

Une étude du décours temporel de la prononciation des mots

Christophe Pallier

Laboratoire de sciences cognitives et psycholinguistique,

EHESS-CNRS URA 1198, 54 bd Raspail 75006 Paris.

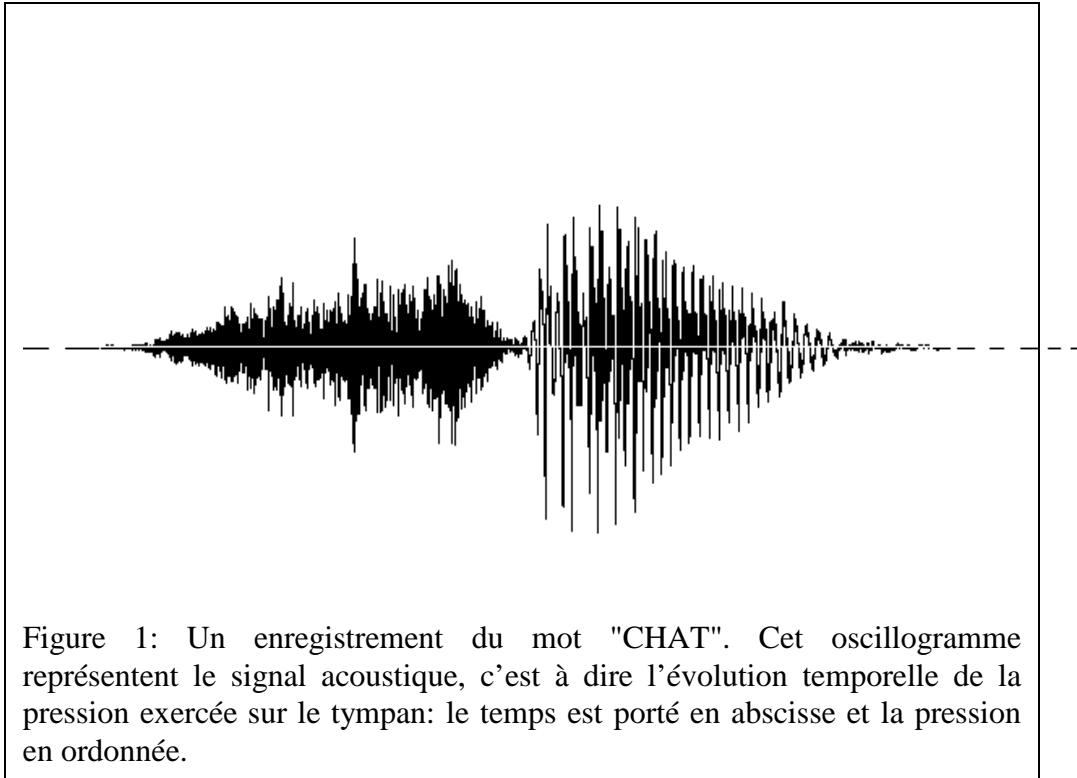
E-mail: pallier@lscp.ehess.fr

Article paru dans les Annales de la fondation Fyssen 1998, n. 13, p. 76-80

Imaginez la situation suivante : une personne est assise, un casque stéréo posé sur les oreilles ; un mot est alors prononcé et la personne doit le répéter le plus rapidement possible. Une fraction de seconde seulement est nécessaire pour effectuer une tâche aussi simple. Pourtant, il faut bien avouer que les processus qui se déroulent dans le cerveau et qui sous-tendent ce type d'action demeurent encore largement mystérieux. C'est à mieux les comprendre que sont consacrés les travaux sur la perception et la production de la parole.

Les toutes premières étapes de la perception sont relativement bien cernées. La parole est transmise à l'auditeur sous forme d'une onde acoustique, c'est à dire de variation temporelle de la pression de l'air (cf. Fig.1). Celle-ci met en mouvement le tympan qui la transmet mécaniquement au fluide qui se trouve dans la cochlée. Dans cet organe, le signal mécanique est converti en signal nerveux qui, par le nerf auditif, est envoyé vers une série de structures sous-corticales pour finalement aboutir au cortex auditif (cf. B.J.C. Moore (1997) pour une présentation générale de ces premières étapes du traitement des sons). Ensuite, plusieurs parties du cortex cérébral se chargent de transformer le message auditif en commandes articulatoires qui vont gouverner les muscles responsables de la production de la parole. Depuis peu, les méthodes d'imagerie cérébrale (Tomographie à émission de positons, Résonance magnétique fonctionnelle; cf. Dehaene (1997)) permettent de visualiser les aires cérébrales mises en jeu lors du traitement du langage (Démonet, 1997). Toutefois, outre le fait que l'interprétation des images d'activation du cerveau est délicate, savoir quelles zones sont activées par une tâche donnée (p. ex. répéter un mot) ne permet pas de déduire directement ce qu'elles font¹. De plus, les méthodes d'imagerie cérébrales qui permettent une localisation assez précise des activations ont une mauvaise résolution temporelle, c'est à dire que l'activité cérébrale est intégrée sur plusieurs secondes voire plusieurs minutes. En conséquence, elles ne sont pas idéales pour étudier des phénomènes qui se déroulent en une fraction de seconde, tels que la perception ou la prononciation des mots. C'est pourquoi les théories actuelles sont spécifiées à un niveau assez abstrait, c'est à dire se limitent à proposer des schéma fonctionnels de traitement de l'information, sans les relier à leur substrat neural ; ce sera le cas de la question théorique de production de la parole que nous discuterons plus loin dans cet article.

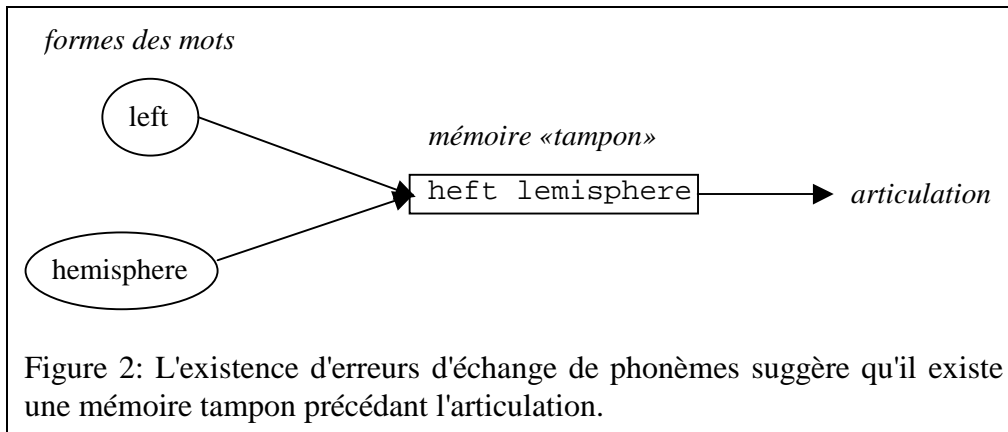
¹ Ce sera seulement après de nombreux recoupements entre images obtenues dans différentes tâches qu'on pourra mieux interpréter les images (Dehaene, 1997).



Le signal acoustique de parole et les commandes musculaires de l'articulation appartiennent à deux domaines physiques complètement différents et il n'y a pas de relation simple, univoque, entre ces deux domaines. D'ailleurs les détails de la réalisation acoustique d'un mot répété, sont, à coup sûr, très différents de l'original, ne serait-ce qu'à cause de la différence de voix. Il est clair que ce qui se passe dans le cerveau a peu de choses à voir avec l'enregistrement sur une bande magnétique et la relecture de cette bande. Dans ce dernier cas, le signal acoustique est enregistré de façon détaillée, sous forme de variation de l'amplitude d'un champ magnétique. Là, la correspondance est immédiate, et c'est pourquoi l'enregistrement "sonne" comme l'original (sauf si la chaîne HIFI est de mauvaise qualité!).

Le cerveau emploie donc d'autres moyens que la mémorisation analogique détaillée de la bande magnétique pour reconnaître et prononcer les mots. La plupart des psycholinguistes admettent que le cerveau "encode" les mots sous forme d'une représentation abstraite, typiquement une suite de phonèmes (En première approximation, les phonèmes sont les sons représentés par les lettres de l'alphabet ; par exemple, le mot "petit" est représenté par une suite de quatre phonèmes : "p", "e", "t" et "i")². Selon cette conception, pour reconnaître les mots, le cerveau doit d'abord extraire les phonèmes du signal acoustique de parole ; pour les prononcer, il doit convertir la suite de phonèmes associée en articulation.

² En français, contrairement par exemple à l'espagnol, de nombreuses lettres ne se prononcent pas et la correspondance lettre-son n'est pas systématique. D'autre part, signalons que d'autres cultures utilisent des écritures où les signes écrits correspondent plutôt aux syllabes (p. ex kana japonais), ou aux mots entiers (p.ex. en chinois) (Coulmas, 1989).



L'un des arguments en faveur de l'idée que les mots sont représentés comme des suites de phonèmes, provient de l'observation des erreurs de production de la parole : un grand nombre d'entre elles consistent en l'échange de deux phonèmes, par exemple, "left hemisphere" devenant "heft lemisphere" (Fromkin, 1973). L'existence de ce type d'erreur suggèrent non seulement que les mots sont représentés comme des suites de phonèmes, mais aussi que les mots ne sont pas prononcés directement : il doit exister une "mémoire tampon", entre les mots et l'articulation, où la phrase avec les phonèmes échangés est représentée (cf. Figure 2). Levelt (1989) a proposé une théorie générale de la production de la parole dans laquelle cette mémoire est nommée "tampon articulatoire". Dans la suite de cet article, nous allons nous poser la question suivante à propos de cette mémoire tampon : *les phonèmes d'un mot à prononcer la remplissent-elle de façon séquentielle, ou bien sont-ils insérés en "parallèle", tous en même temps?*

Un modèle "séquentiel" prédit une asymétrie entre le début et la fin du mot, alors qu'un modèle "parallèle" ne prédit pas de telle asymétrie. Une expérience d'Antje Meyer (1990) a tenté de départager ces deux types de modèles. Les participants devaient apprendre une correspondance arbitraire entre quatre figures et quatre mots. Il y avait trois possibilités : dans le premier cas, les quatre mots commençaient par la même syllabe ; dans le second ils finissaient par la même syllabe ; et, finalement, dans le troisième cas, ils ne partageaient pas de syllabe en commun. Durant l'expérience proprement dite, on montrait successivement les figures aux participants, et ceux-ci devaient nommer aussi rapidement que possible le mot associé. Un ordinateur mesurait la latence de dénomination, c'est à dire la durée entre la présentation de la figure et le début de la prononciation par le participant. Antje Meyer a observé que les latences moyennes étaient plus rapides dans le 1^{er} cas (où les mots partageaient la première syllabe) que dans les deux autres. En fait, le second cas était aussi lent que le dernier : connaître à l'avance la seconde syllabe du mot à prononcer n'aidait pas à le produire plus rapidement.

Les résultats de Meyer sont les seuls, à notre connaissance, à indiquer que l'insertion des phonèmes dans le tampon articulatoire se fait de manière séquentielle. C'est pourquoi il nous a semblé qu'il fallait essayer d'examiner cette question en employant d'autres paradigmes expérimentaux. Dans un premier temps, nous avons pensé réaliser l'expérience suivante : elle consisterait à faire prononcer à des participants les noms d'objet dessinés, et, quelque fois, à leur indiquer à l'avance la dernière syllabe de ce nom. Les durées entre le moment de présentation de l'objet et le début de la prononciation seraient mesurées, puis moyennées et comparées dans les situations où la

dernière syllabe était connue ou inconnue. Si connaître la dernière syllabe aide la préparation de la production, alors on s'attend à ce que les durées soient inférieures dans la première situation.

Cette expérience présente toutefois un défaut : il est probable que présenter la seconde syllabe à l'avance aide à retrouver le nom associé à l'objet présenté. Imaginez-vous devant l'image d'un sextant, avec le mot sur le "bout de la langue" ; savoir qu'il fini par "tant" vous aiderait sans doute à retrouver le mot. Donc la facilitation éventuellement observée pourrait être attribuée à l'étape de récupération du nom, plutôt qu'à l'étape de préparation de sa prononciation. C'est pourquoi l'expérience que nous avons réalisé, en collaboration avec Albert Costa et Nuria Sebastian de l'Université de Barcelone, est un petit peu plus complexe (elle est décrite en détails dans Costa, Sebastian et Pallier (soumis)). L'expérience, inspirée par un paradigme de préparation motrice inventé par Jeff Miller (1982), se déroulait de la façon suivante :

Dans un premier temps, les participants apprenaient à associer, un à un, quatre figures à quatre mots. Les figures étaient des carrés ou des ronds, vert ou rouge. Les mots étaient des noms communs espagnols, formés de deux syllabes consonne-voyelle (la Table 1 donne trois exemples d'associations figures-mots). L'expérience proprement dite consistait en une succession d'"essais" qui duraient chacun environ trois secondes : au début de chaque essai, une des quatre figures était présentée sur l'écran d'un ordinateur et le participant devait prononcer le mot correspondant le plus rapidement possible: l'ordinateur mesurait le temps de latence de dénomination (c'est la variable d'intérêt), puis passait à l'essai suivant.

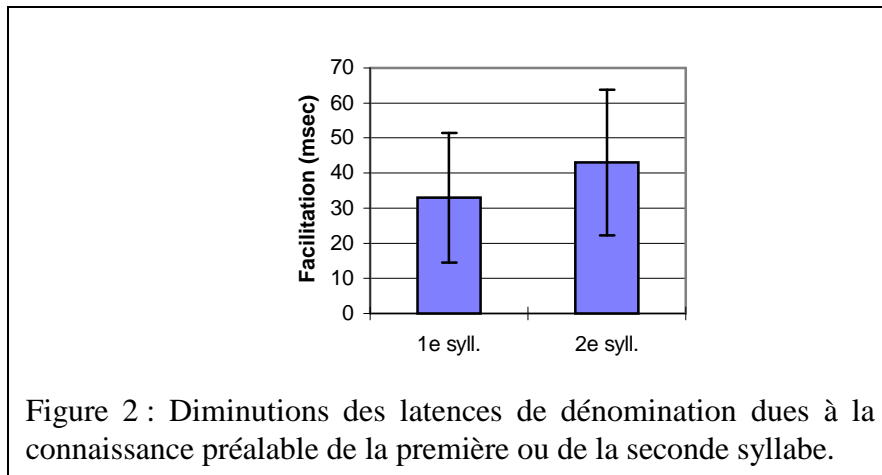
La figure s'affichait en fait en deux temps : d'abord sa forme (carré ou disque) était dessinée, puis 500 millisecondes plus tard la couleur (rouge ou vert) remplissait la forme. C'est seulement quand la couleur apparaissait que le sujet pouvait savoir quel mot il fallait prononcer. Cependant la forme pouvait être informative: dans certaines associations (dites de type "1"), la forme permettait de prévoir la première syllabe du mot cible ; dans d'autre associations (de type "2"), la forme permettait de prévoir la seconde syllabe du mot cible ; enfin, dans le dernier type d'associations (de type "C"), la forme ne permettait de prévoir aucune syllabe du mot (cependant elle permettait de présélectionner deux mots parmi quatre).

Table 1 Les trois types d'associations entre figures et mots à prononcer

Type d'association	figures			
	carré rouge	carré vert	disque rouge	disque vert
1	vida	vino	peso	pera
2	peso	vaso	mina	cuna
C	mina	peso	cuna	vaso

Trente-deux sujets espagnols, étudiants de l'université de Barcelone, ont participé à l'expérience. Chaque participant effectuait deux blocs d'une centaine d'essais; la moitié d'entre eux, formant le groupe 1, ont passé un bloc de type " 1 " et un bloc de type " C "; l'autre moitié, formant le groupe 2, ont passé un bloc de type " 2 " et un bloc de type C³.

³ Le bloc de type C était toujours obtenu en utilisant les même mots que dans l'autre bloc, 1 ou 2, mais en changeant l'association figure-mots (cf. Table 1); De plus,



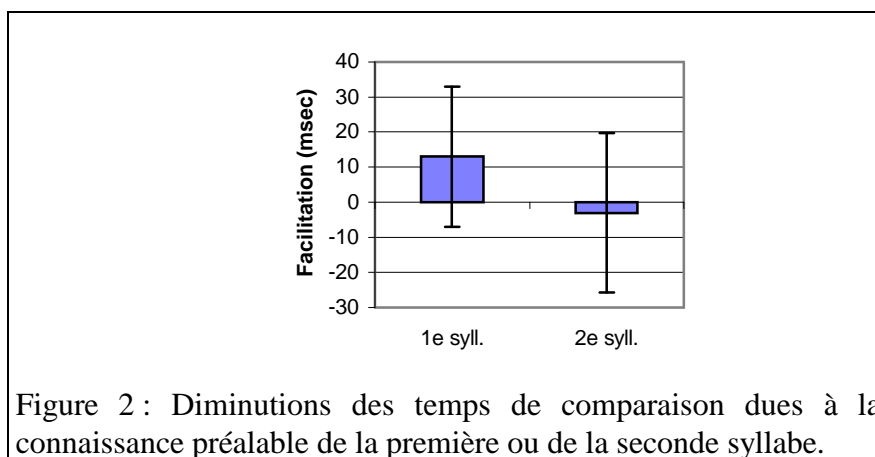
Pour chaque sujet, on a calculé la différence entre les latences de dénomination moyennes dans le bloc dit “ expérimental ” (1 ou 2) et dans le bloc dit “ contrôle ” (C). Les moyennes par sujets de ces différences sont affichées sur la Figure 2 (Syll. 1 = Groupe 1 ; syll. 2 = Groupe 2). On constate que connaître à l’avance la première, ou la seconde, syllabe d’un mot à prononcer permet d’accélérer la préparation de celui-ci (des test de T confirment la significativité statistique à 5%).

L’effet de facilitation dû à la connaissance de la première syllabe peut provenir d’un effet d’anticipation de l’articulation (à la limite, le sujet pourrait commencer à prononcer le début du mot cible dès qu’il voit l’indice de forme; cette critique s’applique également à l’étude de Meyer (1990)). Par contre, l’effet de facilitation dû à la connaissance préalable de la seconde syllabe ne peut s’expliquer ainsi. Ce résultat suggère que connaître à l’avance la dernière syllabe d’un mot aide la phase de préparation de la prononciation de celui-ci.

Il faut néanmoins considérer une hypothèse alternative : il se pourrait que les associations de type “ 1 ” et “ 2 ” soient plus simples à mémoriser que les associations de type “ C ”, et que les sujets n’aient utilisé l’indice de forme que dans les blocs de type “ 1 ” et “ 2 ” et l’aient simplement ignoré dans les blocs de type “ C ”. Dans cette interprétation, la facilitation ne proviendrait pas de l’étape de production du mot, mais d’une étape de traitement située en amont, lors de la sélection du mot associé à la figure présenté. Pour tester cette interprétation, nous avons réalisé une seconde expérience. Elle était identique à la précédente, si ce n’est que les participants ne devaient plus dénommer le mot associé à la figure, mais plutôt le comparer à un autre mot affiché sur l’écran. Plus précisément, les essais débutaient de manière identique mais quand la couleur remplissait la figure, un mot était affiché en dessous de celle-ci, qui avait une probabilité de 50% d’être le mot associé à la figure colorée. La tâche du participant était d’appuyer aussi vite que possible sur une touche du clavier pour signifier si le mot affiché était celui correspondant à la figure. Son temps de décision était mesuré par l’ordinateur.

Cette seconde expérience partage avec la précédente l’étape de sélection du mot associé à la figure présentée sur l’écran. Par contre, il n’y a pas d’étape de préparation de la prononciation du mot. Si l’effet de facilitation observé dans la première expérience est

l’ordre de passation des blocs était contrebalancé entre les sujets à l’intérieur de chaque groupe.



du à l'étape de sélection, on devrait observé le même effet dans la seconde expérience. En fait, les résultats, présentés sur la figure 3, montrent clairement que la sélection du mot n'est pas facilité dans les associations de type "1" ou "2" par rapport aux associations de type "C" (statistiquement, les différences entre les temps de décision ne sont pas significativement différentes de 0).

L'expérience que nous avons réalisée montre que la préparation de la prononciation d'un mot est autant facilité par la connaissance préalable de la dernière syllabe que de la première syllabe d'un mot. Ce résultat milite en faveur de l'hypothèse que les phonèmes d'un mots sont insérés en parallèle, plutôt que séquentiellement, dans le tampon articulatoire. Pourtant, cette conclusion s'oppose à celle de Meyer (1990). Il se peut que les deux tâches expérimentaux impliquent des niveaux de traitement différents : par exemple, notre paradigme aurait pu faciliter l'étape de remplissage du tampon, et celui utilisé par Antje Meyer, la conversion du contenu de ce tampon en articulation. De futurs études devront examiner les facteurs qui distinguent les deux études, parmi lesquels, la langue des participants (hollandais pour Meyer, espagnols pour nous).

Bibliographie

- Costa, A., Sebastian, N., & Pallier, C. (soumis). The time-course of segment-to-frame association in phonological encoding. .
- Coulmas, F. (1989). *The writing systems of the world*. Cambridge MA: Blackwell.
- Dehaene, S. (Ed.). (1997). *Le cerveau en action: imagerie cérébrale fonctionnelle en psychologie cognitive*. Paris: Presses Universitaires de France.
- Démonet, J.-F. (1997). Le traitement du langage. In S. Dehaene (Ed.), *Le cerveau en action : imagerie cérébrale fonctionnelle en psychologie cognitive* (pp. 159-184): Presses Universitaires de France.
- Fromkin, V. (1973). *Speech errors as linguistic evidence*. The Hague: Mouton.
- Levelt, W. J. M. (1989). *Speaking: from intention to articulation*. Cambridge MA: MIT Press.
- Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language*, 29, 524-545.

- 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.
- Moore, B. J. C. (1997). *An Introduction to the Psychology of Hearing*. (4th ed.). London: Academic Press.