

# La syllabe comme unité de perception de la parole : un état de la question

*Alain Content et Uli H. Frauenfelder*

Laboratoire de Psychologie Expérimentale, Université libre de Bruxelles, Ave F.D. Roosevelt 50, B-1050 Bruxelles.  
Email : [alain.content@ulb.ac.be](mailto:alain.content@ulb.ac.be)

Laboratoire de Psycholinguistique Expérimentale, Université de Genève, 40 Bd du Pont d'Arve, CH-1211 Genève 4.  
Email : [ulrich.frauenfelder@pse.unige.ch](mailto:ulrich.frauenfelder@pse.unige.ch)

## ABSTRACT

One highly influential finding that suggests that syllabic units are instrumental in speech perception is the crossover interaction between target type and word type observed in the sequence detection task. In this paper we review our recent studies with French speakers using the same task. Overall the findings fail to replicate the "syllable effect" and indicate that the observed effects are primarily due to the time course of the arrival of phonetic information in the carrier stimuli. These data argue against an early syllabic classification mechanism in speech perception, but other results that we have obtained suggest an important role of syllable structure and more specifically onsets in speech segmentation.

## 1. INTRODUCTION

Une des questions qui traversent toute la psycholinguistique moderne porte sur la nature des unités qui sont à la base du traitement perceptif et linguistique, que ce soit pour la modalité orale ou écrite, les mots ou les phrases, et que ce soit pour la perception ou pour la production. La notion d'unité de traitement provient des modèles de traitement de l'information des années 60, dans lesquels les mécanismes mentaux étaient conceptualisés comme des séquences de transformations de représentations mentales de plus en plus abstraites par rapport aux propriétés physiques de la stimulation sensorielle, de plus en plus durables, et de plus en plus élaborées en termes de leur valeur informationnelle.

Pour la modalité auditive par exemple, les théories dominantes supposaient le stockage immédiat de l'information sensorielle dans une mémoire analogique conçue comme un enregistrement à durée de vie brève (de l'ordre de la seconde). Cette trace sensorielle brute était censée servir de support aux opérations d'analyse par lesquelles l'appareil perceptif identifie des unités discrètes élémentaires, pour accéder ensuite au lexique mental, identifier les mots successifs, retrouver les propriétés morphologiques, syntaxiques et sémantiques, et enfin, construire une interprétation du message.

À la question générale des unités de la perception se superposent d'autres préoccupations liées spécifiquement aux caractéristiques de la parole et des langues naturelles. Ces caractéristiques bien connues des phonéticiens sont celles qui fondent le domaine du décodage acoustico-phonétique. La nature continue, et éminemment variable du signal de parole, l'absence d'indices physiques clairs permettant une segmentation du flux en unités phonétiques, la difficulté rencontrée par les chercheurs pour isoler des propriétés acoustiques ou articulatoires

stables et invariantes permettant une catégorisation non-ambiguë, ont conduit certains psycholinguistes à remettre en cause l'analyse linguistique communément admise, et à envisager d'autres solutions que l'extraction d'une séquence d'éléments discrets correspondant en taille aux segments phonétiques et phonémiques.

L'une des solutions envisagées est la syllabe, et différentes observations empiriques, vers le début des années 70, appuyaient cette hypothèse. Par exemple, le fait que la compréhension d'un extrait présenté en alternance entre l'oreille gauche et l'oreille droite soit particulièrement dégradée lorsque l'alternance dépassait 3 à 4 cps, soit la durée approximative de la syllabe, a été attribué à la perturbation de l'extraction d'unités syllabiques [Hug64, mais voir Sam91]. De même, des expériences de masquage auditif suggéraient que la durée de la trace mnésique initiale correspondait approximativement à celle de la syllabe [Mas72].

Une démonstration plus influente provient d'études basées sur la mesure des temps de détection d'éléments dans des séquences auditives. Savin et Bever [Sav70] ont comparé les temps de détection (mesurés à partir du début de la syllabe critique) de syllabes et de phonèmes dans des suites de syllabes sans signification. Par exemple, les participants entendaient une suite comme « flauk graiv zeem barg drooze » et devaient réagir à la cible /b/ ou « barg ». L'observation était que les auditeurs repéraient plus rapidement la syllabe entière que la consonne initiale de cette même syllabe. Ce résultat surprenant et apparemment paradoxal, puisque la totalité était détectée avant son élément initial, fut initialement interprété comme une démonstration que le format du code perceptif extrait à partir du signal sensoriel devait être de nature syllabique et non-segmental. Les auteurs conclurent que les phonèmes ne constituent pas une unité du code perceptif, mais seraient identifiés tardivement, sur base d'une représentation initiale correspondant à la syllabe.

Les conclusions de Savin et Bever ont cependant été remises en question ultérieurement, notamment du fait de la différence de taille entre les deux types de cibles. Une interprétation alternative [Cut87] est que la spécification d'une cible syllabique fournit une information beaucoup plus détaillée et contextuelle permettant une anticipation plus précise des propriétés phonétiques de la cible. Il se pourrait également que les temps ne reflètent pas tant la nature des processus d'extraction d'information au niveau des phases précoces de traitement perceptif, mais plutôt le temps nécessaire à l'accès de différents types d'informations à des systèmes tardifs de décision [Mor85].

## 2. LA DEMONSTRATION DE L'EFFET SYLLABIQUE

La série d'expériences publiées par Mehler, Dommergues, Frauenfelder et Segui [Meh81] présentait une solution élégante à ce problème. En effet, elle utilisait un principe méthodologique analogue —les auditeurs devaient détecter une cible annoncée au préalable dans une séquence de mots, mais les cibles étaient toutes des syllabes possibles en français. La manipulation expérimentale consistait à varier indépendamment la longueur de la cible (CV ou CVC) et la structure syllabique du porteur, le mot de la séquence qui contenait la cible. L'intérêt de l'étude était dans la comparaison entre les deux cas où la cible correspondait exactement à la syllabe initiale du porteur (par exemple, *ba* dans « *balance* », et *bal* dans « *balcon* »), et ceux où la cible correspondait à une partie de la syllabe initiale (« *ba* » dans « *balcon* ») ou à plus que la syllabe initiale (« *bal* » dans « *balance* »). Les porteurs de structure CV et CVC étaient des mots partageant les trois premiers phonèmes (e.g., *balance-balcon*), de manière que les deux cibles (*ba* et *bal*) puissent être détectées dans les deux types de porteurs.

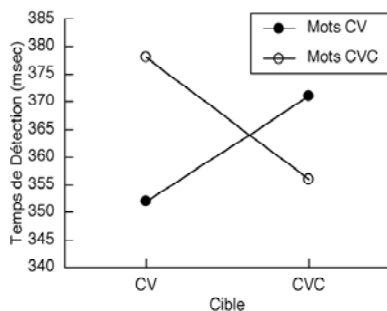


Figure 1. Données de Mehler et al., 1981, Exp. 1

Les résultats, reproduits à la Figure 1 montraient que la détection de cibles CV était plus rapide que celle des cibles CVC dans les porteurs de structure CV. Pris isolément, cette observation est ambiguë, puisqu'elle peut s'expliquer aisément par la différence de longueur des cibles : l'identification des cibles CVC supposerait l'acquisition d'une portion plus longue de signal. Par contre, les résultats obtenus pour les porteurs CVC ne semblaient pas compatibles avec une telle interprétation : la détection des cibles CVC était en effet plus rapide que celle des cibles CV.

Ensemble, les deux observations conduisent à une généralisation simple : les temps de détection des cibles seraient déterminés par la correspondance entre celles-ci et la structure syllabique des porteurs. C'est cette interprétation qui appuie l'idée d'un mécanisme d'identification basé sur l'extraction d'un code syllabique : la détection serait facilitée chaque fois que la cible correspond à une unité de la représentation perceptive immédiate. Par contre, des opérations complémentaires d'analyse et de synthèse seraient nécessaires pour détecter les cibles correspondant à une partie de syllabe ou à la syllabe augmentée de l'attaque suivante. L'étude apparaît particulièrement convaincante par l'utilisation d'un plan expérimental qui, contrairement aux études antérieures utilise les *mêmes* cibles et les *mêmes* porteurs dans les conditions syllabique et non-syllabique.

Dans une étude ultérieure, Cutler et al. [Cut83, Cut86] ont comparé les résultats de l'expérience initiale à ceux d'anglophones testés avec le même matériel expérimental (français), et à des locuteurs français et anglais qui recevaient des stimuli analogues en anglais. Seuls les francophones montraient l'effet syllabique. Il semblait donc que le phénomène ne puisse se réduire à des différences acoustico-phonétiques entre porteurs CV et CVC. De plus, le contraste entre locuteurs francophones et anglophones amenait une réponse nouvelle à la question des unités, qui mettait en relief les particularités des langues et le rôle de l'apprentissage dans la construction mentale de systèmes de représentations phonologiques adaptés. Ainsi, par exemple, la syllabe constituerait une unité de codage importante dans les langues romanes, souvent caractérisées comme des langues à rythme syllabique et à structure syllabique simple. Elle ne jouerait pas un rôle aussi important dans les langues germaniques.

Malheureusement, les études qui ont poursuivi cette perspective, [Kol98, Cut01 pour des revues] n'ont pas confirmé l'hypothèse. L'interaction porteur x cible caractéristique n'apparaît pas systématiquement dans les langues romanes, et inversement, elle est parfois obtenue dans les langues germaniques. De plus, certaines limitations méthodologiques de l'étude initiale nous ont conduits à ré-examiner le phénomène en français.

## 3. NOS RECHERCHES

La première série d'expériences que nous avons réalisée [Con01a] visait à examiner les patrons de temps de détection pour des pseudomots, selon une manipulation comparable à l'étude d'origine. Mais alors que celle-ci utilisait seulement quelques paires de mots, comportant toutes la voyelle /a/ suivie d'une consonne l ou r, nous avons voulu faire une analyse plus systématique, en variant la nature de la voyelle (a i u y) et la catégorie de la consonne pivotale (liquides, *tu.li-toul.mé*; fricatives, *ta.fi-taf.no*; plosives, *vi.por-vip.tère*). Pour chaque type de consonne, huit paires de pseudomots bisyllabiques ont été construites, dont deux avec chacune des quatre voyelles. Les deux pseudomots d'une paire partageaient la séquence CVC initiale et variaient par leur structure syllabique (CV.CV, *tu.li* vs CVC.CV, *tul.me*). Deux expériences ont été menées avec le même matériel et les mêmes enregistrements. Les participants (environ 60 dans chaque expérience) étaient des étudiants de premier cycle à l'Université de Genève. Dans la première, tous les stimuli étaient mélangés, et nous avons examiné l'effet de la présence "d'essais-pièges" dans lesquels le stimulus commence par les mêmes phonèmes que la cible, mais comporte une différence ensuite (par exemple, pour la cible "par", *paltoque*). On sait en effet que la présence de tels pièges ralentit la détection [Nor88, Mil80], et certains travaux suggèrent que l'effet syllabique apparaît plus nettement lorsque la détection est ralentie [Dup90]. Dans la seconde expérience, les essais étaient regroupés soit selon la nature de la voyelle (quatre blocs successifs, un pour chaque voyelle), soit selon la nature de la classe de la consonne pivotale.

Les résultats n'ont pas reproduit l'effet syllabique. Dans la première expérience (Figure 2), toutes les conditions montraient un effet clair de la longueur de la cible (temps de détection plus longs pour les cibles CVC), ainsi qu'un avantage pour les porteurs CVC, surtout pour les cibles CVC. Dans la seconde expérience (Figure 3), les mêmes tendances étaient présentes. En outre, une interaction porteur x cible était observée dans la plupart des cas, mais elle ne correspondait généralement pas au patron croisé caractéristique de l'effet syllabique dans l'étude d'origine. Dans nos données, l'interaction s'expliquait par des temps de détection plus courts pour les porteurs CVC par rapport aux porteurs CV, et ce, particulièrement, pour la détection des cibles CVC (points de droite sur chaque graphique). Par contre, l'avantage des porteurs CV pour la détection de CV (portion de gauche des graphiques) n'apparaissait pas systématiquement. Il était néanmoins observé, mais uniquement dans des conditions très circonscrites, à savoir, pour les consonnes liquides, et dans les conditions de présentation homogène de l'expérience 2.

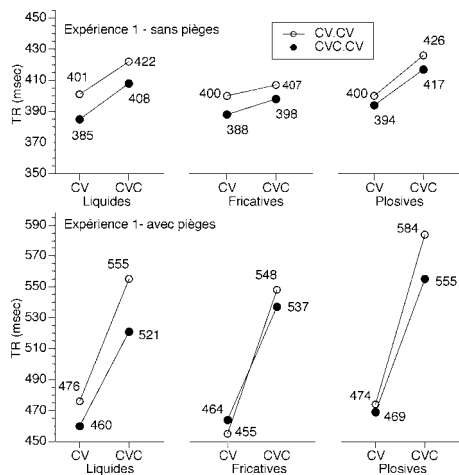


Figure 2. Résultats de Content et al., 2001, Exp.1

L'effet syllabique ne semble donc pas apparaître de manière systématique, tout au moins dans le cas des pseudomots. Le fait que la configuration caractéristique ne soit obtenue que dans le cas des consonnes liquides est intéressant, puisqu'il s'agit des stimuli les plus proches de ceux de l'étude initiale. On peut donc penser que l'interaction croisée observée reflète la manière dont les locuteurs tirent parti d'informations phonétiques fines présentes dans le signal pour anticiper la nature de la cible. Le fait que l'effet ne se généralise pas à d'autres types de consonnes indique qu'il s'agit de propriétés des consonnes liquides plutôt que d'un effet général de la structure syllabique.

Au-delà de cette conclusion négative —la non-réplication de la configuration d'effets de l'étude initiale, il était intéressant d'analyser les données afin de déterminer si certains aspects des performances observées appuyaient néanmoins l'hypothèse d'un codage syllabique. De ce point de vue, les résultats semblent ambigus en première analyse. En effet, l'avantage des porteurs CVC par rapport aux porteurs CV, pour la détection de CVC, pourrait être indicatif d'un codage syllabique. Néanmoins, cet aspect pourrait également s'expliquer par les caractéristiques temporelles des stimuli. Selon le principe d'isochronie

syllabique, on pourrait en effet s'attendre à ce que les syllabes CV (i.e., *ba* dans *balance*) et CVC (*bal* dans *balcon*) aient des durées comparables. De ce fait, la consonne pivotale apparaîtrait plus précocément dans les syllabes CVC. Cette différence dans le déroulement temporel de l'information pourrait évidemment rendre compte de l'avantage observé pour les porteurs CVC, sans qu'il soit nécessaire de supposer un mécanisme mental de codage syllabique.

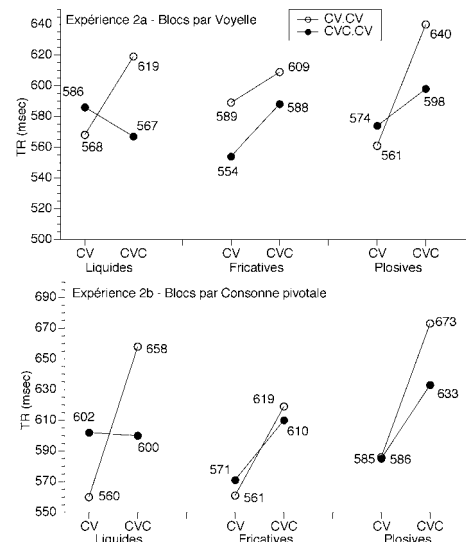


Figure 3. Résultats de Content et al. (2001, Exp. 2).

Des mesures phonétiques et des estimations perceptives ont effectivement confirmé l'existence de cette tendance. Les estimations perceptives étaient basées sur une technique de dévoilement progressif (tâche de "gating"), grâce à laquelle nous avons estimé la durée minimale de signal nécessaire pour que des locuteurs perçoivent correctement la consonne pivotale. Du point de vue de l'analyse phonétique du signal, l'occurrence de la consonne pivotale était en moyenne de 27 ms plus précoce dans les stimuli de structure CVC.CV (231 ms) que dans les CV.CV (258 ms). Du point de vue perceptif, une différence plus marquée (54 ms) était observée.

De plus, des analyses de régression multiple sur les temps moyens de détection des cibles CVC par items ont démontré que 70 à 85% de la variance était expliquée par la position temporelle de la première voyelle et de la consonne pivotale. Aucun effet *direct* de la structure syllabique n'était obtenu. En conséquence, l'ensemble de nos résultats pour la détection des cibles CVC indique que les temps de réponse des locuteurs s'expliquent de façon très directe et très précise sur la base des variations dans le déroulement temporel de l'information phonétique entre les stimuli. Il semble donc que l'avantage des porteurs CVC vis-à-vis des porteurs CV pour la détection de cibles CVC reflète les caractéristiques temporelles des stimuli et ne soit donc pas indicatif par rapport à la question du codage mental syllabique.

Reste évidemment à établir si les résultats obtenus sont propres aux pseudomots. D'une part, on pourrait envisager que les effets syllabiques résultent de phases tardives du traitement, postérieures à l'identification des mots. D'autre part, il est également concevable que des

propriétés articulatoires des pseudomots empêchent l'effet de se manifester. Plus récemment, avec Grégory Leclercq, nous avons réalisé une nouvelle série d'expériences, destinées à comparer directement la détection de cibles pour des mots et des pseudomots appariés. Pour reproduire au plus près les conditions dans lesquelles nous avons précédemment observé l'effet syllabique, tous les stimuli utilisaient des voyelles pivotales liquides. Les détails de l'étude sont présentés dans une autre contribution [Lec02], mais la conclusion principale est que ni pour les mots, ni pour les pseudomots, nous n'avons reproduit l'effet syllabique. Les résultats obtenus dans cette nouvelle étude sont parfaitement compatibles avec notre interprétation : l'avantage des porteurs CV pour la détection de CV est reproduit avec un nouveau matériel, tant pour les mots que pour les pseudomots. Par contre, l'avantage des porteurs CVC pour la détection de CVC n'apparaît pas, mais l'analyse phonétique indique que les différences de durées de la séquence CV initiale entre porteurs CV et CVC est beaucoup plus réduite que dans nos études antérieures.

#### 4. CONCLUSION

Malgré l'influence profonde que l'étude de Mehler et collaborateurs a exercée sur les travaux psycholinguistiques des 20 dernières années, aucune reproduction directe de l'effet n'a été publiée. Nos propres tentatives de reproduire le phénomène avec des pseudomots ou des mots se sont soldées par des échecs.

Les résultats que nous avons obtenus montrent qu'une partie de l'effet s'explique par les propriétés temporelles des syllabes en production sans requérir la notion de codage mental syllabique. Par ailleurs, ils suggèrent que l'effet syllabique ne serait apparent que pour des consonnes pivotales liquides, ce qui nous a amenés à penser qu'il reflète la présence de dissemblances fines entre les cibles et leurs réalisations phonétiques dans les porteurs qui rendraient la détection plus ardue. La nature de ces indices reste toutefois une question ouverte.

Enfin, la ré-interprétation de l'effet syllabique que nous avons proposée ne signifie pas que la structure syllabique n'ait aucun rôle à jouer dans la perception de la parole. Au contraire, nous avons avancé dans d'autres publications [Con01b, Dum02] l'hypothèse que la perception des attaques de syllabes pourrait servir d'ancrage aux processus d'accès lexical dans la parole continue.

#### BIBLIOGRAPHIE

- Con01a Content, A., Meunier, C., Kearns, R. K., & Frauenfelder, U. H. (2001). Sequence detection in pseudowords in French: Where is the syllable effect? *Lang. Cogn. Proc.*, 16, 609-636.
- Con01b Content, A., Kearns, R., & Frauenfelder, U. (2001). Boundaries versus onsets in syllabic segmentation. *J. Mem. Lang.*, 45, 177-199.
- Cut83 Cutler, A., Mehler, J., Norris, D., & Segui, J. (1983). A language-specific comprehension strategy. *Nature*, 304, 159-160.
- Cut86 Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *J. Mem. Lang.*, 25, 385-400.
- Cut87 Cutler, A., Norris, D., & Williams, J. N. (1987). A note on the role of phonological expectations in speech segmentation. *J. Mem. Lang.*, 26, 480-487.
- Dum01 Dumay, N., Frauenfelder, U., & Content, A. (2001). The role of the syllable in lexical segmentation in French: Word-spotting data. *Brain and Language*, in press.
- Dup89 Dupoux, E. *Identification des mots parlés*. Thèse de doctorat non publiée, E.H.E.S.S., Paris.
- Hug64 Huggins, A.W.F. (1964). Distortion of the temporal pattern of speech: interruption and alternation. *J. Acoust. Soc. Amer.*, 36, 98-104.
- Kol98 Kolinsky, R. (1998). Spoken word recognition: A stage-processing approach to language differences. *Eur. J. Cognit. Psych.*, 10, 1-40.
- Lec02 Leclercq, G., Content, A. & Frauenfelder, U. (2002). L'effet syllabique dans les mots et les pseudomots en français. *Actes des XXIVèmes Journées d'Etude de la Parole*, Nancy.
- Mas72 Massaro, D. W. (1972). Preperceptual images, processing time, and perceptual units in auditory perception. *Psych. Rev.*, 79, 124-145.
- Meh81 Mehler, J., Dommergues, J. Y., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *J. Verb. Learn. Verb. Behav.*, 20, 298-305.
- Mil80 Mills, C. B. (1980). Effects of the match between listener expectancies and coarticulatory cues on the perception of speech. *J. Exp. Psych. Hum. Perc. Perf.*, 3, 528-535.
- Mor85 Morais, J. (1985). Literacy and awareness of the units of speech: Implications for research on the units of perception. *Linguistics*, 23, 707-721.
- Nor88 Norris, D., & Cutler, A. (1988). The relative accessibility of phonemes and syllables. *Percept. Psychophys.*, 43, 541-550.
- Sam91 Samuel, A. G. (1991). Perceptual degradation due to signal alternation: implications for auditory pattern processing. *J. Exp. Psych. Hum. Perc. Perf.*, 17, 392-403.
- Sav70 Savin, H. B., & Bever, T. G. (1970). The nonperceptual reality of the phoneme. *J. Verb. Learn. Verb. Behav.*, 9, 295-302.