S. (1989). « Implicit Learning and Tacit Knowledge ». *Psychological Review*, *118*(3), 219–235.

REDDY, D. R., ERMAN, L. D., FENNEL, R. D., & NEELY, R. B. (1973). « The Hearsay speech understanding system: An exemple of the recognition process ». In *Proceedings of the International Conference on Artificial Intelligence*, pp. 185–194, Stanford, CA.

REILLY, R. G., & SHARKEY, N. E. (Eds.). (1992). *Connectionnist approaches to natural language processing*. Hillsdale, NJ: Erlbaum.

REMEZ, R. E. (1987). « Neural models of speech perception: a case history ». In S. Harnad (Ed.), *Categorical Perception: the groundwork of cognition*, pp. 199–225. Cambridge University Press.

▹REMINGTON, R. (1977). « Processing of phonemes in speech: A speed-accuracy study ». *Journal of the Acoustical Society of America*, *62*(5), 1279–1290.

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

⋄REPP, B. H., & LIBERMAN, A. M. (1987). « Phonetic boundaries are flexibles ». In S. Harnad (Ed.), *Categorical Perception: the groundwork of cognition*, pp. 89–111. Cambridge, UK: Cambridge University Press.

RUBIN, P., TURVEY, M. T., & GELDER, P. V. (1976). « Initial phonemes are detected faster in spoken words than in nonwords ». *Perception & Psychophysics*, *19*, 394–398.

SAMUEL, A. (1990). « The Architecture of Perception ». In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: psycholinguistic and computational perspectives*, chap. 14, pp. 295–314. Cambridge, Mass: MIT Press.

SAMUEL, A. G. (1977). « The effect of discrimination training on speech perception: Noncategorial perception ». *Perception & Psychophysics*, *22*, 321–330.

SAMUEL, A. G. (1989). « Insights from a failure of selective adaptation : Syllable-initial and syllable-final consonants are different ». *Perception & Psychophysics*, *45*, 485–493.

▷SAMUEL, A. G. (1991). « Perceptual degradation due to signal alternation: implications for auditory pattern processing ». *Journal of Experimental Psychology: Human Perception and Performance*, *17*(2), 392–403.

SAPIR, E. (1921). *Language*. New York: Harcourt Brace Jovanovich. Trad. française : *Le Langage* Paris, Payot, 1967.

SAVIN, H., & BEVER, T. (1970). « The nonperceptual reality of the phoneme ». *Journal of Verbal Learning and Verbal Behavior*, *9*, 295–302.

SEBASTIAN-GALLÉS, N., DUPOUX, E., SEGUI, J., & MEHLER, J. (1992). « Contrasting syllabic effets in catalan and spanish ». *Journal of Memory and Language*, *31*, 18–32.

SEGUI, J. (1984). « The syllable: A basic perceptual unit in speech perception ? ». In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and Performance X*, pp. 165–181. Hillsdale, NJ: Erlbaum.

▷SEGUI, J., DUPOUX, E., & MEHLER, J. (1990). « The role of the syllable in speech segmentation,phoneme identification, and lexical access ». In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: psycholinguistic and computational perspectives*, chap. 12, pp. 263–280. Cambridge, Mass.: MIT Press.

SELKIRK, E. O. (1982). « The syllable ». In van der Hulsts & Smith (Eds.), *The structure of phonological representation*, Vol. I. Dordrecht: Foris Publication.

SHAND, M. A. (1976). « Syllabic vs segmental perception: On the inability to ignore “irrelevant” stimulus parameter ». *Perception & Psychophysics*, *20*(6), 430–432.

SHATTUCK-HUFNAGEL, S. (1979). « Speech errors as evidence for a serial order mechanism in sentence production ». In W. E. Cooper & E. T. C. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, NJ: Erlbaum.

◇SHEPARD, R. N. (1972). « Psychological Representation of Speech Sound ». In E. E. David & P. B. Denes (Eds.), *Human Communication: A unified view*, pp. 67–111. New York: McGraw-Hill.

STEVENS, K. N. (1986). « Models of phonetic recognition II: A feature-based model of speech recognition ». In P. Mermelstein (Ed.), *Proceedings of the Montreal Satellite Sympsosium on Speech Recognition*, pp. 67–68. Twelfth International Congress on Acoustic.

◇STEVENS, N. K., & BLUMSTEIN, S. E. (1981). « The search for invariant acoustic correlates of phonetic features ». In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech*. Hillsdale, NJ: Erlbaum.

STOFFER, T. H. (1985). « Representation of Phrase Structure in the perception of music ». *Music Perception*, *3*(2), 191–220.

STRANGE, W. (1987). « Information for vowels in formant transitions ». *Journal of Memory and Language*, *26*(5), 550–557.

STUDDERT-KENNEDY, M., LIBERMAN, A. M., HARRIS, K. S., & COOPER, F. S. (1970). « Motor theory of speech perception: A reply to Lane's critical review ». *Psychological Review*, *77*, 234–249.

SWINNEY, D. A., & PRATHER, P. (1980). « Phonemic identification in a phoneme monitoring experiment: The variable role of uncertainty about vowel contexts ». *Perception & Psychophysics*, *27*, 104–110.

TAFT, M., & HAMBLY, G. (1985). « The influence of orthography on phonological representations in the lexicon ». *Journal of Memory and Language*, *24*, 320–325.

TAFT, M., & HAMBLY, G. (1986). « Exploring the Cohort Model of Spoken word recognition ». *Cognition*, *22*, 259–282.

▷TOMIAK, G. R., MULLENIX, J. W., & SAWUSCH, J. R. (1987). « Integral perception of phonemes: Evidence for a phonetic mode of perception ». *Journal of the Acoustical Society of America*, *81*, 755–764.

TREIMAN, R. (1983). « The structure of spoken syllables: Evidences from novel word games ». *Cognition*, *15*, 49–74.

TREIMAN, R. (1986). « The division between onsets and rimes in English syllables ». *Journal of Memory and Language*, *25*, 476–491.

TREIMAN, R., & BREAUX, A. M. (1982). « Common phoneme and overall similarity relations among spoken syllabes: their use by children and adults ». *Journal of Psycholinguistic Research*, *11*, 581–610.

TREIMAN, R., & DANIS, C. (1988a). « Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables ». *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14*, 145–152.

▷TREIMAN, R., & DANIS, C. (1988b). « Syllabification of intervocalic consonants ». *Journal of Memory and Language*, *27*, 87–104.

TREIMAN, R., STRAUB, K., & LAVERY, P. (1994). « Syllabification of bisyllabic nonwords : evidence from short-term memory errors ». *Language and Speech*, *37*(1), 45–60.

TREIMAN, R., & ZUKOWSKI, A. (1990). « Toward an Understanding of English Syllabification ». *Journal of Memory and Language*, *29*, 1990.

TYLER, L. K. (1990). « The relationship between sentential context and sensory input: comments on Connine's and Samuel's chapter ». In G. T. M. Altmann (1990), pp. 315–323.

WALLEY, A. C., SMITH, L. B., & JUSCZYK, P. W. (1986). « The role of phonemes and syllables in the perceived similarity of speech sounds for children ». *Memory & Cognition*, *14*(3), 220–229.

WARREN, P., & MARSLEN-WILSON, W. D. (1988). « Cues to lexical choice: Discriminating place and voice ». *Perception & Psychophysics*, *43*, 21–30.

WARREN, R. M. (1970). « Perceptual restauration of missing speech sounds ». *Science*, *167*, 392–393.

WARREN, R. M., & ACKROFF, J. M. (1976). « Dichotic verbal transformations and evidence of separate processors for identical stimuli ». *Nature*, *259*, 475–476.

WARREN, R. M., & WARREN, R. P. (1970). « Auditory illusions and confusions ». *Scientific American*, *223*, 30–36.

WERKER, J. F., & TEES, R. C. (1984). « Cross-language speech perception: Evidence for perceptual reorganization during the first year of life ». *Infant Behavior and Development*, *7*, 49–63.

▷WHALEN, D. H. (1989). « Vowel and consonant judgements are not independent when cued by the same information ». *Perception & Psychophysics*, *46*(3), 284–292.

WOOD, C. C. (1974). « Parallel processing of auditory and phonetic information in speech discrimination ». *Perception & Psychophysics*, *15*, 501–508.

WOOD, C. C. (1975). « Auditory and phonetic levels of processing in speech perception: Neuropsychological and information-processing analyses ». *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 3–20.

▷WOOD, C. C., & DAY, R. S. (1975). « Failure of selective attention to phonetic segments in consonant-vowel syllables ». *Perception & Psychophysics*, *17*, 346–350.

YOUNG, S. R. (1990). « Use of dialogue, pragmatics and semantics to enhance speech recognition ». *Speech Communication*, *9*, 551–564.

ZWITSERLOOD, P., SCHRIEFERS, H., LAHIRI, A., & VAN DONSELAAR, W. (1993). « The role of the syllabe in the perception of spoken Dutch ». *Journal of Experimental Psychology: Learning, Memory and Cognition*, *19*(2), 260–271.

Annexes

Annexe 1: Anova cumulant Exp. 1 et 2

```
PLAN  S8<G2>*M2*O2*X2
G = Groupe (=expérience)
X = Expérimental
O = Ordre
M = Moitié de bloc (M1 = début et M2 = fin)
G              F(1,14)=  3.68    MSE=36712.1    p=0.0757
X              F(1,14)= 33.06    MSE=1763.69    p=0.0001
X.G            F(1,14)=  0.37    MSE=1763.69    p=0.4473
O              F(1,14)=  4.77    MSE=4333.45    p=0.0465
O.G            F(1,14)=  0.52    MSE=4333.45    p=0.4827
O.X            F(1,14)=  0.07    MSE=3673.11    p=0.2048
O.X.G          F(1,14)=  0.81    MSE=3673.11    p=0.3833
M              F(1,14)=  1.90    MSE=1261.04    p=0.1897
M.G            F(1,14)=  1.14    MSE=1261.04    p=0.3037
M.X            F(1,14)=  2.01    MSE=780.337    p=0.1781
M.X.G          F(1,14)=  0.43    MSE=780.337    p=0.4774
M.O            F(1,14)=  0.20    MSE=536.895    p=0.3384
M.O.G          F(1,14)=  2.69    MSE=536.895    p=0.1232
M.O.X          F(1,14)=  0.13    MSE=877.188    p=0.2762
M.O.X.G        F(1,14)=  0.04    MSE=877.188    p=0.1556
X/M1           F(1,14)= 27.79    MSE=1421.72    p=0.0001
X/M2           F(1,14)= 18.14    MSE=1122.3     p=0.0008
X/O1           F(1,14)=  8.63    MSE=3827.91    p=0.0108
X/O2           F(1,14)= 15.85    MSE=1608.89    p=0.0014
```

Annexe 2: Mots inducteurs des expériences 4, 4bis et 5

Coda : calmant, capture, carton, certaine, culture, factice, facture, filmer, furtive, garder, lacté, magma, magnum, malsain, marquer, marteau, martyr, mortel, nectar, normal, pactole, parcours, pardon, parfum, parquet, personne, pigmé, pigment, sceptique, secteur, septembre, serpent, sourdine, subjugue, submerge, subsiste, sultan, surface, surtout, tactile, technique, turban, verbal, verdure, vertige, victime, virgule, volcan, voltige, vulgaire.

Attaque : citron, coffrage, cycliste, cyprès, débris, déclin, degré, déplaire, déprime, doublure, fabrique, faiblesse, fébrile, goudron, libraire, maigrir, maîtrise, microbe, motrice, mystère, nacrée, naufrage, navrée, névrose, patron, pétrole, public, réclame, recrue, réflexe, reflux, refrain, regret, repli, retraite, sacrée, safran, secret, sevrée, siffler, souplesse, sucrée, suffrage, supplice, suprême, tableau, tigresse, vibrant, vitrine.

Annexe 3: Mots inducteurs de l'expérience 6

rictus, fermette, vecteur, mesquine, soldeur, cordage, calmante, mortel, victime, servile, bifteck, virgule, charnel, captive, charlotte, courbette, fiscal, malgache, cornette, parfaite, charpente, verbal, sceptique, tardif, valseur, calvaire, nordique, sulfate, docteur, formule, dolmen, bulgare, secteur, muscade, mascotte, marquage, captif, courtine, bismuth, certaine, berceuse, malsaine, parlante, dorsal, formel, thermos, vecteur, fictive, dispute, barman, voltige, pactole, carbone, facteur, marquante, tardive, disquaire, largesse, golfeur, subtil, capsule, cheftaine, lecteur, formol, perfide, capture, bascule, catcheur, fortune, culbute.

Annexe 4: Anova de l'expérience 6

```
Analyse par sujets : PLAN  S10<I4>*P4
I = Induction attention
P = Position de la cible
Temps de réaction
I              F(3,36)=  1.11    MSE=42720.1    p=0.3578
P              F(3,108)=109.3    MSE=3316.56    p=0.0000
```

```
   P.I            F(9,108)=  8.7    MSE=3316.56    p=0.0000
   I1 I2.P1 P2    F(1,18)= 42.39    MSE=1356.42    p=0.0000
   I1 I3.P1 P3    F(1,18)= 10.28    MSE=2854.09    p=0.0049
   I1 I4.P1 P4    F(1,18)= 48.17    MSE=1894.48    p=0.0000
   I2 I3.P2 P3    F(1,18)=  6.25    MSE=4077.79    p=0.0223
   I2 I4.P2 P4    F(1,18)= 35.52    MSE=3316.98    p=0.0000
   I3 I4.P3 P4    F(1,18)= 21.50    MSE=2989.59    p=0.0002
   I1,I2 I3 I4.P1,P2 P3 P4  F(1,36)= 23.16  MSE=2294.72
p=0.0000
   I2,I1 I3 I4.P2,P1 P3 P4  F(1,36)= 16.70  MSE=3747.94
p=0.0002
   I3,I2 I1 I4.P3,P2 P1 P4  F(1,36)=  6.03  MSE=3475.05
p=0.0190
   I4,I2 I3 I1.P4,P2 P3 P1  F(1,36)= 33.49  MSE=3748.55
p=0.0000
   I1 I2.P3 P4    F(1,18)=  4.35    MSE=1938.46    p=0.0515
   I1 I3.P2 P4    F(1,18)=  3.90    MSE=6386.72    p=0.0638
   I1 I4.P2 P3    F(1,18)=  0.32    MSE=4255.39    p=0.4214
   I2 I3.P1 P4    F(1,18)=  0.73    MSE=4669.1     p=0.4041
   I2 I4.P1 P3    F(1,18)=  3.74    MSE=4259.53    p=0.0690
   I3 I4.P1 P2    F(1,18)=  0.85    MSE=1800.22    p=0.3687
   Erreurs:
   I              F(3,36)=  1.24    MSE=0.6257   p=0.3095
   P              F(3,108)= 16.0    MSE=0.5887   p=0.0000
   P.I            F(9,108)=  1.4    MSE=0.5887   p=0.1721
   I1 I2.P1 P2    F(1,18)=  0.51    MSE=0.1944   p=0.4843
   I1 I3.P1 P3    F(1,18)=  0.40    MSE=0.25     p=0.4650
   I1 I4.P1 P4    F(1,18)=  0.76    MSE=1.1778   p=0.3948
   I2 I3.P2 P3    F(1,18)=  0.00    MSE=0.1111   p=1.0000
   I2 I4.P2 P4    F(1,18)=  7.42    MSE=0.7583   p=0.0139
   I3 I4.P3 P4    F(1,18)=  2.50    MSE=0.6389   p=0.1313
   I1,I2 I3 I4.P1,P2 P3 P4  F(1,36)=  0.15  MSE=0.2345
p=0.2992
   I2,I1 I3 I4.P2,P1 P3 P4  F(1,36)=  2.07  MSE=0.4104
p=0.1589
   I3,I2 I1 I4.P3,P2 P1 P4  F(1,36)=  0.10  MSE=0.3363
p=0.2463
   I4,I2 I3 I1.P4,P2 P3 P1  F(1,36)=  3.48  MSE=1.3734
p=0.0703
   I1 I2.P3 P4    F(1,18)=  1.57    MSE=1.2917   p=0.2262
   I1 I3.P2 P4    F(1,18)=  0.29    MSE=1.3611   p=0.4032
   I1 I4.P2 P3    F(1,18)=  0.23    MSE=0.4333   p=0.3627
   I2 I3.P1 P4    F(1,18)=  0.95    MSE=0.6583   p=0.3426
```

```
   I2 I4.P1 P3    F(1,18)=  0.00    MSE=0.05      p=1.0000
   I3 I4.P1 P2    F(1,18)=  0.72    MSE=0.1389    p=0.4073
   Analyse par items: PLAN  S<P4>*I4
   I              F(3,105)= 17.6    MSE=2862.41    p=0.0000
   P              F(3,35)= 19.90    MSE=17022.5    p=0.0000
   P.I            F(9,105)=  9.6    MSE=2862.41    p=0.0000
   I1 I2.P1 P2    F(1,18)= 40.24    MSE=1428.8     p=0.0000
   I1 I3.P1 P3    F(1,18)=  9.18    MSE=3197.5     p=0.0072
   I1 I4.P1 P4    F(1,17)= 43.96    MSE=1967.01    p=0.0000
   I2 I3.P2 P3    F(1,18)=  9.44    MSE=2698.85    p=0.0066
   I2 I4.P2 P4    F(1,17)= 82.42    MSE=1354.14    p=0.0000
   I3 I4.P3 P4    F(1,17)= 24.06    MSE=2530.97    p=0.0001
   I1,I2 I3 I4.P1,P2 P3 P4  F(1,35)= 12.72  MSE=4239.47
p=0.0011
   I2,I1 I3 I4.P2,P1 P3 P4  F(1,35)= 60.28  MSE=998.238
p=0.0000
   I3,I2 I1 I4.P3,P2 P1 P4  F(1,35)=  5.08  MSE=3936.41
p=0.0306
   I4,I2 I3 I1.P4,P2 P3 P1  F(1,35)= 50.92  MSE=2275.51
p=0.0000
   I1 I2.P3 P4    F(1,17)=  2.69    MSE=2972.44    p=0.1194
   I1 I3.P2 P4    F(1,17)=  3.40    MSE=6946.39    p=0.0827
   I1 I4.P2 P3    F(1,18)=  0.28    MSE=4861.99    p=0.3968
   I2 I3.P1 P4    F(1,17)=  1.77    MSE=1817.18    p=0.2010
   I2 I4.P1 P3    F(1,18)= 11.99    MSE=1330.13    p=0.0028
   I3 I4.P1 P2    F(1,18)=  0.47    MSE=3266.48    p=0.4983
   Erreurs:
   I              F(3,105)=  1.5    MSE=0.5208    p=0.2136
   P              F(3,35)= 17.07    MSE=0.6514    p=0.0000
   P.I            F(9,105)=  1.8    MSE=0.5208    p=0.0677
   I1 I2.P1 P2    F(1,18)=  1.20    MSE=0.0833    p=0.2878
   I1 I3.P1 P3    F(1,18)=  0.47    MSE=0.2111    p=0.4983
   I1 I4.P1 P4    F(1,17)=  4.37    MSE=0.2487    p=0.0519
   I2 I3.P2 P3    F(1,18)=  0.00    MSE=0.1056    p=1.0000
   I2 I4.P2 P4    F(1,17)= 12.22    MSE=0.5242    p=0.0028
   I3 I4.P3 P4    F(1,17)=  1.68    MSE=1.1441    p=0.2122
   I1,I2 I3 I4.P1,P2 P3 P4  F(1,35)=  0.06  MSE=0.6535
p=0.1921
   I2,I1 I3 I4.P2,P1 P3 P4  F(1,35)=  2.76  MSE=0.3158
p=0.1056
   I3,I2 I1 I4.P3,P2 P1 P4  F(1,35)=  0.05  MSE=0.797
p=0.1756
```

```
   I4,I2 I3 I1.P4,P2 P3 P1  F(1,35)= 17.36  MSE=0.317
p=0.0002
   I1 I2.P3 P4    F(1,17)=  2.21    MSE=1.0029    p=0.1554
   I1 I3.P2 P4    F(1,17)=  0.26    MSE=1.6222    p=0.3833
   I1 I4.P2 P3    F(1,18)=  0.24    MSE=0.4167    p=0.3699
   I2 I3.P1 P4    F(1,17)=  0.88    MSE=0.7977    p=0.3613
   I2 I4.P1 P3    F(1,18)=  0.00    MSE=0.0556    p=1.0000
   I3 I4.P1 P2    F(1,18)=  0.62    MSE=0.1611    p=0.4413
```

Annexe 5: Matériel espagnol pour la détection de phonème attentionnelle

Mots

Attaque :

Tests : sagrado, sobrina, copla, capricho, microbio, libro, metralla, reprimir, preclaro, flagrante, triplicar, proclive, triple, progre, problema, programa.

Inducteurs : ciclo, fabrica, cetrino, degradar, siglo, libreta, bucle, febrero, medrar, sacristan, cobre, replica, moflete, soprano, febril, lobrego, cipres, tigre, seglar, lograr, ciclico, reproche, vitrina, refrescar, sofrito, replegar, ladrar, rebrotar, lagrima, cifra, cable, vibrar, cabra, cedro, quebrantar, madrugar, decreto, sublime, depresion, madrigal, padrino, patron, sidra, secreto, migra, labrador, biblico, fibra, metrico.

Coda :

Tests : cansado, segmento, ruptura, lince, nectar, reptil, dogma, rectitud, fluctua, blonda, fragmento, tracto, glandula, fractura, cripta, tractor.

Inducteurs : pigmeo, gamba, factor, septimo, captar, lapso, campestre, furtivo, magno, bacteria, capturar, martir, victima, lastima, doctor, sensacion, magnitud, sector, cadmio, sublingual, raptado, cultivo, magnesio, rector, magma, masticar, capsula, subsuelo, candor, tecnico, lacteo, raptar, pigmento, lectura, vector, facturar, tactil, subcostal, victoria, doctrina, culpa, subsanar, ductil, subsistir, jactarse, digno, sigma, subvencion, secta.

Non-Mots

Attaque

Tests : coflinciar, piDriente, teGleral, piTronal, toclanto, cibreda, maplecho, muTrisar. froBlante, cleProna, traPleno, driclivo, glopluron, treBlinar, ploGlifar, ploGlidal.

Inducteurs : taclefon, chabreto, cadrafa, ruglutel, tapresior, siproche, fidrante, sibrenor, picraton, codrinar, pegraspar, cotruico, cibronal, digropar, pibrerol, tapladar, chubrigo, poclino, maFreto, chuFruisa, tuPligo, teGrina, sodruogal, jodrastor, ficronten, pibrante, doclidion, , puglarda, tufra, soblentia, cotracha, picrismon, digruaval, tebrentol, quedrieson, topral, sebritua, sebrador, caflegil, sigrear, chibruito, soblefar, puglosia, nibreras, setriza, sibreto, batronciar, tipresil, fablico, sebra.

Coda

Tests : cemboilga, deptonia, sobdula, ralvinuar, tuntoba, cuncero, dunsielar, ceptoril. plondinar, cloptuicon, plestomia, pructosa, truptano, drectural, droctivar, bliptonsar.

Inducteurs : deptorial, pagmorsia, munsaza, dopterdo, redmida, dempostro, sipturdor, fistucar, pugmesol, dectreba, randina, deptordil, poctora, roctona, segniral, poctular, cirtoval, tigmiense, birdiela, pusgotriel, dolties, sombionar, colpino, mocturon, vacmaler, buctoria, sagmelar, cupsorte, relpordar, saptencar, roptacor, ralparron, mirtila, gembora, foctaso, murlido, rignedio, cuntida, pagnodiur, muctilda, siptala, mapsola, pindiural, dectueso, tistomal, chaptemon, sobledir, repsulon, pagnida, dagtordil.

Annexe 6: Matériel américain (Inducteurs) pour la détection de phonème attentionnelle (exp. 7 et 8)

Type "coda" :
bandit (b'@n-d|t), standardize (st'@n-dX-d`Yz), falcon (f'@l-k|n), comfort (k'^m-fXt), mansion (m'@n-C|n), bramble (br'@m-bL), synthesis (s'In-T|-s|s), lantern (l'@n-tXn), panda (p'@n-dx), sensible (s'En-s|-bL), fulminate (f'Ul-mx-n`et), hinder (h'In-dX), trumpet (tr'^m-p|t), cancer (k'@n-sX), filter (f'Il-tX), brandy (br'@n-di), culminate (k'^l-mx-n`et), candidate (k'@n-d|-d`et), number (n'^m-bX), winter (w'In-tX), cultivate (k'^l-t|-v`et), banter (b'@n-tX), thunder (T'^n-dX), vulgar (v'^l-gX), sulphurous (s'^l-fyU-rxs), scandal (sk'@n-dL), hamlet (h'@m-l|t), salvage (s'@l-vIJ), velvet (v'El-vxt), remnant (r'Em-n|nt), calcium (k'@l-si-xm), culture (k'^l-CX), dentist (d'En-t|st), chimney (C'Im-ni), scalpel (sk'@l-pL), tumble (t'^m-bL), sentiment (s'En-tx-mxnt), sultan (s'^l-tN), lumber (l'^m-bX), bundle (b'^n-dL), sample (s'@m-pL), fantasy (f'@n-t|-si), helmet (h'El-mxt), campus (k'@m-pxs), crumple (kr'^m-pL), frantic (fr'@n-tIk), silver (s'Il-vX), dampen (d'@m-pxn), sympathize (s'Im-px-T`Yz), tension

(t'En-C|n), transitive (tr'@n-s|-tIv).

Type “attaque” :

talon (t'@-l|n), sinister (s'I-n|s-tX), calibrate (k'@-l|-br`et), ballot (b'@-l|t), minute (m'I-n|t), clamor (kl'@-mX), skeleton (sk'E-l|-tN), punish (p'^-nIS), lemon (l'E-mxn), pelican (p'E-lI-k|n), planet (pl'@-n|t), tenant (t'E-n|nt), banister (b'@-n|s-tX), limit (l'I-m|t), manifest (m'@-nx-f`Est), melon (m'E-l|n), stomach (st'^-mxk), granular (gr'@n-yU-lX), validate (v'@-lx-d`et), banner (b'@-nX), summit (s'^-mxt), panel (p'@-nL), summarize (s'^-mX-`Yz), damage (d'@-mIJ), malady (m'@-lx-di), cannibal (k'@-nx-bL), shimmer (S'I-mX), timid (t'I-mxd), relic (r'E-lIk), tremor (tr'E-mX), money (m'^-ni), dinner (d'I-nX), felon (f'E-l|n), channel (C'@-nL), million (m'I-li-xn), senator (s'E-n|-tX), jealous (J'E-lxs), camera (k'@-mX-x), salad (s'@-lxd), calendar (k'@-l|n-dX), spinach (sp'I-nIC), palace (p'@-lxs), chemist (k'E-m|st), diligent (d'I-lx-J|nt), democrat (d'E-mx-kr`@t), talent (t'@-l|nt), color (k'^-lX), camel (k'@-mL), gullet (g'^-l|t), stellar (st'E-lX), benefit (b'E-nx-f`It).

Annexe 7: Instructions de détection de phonème pour les américains

In the following experiment, you will have to detect speech sounds (targets) inside spoken words. You will be listening to a continuous list of words through the headphones. Before each word, a letter (e.g. “N”), standing for the target sound, will be displayed at the center of the computer screen. If you hear the target sound in the word, you have to press the ’YES’ button (The *right* one) *as fast as possible*. Some words will not contain the target, however. In this case, depress the ’NO’ button (the *left* one).

It is important that you respond as fast as possible because we are measuring your reaction time to detect the sound. Respond as soon as you detect the sound, even if the word is not finished. But, do not press the button BEFORE you hear the sound.

If you make too many errors, we will not be able to use your data. Making a few errors is normal, so, do not deconcentrate when you happen to make one.

Last but not least, let us emphasize that it is the sound that you must detect, not the letter. For example the letter ’K’ stands for the sound appearing at the beginning of ’cat’ or at the end of ’mock’. When you feel ready, press the spacebar to perform a short training.

Annexe 8: Instruction de détection de phonème pour sujets Français (exp.8)

Dans cette expérience, vous entendrez une liste de mots ANGLAIS, à raison d'un toutes les 5 secondes. Précédant chacun de ces mots, une lettre sera affichée au centre de l'écran. Quand cette lettre apparaîtra, vous devrez penser au SON correspondant: Par exemple:

P comme dans pot, supper, stop
K comme dans cat, token, make
L comme dans low, mellow, all

Puis, si vous entendez ce son dans le mot qui suit, vous devrez appuyer sur le bouton de réponse OUI. Certains mots ne contiennent pas le son correspondant, dans ce cas vous appuyez sur le bouton NON.

Remarquez qu'il n'est pas important que vous connaissiez le mot anglais et en particulier son orthographe puisque c'est le SON qu'on vous demande de détecter (cf 'k' dans 'cat').

Nous mesurons votre temps de réaction, c'est pourquoi il faut que vous répondiez AUSSI VITE que possible, sans attendre la fin du mot, et surtout sans faire d'erreurs. Tâchez d'être attentif, et en particulier de ne pas rater les lettres qui apparaissent à l'écran. Si jamais vous faites une erreur, NE VOUS DECONCENTREZ PAS, et préparez-vous pour l'essai suivant.

Annexe 9: Anova de l'expérience 11 (« PATABADA »)

```
PLAN  S8<I2*O2*F2>*C2
C = condition (Bloc Expérimental vs bloc contrôle)
F = Trait (F1=voisement ; F2=place)
O = Ordre (Bloc Expérimental ou Bloc Contrôle d'abord)
I = Index (I1= congruent; I2= non congruent)
case: Moyenne (SDM/n)
C1I1O1F1: 683.1 (27.4/8)
C2I1O1F1: 746.3 (35.9/8)
C1I2O1F1: 706.4 (19.7/8)
C2I2O1F1: 724.9 (12.6/8)
C1I1O2F1: 734.1 (28.7/8)
C2I1O2F1: 700.6 (10.5/8)
```

C1I2O2F1 : 728.4 (17.7/8)
C2I2O2F1 : 733.3 (14.6/8)
C1I1O1F2 : 695.7 (19.8/8)
C2I1O1F2 : 763.5 (21.9/8)
C1I2O1F2 : 698.5 (21.3/8)
C2I2O1F2 : 737.1 (41.2/8)
C1I1O2F2 : 738.6 (14.6/8)
C2I1O2F2 : 743.4 (24.1/8)
C1I2O2F2 : 697.9 (29.0/8)
C2I2O2F2 : 707.5 (35.1/8)

Temps de réaction

F	F(1,56)=	0.04	MSE=7617.96	p=0.1578
O	F(1,56)=	0.05	MSE=7617.96	p=0.1761
O.F	F(1,56)=	0.12	MSE=7617.96	p=0.2697
I	F(1,56)=	0.34	MSE=7617.96	p=0.4378
I.F	F(1,56)=	1.09	MSE=7617.96	p=0.3010
I.O	F(1,56)=	0.05	MSE=7617.96	p=0.1761
I.O.F	F(1,56)=	0.40	MSE=7617.96	p=0.4703
C	F(1,56)=	6.48	MSE=2331.15	p=0.0137
C.F	F(1,56)=	0.98	MSE=2331.15	p=0.3265
C.O	F(1,56)=	8.77	MSE=2331.15	p=0.0045
C.O.F	F(1,56)=	0.07	MSE=2331.15	p=0.2077
C.I	F(1,56)=	0.20	MSE=2331.15	p=0.3436
C.I.F	F(1,56)=	0.07	MSE=2331.15	p=0.2077
C.I.O	F(1,56)=	2.94	MSE=2331.15	p=0.0919
C.I.O.F	F(1,56)=	0.51	MSE=2331.15	p=0.4781
C/F1	F(1,28)=	1.37	MSE=2050.71	p=0.2517
I/F1	F(1,28)=	0.14	MSE=6056.76	p=0.2889
C.I/F1	F(1,28)=	0.02	MSE=2050.71	p=0.1115
C.O/F1	F(1,28)=	5.93	MSE=2050.71	p=0.0215
I.O/F1	F(1,28)=	0.10	MSE=6056.76	p=0.2458
C.I.O/F1	F(1,28)=	3.36	MSE=2050.71	p=0.0775
C/F2	F(1,28)=	5.58	MSE=2611.58	p=0.0254
I/F2	F(1,28)=	1.09	MSE=9179.17	p=0.3054
C.I/F2	F(1,28)=	0.23	MSE=2611.58	p=0.3648
C.O/F2	F(1,28)=	3.24	MSE=2611.58	p=0.0826
I.O/F2	F(1,28)=	0.31	MSE=9179.17	p=0.4179
C.I.O/F2	F(1,28)=	0.44	MSE=2611.58	p=0.4875

Erreurs

F	F(1,56)=	0.38	MSE=0.0052	p=0.4599
O	F(1,56)=	0.06	MSE=0.0052	p=0.1926

O.F	F(1,56)=	2.70	MSE=0.0052	p=0.1060
I	F(1,56)=	1.17	MSE=0.0052	p=0.2840
I.F	F(1,56)=	0.07	MSE=0.0052	p=0.2077
I.O	F(1,56)=	0.09	MSE=0.0052	p=0.2347
I.O.F	F(1,56)=	3.21	MSE=0.0052	p=0.0786
C	F(1,56)=	7.61	MSE=0.0034	p=0.0078
C.F	F(1,56)=	0.09	MSE=0.0034	p=0.2347
C.O	F(1,56)=	1.00	MSE=0.0034	p=0.3216
C.O.F	F(1,56)=	3.53	MSE=0.0034	p=0.0655
C.I	F(1,56)=	0.07	MSE=0.0034	p=0.2077
C.I.F	F(1,56)=	0.09	MSE=0.0034	p=0.2347
C.I.O	F(1,56)=	0.77	MSE=0.0034	p=0.3840
C.I.O.F	F(1,56)=	2.12	MSE=0.0034	p=0.1510
C/F1	F(1,28)=	5.19	MSE=0.0031	p=0.0305
I/F1	F(1,28)=	0.44	MSE=0.0038	p=0.4875
C.I/F1	F(1,28)=	0.00	MSE=0.0031	p=0.0000
C.O/F1	F(1,28)=	4.59	MSE=0.0031	p=0.0410
I.O/F1	F(1,28)=	1.51	MSE=0.0038	p=0.2294
C.I.O/F1	F(1,28)=	3.01	MSE=0.0031	p=0.0937
C/F2	F(1,28)=	2.75	MSE=0.0037	p=0.1084
I/F2	F(1,28)=	0.72	MSE=0.0065	p=0.4033
C.I/F2	F(1,28)=	0.15	MSE=0.0037	p=0.2985
C.O/F2	F(1,28)=	0.35	MSE=0.0037	p=0.4411
I.O/F2	F(1,28)=	1.72	MSE=0.0065	p=0.2003
C.I.O/F2	F(1,28)=	0.15	MSE=0.0037	p=0.2985

Table des matières