

Etude préliminaire de la perception précoce des voyelles labialisées par des auditeurs déficients visuels

Fabrice Hirsch¹, Henri Dreyfus², Rudolph Sock¹, Béatrice Vaxelaire¹,
Camille Fauth¹, Fayssal Bouarourou¹, Marion Béchet¹

¹ Institut de Phonétique de Strasbourg & E.A. 1339 - LiLpa, Composante Parole et Cognition
Université de Strasbourg, 22, rue René Descartes, 67084 Strasbourg

² Unité de Recherche INSERM U.44, 12, rue de Copenhague, 67000 Strasbourg
f.hirsch@unistra.fr

ABSTRACT

The present research deals with auditory effects of anticipatory lip movements in [V¹CV²] sequences, where V¹ is vowel [i], C is the fricative [s] and V² a rounded vowel [V^{lab}]. The aim of the investigation is triple: first, we wanted to observe for potential anticipatory acoustic effects of vowel rounding, when a speaker pronounces sequences which include vowels [y], [ɛ], [u] or [o] in the V² position. Secondly, attention is paid to the moment when sighted listeners may start to perceive probable anticipatory effects of the rounded vowel. Thirdly, the core purpose of this work was to analyse the behaviour of blind subjects carrying out the same experimental tasks. Results show that vowels [y] and [u] are perceived earlier than [ɛ] and [o] by both control and blind subjects. When data for the control group and the group of blind subjects are compared, it is noticed that blind subjects perceive rounded vowels earlier than control sighted subjects.

Keywords: auditory perception, blindness, anticipation, rounded vowels

1. INTRODUCTION

Un grand nombre d'études ([1] par exemple) a montré que les mouvements propres à chaque son d'une séquence [CV] ne sont pas produits de manière successive. Au contraire, les études articulatoires menées sur ce sujet ont observé un phénomène de coarticulation, et plus précisément celui de la coarticulation anticipatoire. Ce phénomène de coarticulation anticipatoire que nous venons d'évoquer a des répercussions sur la perception de la parole. En effet, l'anticipation des gestes de la parole peut être perçue auditivement ([2] ; [3]) et visuellement ([4]), notamment pour des syllabes [CV], où la voyelle est labialisée. En d'autres termes, dans une séquence [su] par exemple, la voyelle [u] peut être perçue, par des auditeurs n'ayant aucun trouble auditif ou visuel, dans le [s], alors même que cette voyelle n'est pas encore rendue acoustiquement.

Si des auditeurs bien-voyants ont la capacité de percevoir précocement un son qui est sur le point d'être produit grâce à certains indices articulatoires et acoustiques, qu'en est-il d'auditeurs mal-voyants ou non-voyants ? La

question mérite d'être posée, étant donné que certaines études ont montré des particularités au niveau de la perception des sons et du message linguistique chez les personnes atteintes de cécité. Moos *et al.* [5] ont, par exemple, observé que les auditeurs non-voyants sont capables de percevoir le sens d'une parole ultra-rapide à une vitesse d'élocution incompréhensible pour des auditeurs sans trouble visuel. Il en est de même pour Menard *et al.* [6] qui ont notamment mis en avant le fait que les personnes souffrant de cécité présentaient des scores de discrimination plus élevés lorsqu'il s'agissait de catégoriser des voyelles par rapport à des auditeurs de contrôle. Ainsi, il semblerait que les auditeurs non-voyants ou malvoyants développeraient des facultés compensatrices en matière d'audition.

L'*objectif* de notre étude sera donc de vérifier deux *hypothèses* : premièrement, nous pensons que les auditeurs non-voyants percevraient la voyelle protruse avant les auditeurs sans trouble visuel, et de manière plus certaine (selon un seuil de confiance), dans des séquences [isV^{lab}], cela grâce aux réajustements auditifs compensatoires liés à la perturbation du canal visuel ; Secondement, nous posons le fait que la perception de la voyelle serait moins précoce à mesure que cette dernière soit moins labialisée, et cela aussi bien pour les auditeurs non-voyants que voyants. Ainsi, l'élément vocalique [y], par exemple, serait discerné avant le [ɛ].

2. MÉTHODE

2.1. Corpus, enregistrement et mesures acoustiques

Afin de vérifier nos hypothèses, des séquences [isV^{lab}] insérées dans une phrase porteuse du type « C'est [isV^{lab}] ça » ont été enregistrées dans la chambre insonorisée de l'Institut de Phonétique de Strasbourg. Il est à noter que la voyelle [V^{lab}] était soit [y], [ɛ], [u] ou [o]. Chaque phrase a été prononcée à trois reprises par un locuteur. Le choix de ce corpus permettait de tester l'effet de l'aperture et de la labialité sur la perception anticipée de la voyelle, et de vérifier si les résultats observés sont les mêmes selon que les éléments vocaliques soient antérieurs ou postérieurs.

Une fois les données acquises, des mesures qualitatives et quantitatives ont été effectuées. La structure formantique

(F1, F2 et F3) a ainsi été mesurée au milieu du [i] et au milieu de la voyelle labialisée. Parallèlement à cela, la fréquence de la limite inférieure du bruit de friction du [s] a été relevée toutes les 10 ms à partir du début de la structure formantique stable de la voyelle protruse. Notons que la fréquence du bruit de friction a tendance à diminuer à mesure que l'on se rapproche d'un élément labialisé (Hirsch, [3] ; Calliope, [7]).

En outre, la durée du [i], de [V^{lab}] et du [s] a été quantifiée. De même, nous indiquerons la durée de l'intervalle compris entre le début de l'inflexion du bruit de friction (si elle a lieu) et le début de la structure formantique stable de la voyelle labialisée en termes de valeurs absolues et relatives, cela afin d'évaluer l'effet de l'élasticité temporelle du signal acoustique.

2.2. Tests de perception et auditeurs

Afin d'étudier le moment où les voyelles labialisées commencent à être perceptibles, quatre phrases ont été sélectionnées parmi le corpus présenté, comportant chacune une voyelle labialisée différente. Le paradigme du *gating* ou du dévoilement progressif du signal, a été utilisé puisque les séquences étudiées ont été tronquées toutes les 10 ms en partant de la structure formantique stable de la voyelle labialisée, et en allant vers le [i]. Un montage a été réalisé pour chacune des phrases proposées avec, entre chaque stimulus, un bip prévenant de l'imminence d'une séquence placé 1,5 sec. avant toute séquence. Les séquences tronquées étaient présentées en ordre aléatoire.

20 auditeurs ont été recrutés pour écouter les stimuli : 10 personnes sans trouble visuel ni auditif, et 10 personnes déficientes visuelles (8 non-voyants et 2 mal-voyants). Etant donné que certains sujets non-voyants n'étaient pas en mesure de cocher une réponse ou de l'écrire, les auditeurs avaient tous pour consigne d'écouter, dans un premier temps la séquence tronquée, puis de donner oralement la voyelle suivant le [s] à l'expérimentateur qui prenait les notes. Les auditeurs avaient le choix entre trois possibilités : [e], [a] (retenues comme « distracteurs ») ou [V^{lab}]. En outre, il était demandé d'allouer une note de confiance (nc) à la réponse donnée, note qui pouvait aller, dans une échelle subjective, de 0 à 5, selon que la voyelle ait été donnée au hasard (0) ou avec une assurance totale (5).

3. RÉSULTATS

3.1. Observations acoustiques

Deux tendances se dégagent lorsque l'on observe la limite inférieure du bruit de friction : pour [y] et [u], la fréquence de la limite inférieure du bruit de friction du [s] a tendance à diminuer très légèrement dès la fin du [i], avant qu'une inflexion plus prononcée ait lieu jusqu'à l'arrivée de la structure formantique de la voyelle labialisée. Dans le cas de la séquence [isy], la limite inférieure du bruit de friction s'élève à 3864 Hz au premier relevé situé à proximité du [i] avant d'être évalué

à 3473 Hz, 90 ms avant le début acoustique du [y] ; à cette date, une forte inflexion est observée, puisque la limite inférieure du bruit de friction diminue fortement pour atteindre 1337 Hz, à 10 ms de l'émergence de la structure formantique stable du [y]. Notons que la date de l'inflexion correspond à 51 % de l'intervalle obstruent [s].

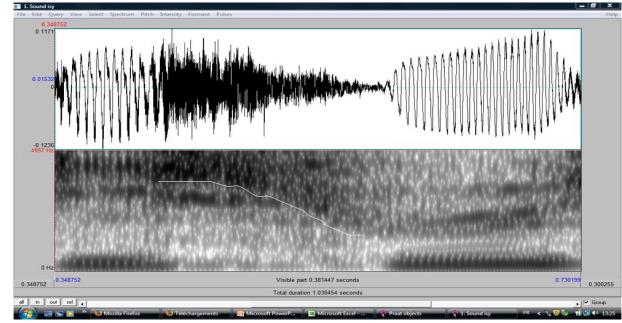


Figure 1 : Signal acoustique et spectrogramme d'une séquence [isy]. La fréquence de la limite inférieure du bruit de friction diminue en deux temps : une légère baisse d'abord avant qu'une inflexion brutale ne soit observée.

Notons que nous avons observé le même phénomène pour [isu]. En ce qui concerne les séquences comportant les voyelles [ʌ] et [o], la tendance est différente dans la mesure où aucune inflexion forte n'a pu être constatée. En effet, seule une légère baisse de la limite inférieure du bruit de friction est observée, celle-ci passant de 4027 Hz (à la date la plus proche du [i]) à 3277 Hz.

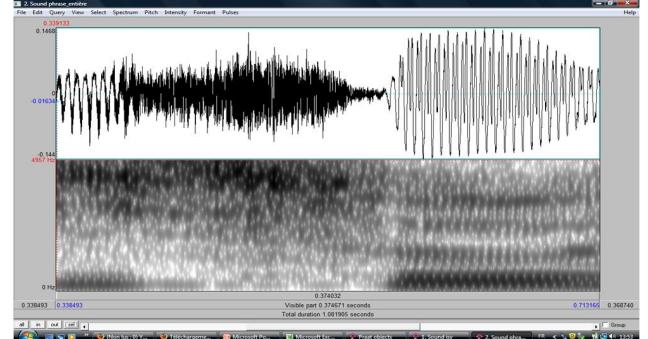


Figure 2 : Signal acoustique et spectrogramme d'une séquence [isʌ]. La fréquence de la limite inférieure du bruit de friction diminue légèrement mais aucune inflexion n'est visible.

3.2. Tests de perception : les résultats

Résultats pour les auditeurs de contrôle

Les résultats (Figure 3) révèlent que la voyelle [y] commence à être perçue 90 ms avant son début acoustique, et cela par 70% des sujets témoins (nc = 2). Il est intéressant de rappeler ici que c'est à cette même date que le début de l'inflexion du bruit de friction du [s] a lieu. Notons aussi qu'à la date suivante, c'est-à-dire à 100 ms de la structure formantique stable de la voyelle arrondie, le pourcentage d'auditeurs qui ont donné une réponse correcte n'est plus que de 50% (nc = 2), ce qui signifie que le seuil du hasard a été franchi.

En outre, il a également été possible d'observer que le seuil de confiance a tendance à diminuer à mesure que l'on s'éloigne de la voyelle labialisée.

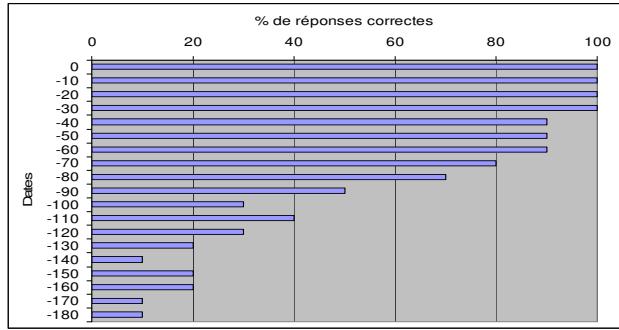


Figure 3 : Pourcentage de réponses correctes pour les auditeurs de contrôle (abscisse) en fonction des points de troncation (ordonnée). La voyelle [y] est perçue 90 ms avant son début acoustique.

Pour ce qui concerne les résultats obtenus pour la voyelle [ʌ] (Figure 4), ils ne permettent pas de parler de perception précoce, étant donné qu'elle n'a été identifiée, par 90 % des auditeurs, qu'à l'apparition de sa structure formantique clairement définie (nc = 3). 10 ms plus tard, le pourcentage de réponses correctes diminue déjà pour atteindre 50% (nc = 1).

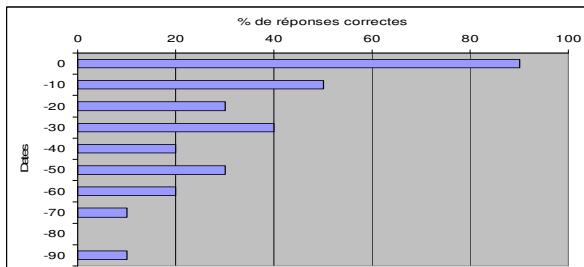


Figure 4 : Pourcentage de réponses correctes pour les auditeurs de contrôle (abscisse) en fonction des points de troncation (ordonnée). La voyelle [ʌ] n'est pas perçue avant son début acoustique.

Pour ce qui est de l'autre couple étudié, les séquences [isu] et [iso], le schéma reste le même que pour les voyelles antérieures, à savoir que la voyelle de petite aperture et très labialisée [u] est perçue de manière anticipée 80 ms avant son début acoustique (nc = 2), alors que le [o] est difficilement perceptible avant le début de sa structure formantique stable.

Comparaison des résultats pour les auditeurs de contrôle vs. les auditeurs déficients visuels

La comparaison des résultats des auditeurs de contrôle avec ceux des auditeurs ayant une déficience visuelle montre que les non-voyants et les mal-voyants tendent à percevoir la voyelle labialisée avant le groupe de contrôle. C'est le cas pour le [y] qui commence à être reconnu par 70 % des auditeurs atteints de déficience visuelle, à 110 ms de son début acoustique (nc = 3). 10 ms plus tard, la voyelle est encore perçue par 60% des auditeurs déficients visuels, score que nous estimons trop bas pour être

robuste. Quant aux sujets de contrôle, rappelons qu'ils ne percevaient cette même voyelle que 90 ms avant son début acoustique.

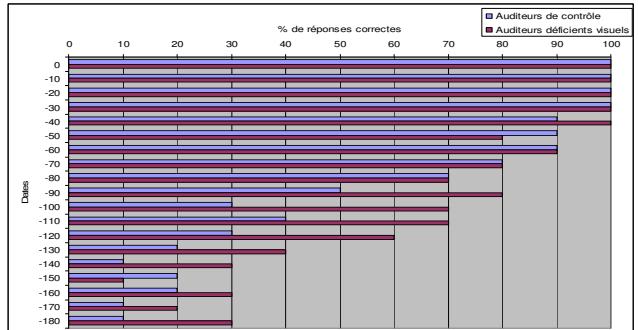


Figure 5 : Comparaison du pourcentage de réponses correctes (abscisse) chez les auditeurs de contrôle et les sujets déficients visuels en fonction des points de troncation (ordonnée) pour la voyelle [y]. Le groupe d'auditeurs non-voyants et mal-voyants perçoit la voyelle arrondie avant les auditeurs de contrôle.

Quant à la voyelle [ʌ], elle est identifiée par 70% des auditeurs 10 ms avant son début acoustique (nc = 3). Cela dit, nous notons une légère baisse du pourcentage de réponses correctes à 20 ms du [ʌ], étant donné que seuls 60% des sujets présentant un trouble visuel sévère ont identifié la voyelle correctement (nc = 3). A 30 ms du début de la voyelle labialisée, le nombre d'auditeurs qui ont indiqué une réponse juste remonte légèrement pour atteindre à nouveau 70% (nc = 2). Par la suite, le pourcentage de réponses correctes diminue pour atteindre des scores non-significatifs (50% des auditeurs déficients visuels ont identifié la voyelle à 40 ms du [ʌ], puis 30 % à la date suivante).

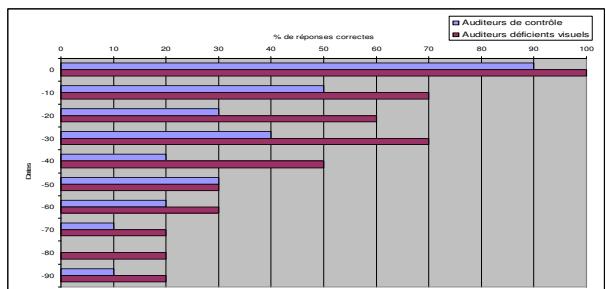


Figure 6 : Comparaison du pourcentage de réponses correctes (abscisse) chez les auditeurs de contrôle et les sujets déficients visuels en fonction des points de troncation (ordonnée) pour la voyelle [ʌ]. Le groupe d'auditeurs non-voyants et malvoyants perçoit la voyelle arrondie avant les auditeurs de contrôle.

Les mêmes tendances ont pu être observées pour les voyelles postérieures du corpus puisque le [u] est identifié plus tôt par les auditeurs non-voyants et malvoyants (à 100 ms avant le début acoustique de la voyelle vs. à 80 ms pour les auditeurs de contrôle), tout comme le [o] qui est reconnu 20 ms (nc = 2) avant l'apparition de la structure formantique, par le groupe présentant des troubles sévères de la vue, alors que les sujets témoins n'avaient pas réussi à identifier cette même voyelle.

Quant aux notes de confiance, elles sont généralement plus élevées chez les auditeurs déficients visuels (2,8 en moyenne) qui semblent, de ce fait, plus sûrs de leur réponse que le groupe de contrôle (2,2 en moyenne). De manière générale, tous les sujets ont été sévères dans l'attribution de leurs notes de confiance, puisque les valeurs moyennes restent relativement bas, même pour des scores d'identification correcte élevés.

4. SYNTHÈSE ET DISCUSSION

Au commencement de ce travail, nous avions formulé deux *hypothèses* : nous pensions tout d'abord que la perception de voyelles labialisées serait moins précoce à mesure que celles-là soient moins protruse. Deuxièmement, nous supposions que des auditeurs non-voyants percevraient la voyelle protruse avant les auditeurs sans trouble visuel.

L'étude acoustique et perceptive des séquences [isV^{lab}], menée auprès des *sujets témoins*, révèle deux phénomènes, l'un concernant les voyelles très labialisées et de petite aperture, l'autre les voyelles moins labialisées de moyenne aperture. Pour les premières citées ([y] et [u]), la fréquence de la limite inférieure du [s] diminue d'abord légèrement, et cela dès la fin de la voyelle [i], avant qu'une inflexion forte ait lieu. La date à laquelle se trouve l'infexion correspond, plus ou moins, au moment où les auditeurs de contrôle commencent à percevoir la voyelle labialisée. Notons que ce résultat peut être mis en parallèle avec ceux obtenus par Vaxelaire *et al.* [8] qui avaient remarqué que, pour une séquence comportant une fricative suivie d'un [y], le moment où les auditeurs perçoivent précocement la voyelle labialisée correspond, au niveau acoustique, au début de l'infexion du bruit de friction qui coïncide, lui-même, avec le pic d'accélération du mouvement labial au niveau articulatoire. Par conséquent, les sujets témoins sembleraient sensibles à certains événements cinématiques qui leur permettraient d'identifier auditivement une voyelle avant son émergence acoustique. Pour le second groupe de voyelles ([ɛ] et [o]), la limite inférieure du bruit de friction diminue, de manière plus ou moins régulière, de la fin du [i] à la voyelle [V^{lab}], mais aucune inflexion n'est visible. Cela laisse supposer qu'une augmentation de l'aperture provoquerait une diminution de la protrusion, ce qui entraînerait une modification des paramètres cinématiques. En d'autres termes, étant donné que la protrusion est moins prononcée, il est envisageable que le pic d'accélération intervienne plus tardivement, ce qui expliquerait pourquoi il ne provoquerait pas d'infexion au niveau de la limite inférieure du bruit de friction. En ce qui concerne les auditeurs *déficients visuels*, nous avons pu observer les mêmes tendances, les voyelles très labialisées étant perçues bien avant les voyelles labialisées de moyenne aperture. En comparant les résultats obtenus pour ces derniers avec ceux des auditeurs de contrôle, il a été constaté que le groupe atteint d'un trouble sévère de la vue perçoit plus tôt les éléments vocaliques labialisés par rapport aux sujets témoins. Ce résultat pourrait s'expliquer par le fait que les

auditeurs déficients visuels développeraient une sensibilité auditive plus prononcée leur permettant de percevoir des changements plus fins au niveau de la limite inférieure du bruit de friction.

En perspective à cette étude, il semblerait intéressant de continuer cette recherche en recrutant davantage d'auditeurs dans chaque groupe. En outre, nous souhaiterions vérifier si les auditeurs témoins seraient capables de rattraper leur « retard » de perception lorsqu'on leur présente des informations audio-visuelles ou lorsqu'on leur masque la vue. Dans tous les cas, nos données acoustiques et auditives sont à rapprocher des celles connexes, obtenues dans les domaines articulatoire ([3] ; [8]) et cinématique ([4]), cela pour pouvoir déceler les relations sensori-motrices impliquées dans la perception précoce de faits anticipatoires.

Remerciements : Programme MISHA (Maison Interuniversitaire des Sciences de l'Homme d'Alsace) « Perturbations et réajustements : parole normale vs. parole pathologique » 2008-2012 ; ANR-07-CORP-018-01, DOCVACIM, 2007-2011 ; Association des Aveugles d'Alsace et de Lorraine (AAAL) ; Coralie Vincent.

BIBLIOGRAPHIE

- [1] C. Abry and T. Lallouache. Le MEM : un modèle d'anticipation paramétrable par locuteur : données sur l'arrondissement en français. *Bulletin de la communication parlée*, volume 3, pages 85-89, 1995.
- [2] V. Ferbach-Hecker. *La perception auditive de l'anticipation des gestes vocaliques en français*. Thèse de doctorat Nouveau Régime soutenue à l'Université de Strasbourg, 2002.
- [3] F. Hirsch, R. Sock, P.-Y. Connan and G. Brock. Auditory effects of anticipatory rounding in relation with vowel height in French. In *Proceedings of the International Phonetic Sciences*, pages 1445-1448, 2003.
- [4] J.P. Roy. Visual perception of anticipatory rounding gestures in French. In *Proceedings of the 9th European Conference on Speech Communication and Technology*, pages 2949-2952, 2005.
- [5] A. Moos, I. Hertrich, S. Dietrich, J. Trouvain and H. H. Ackermann. Perception of ultra-fast speech by a blind listener - Does he use his visual system?. In *Proceedings of the International Speech Production Seminar, Strasbourg*, pages 297-300, 2008.
- [6] L. Menard, S. Dupont, S.R. Baum and J. Aubin. Production and perception of French vowels, by congenitally blind adults and sighted adults. In *Journal of acoustical society of America*, volume 126 (3), pages 1406-1414, 2009.
- [7] Calliope. *La parole et son traitement automatique*. Editions Masson, 1992.
- [8] B. Vaxelaire, R. Sock, F. Hirsch, V. Ferbach-Hecker, J.P. Roy and F. Bouarourou. *The Anticipatory Perception Based on Events (APE) Hypothesis*. Communication présentée aux Vè Convegno Nazionale AISV, Zurich, 2009.