Perception anticipatoire catégorielle des voyelles arrondies du français

Fabrice Hirsch Rudolph Sock Johanna-Pascale Roy Mélanie Canault

Institut de Phonétique de Strasbourg – E.A. 3403
22, rue Descartes - 67084 Strasbourg
Tél : ++33 (0) 3.88.41.73.64
Mél : fabrice_hirsch@yahoo.fr

ABSTRACT

Most research works that have dealt with anticipatory auditory perception of French rounded vowels, offered a choice between identification of a rounded vowel among several other un-rounded ones. If such an approach indeed allows verifying subjects' sensitivity to vowel labialisation, one cannot be sure that listeners would succeed in precisely categorising the rounded vowel. The aim of this research is to find out if perception of certain acoustic properties of vowel protrusion, contained in a preceding fricative, is accompanied by vowel categorisation, or if auditors simply react to those properties without being able to successfully carry out an identification and discrimination task.

1. INTRODUCTION

La raison d'être principale des gestes anticipatoires semble correspondre à leur efficacité auditive et visuelle. L'apparition de tels gestes correspondrait soit à des stratégies coarticulatoires déployées par les locuteurs, soit à des contraintes biomécaniques inhérentes aux articulateurs eux-mêmes, ainsi qu'à leur coordination. L'interlocuteur aurait appris à exploiter ces caractéristiques motrices anticipatoires dans la communication linguistique, leur perception étant principalement auditive et visuelle.

L'exploitation en production-perception de la parole de tels mécanismes coarticulatoires a été étudiée sur les plans acoustiques-auditifs (cf., par ex., BENGUEREL et ADELMAN [1] ; LUBKER et LINDGREN [4]) et dans le domaine visuel. Cependant, il a été montré (cf., par ex., SCHWARTZ [6]) que, pour comprendre correctement ces phénomènes, il fallait étudier en parallèle les dimensions motrice (articulatoire) et sensorielle (acoustique).

L'étude présentée ici suit cette démarche sensori-motrice et s'inscrit dans le cadre d'un ensemble de travaux portant sur l'efficacité perceptive des gestes anticipatoires en production de la parole, réalisés à l'Institut de Phonétique de Strasbourg (FERBACH-HECKER [2] ; VAXELAIRE *et al*. [7] ; ROY *et al*. [5]) dans le cadre d'un programme de recherche. De manière plus ciblée, il s'agit de la suite donnée à nos précédentes recherches (HIRSCH *et al.* [3]), dans lesquelles nous avons tenté d'interpréter systématiquement nos données acoustiques en termes articulatoires. Ayant tronqué une voyelle labialisée cible du signal acoustique, nous avons pu mettre en avant la relation entre le moment où un effet de la labialisation vocalique était perçu et le moment où la fréquence du bruit de friction d'un [s], dans une séquence [i + s + Vlab], entamait une baisse plus où moins marquée, selon que la voyelle fût très ou peu protruse. Ainsi, nous avons pu observer que la perception auditive de l'anticipation des mouvements des lèvres débutait plus tôt pour la voyelle [y], par rapport aux autres voyelles labialisées. Cette perception anticipatoire de la labialité était en fonction de l'aperture, diminuant à mesure que celle-ci augmentait. Néanmoins, cette étude, comme d'autres portant sur le même sujet, proposait aux auditeurs un test de perception, ceux-ci ayant pour consigne de sélectionner la voyelle labialisée tronquée parmi plusieurs autres voyelles non-labialisées. Si une telle approche permet d'observer la réaction des auditeurs quant à la perception anticipatoire d'une éventuelle composante de la labialité vocalique, elle ne permet pas pour autant de conclure à une identification catégorielle de la voyelle. En réalité, c'est la capacité des auditeurs à détecter les propriétés d'un élément vocalique labialisé, mais tronqué du signal acoustique, qui a été démontrée.

Cette présente étude, portant sur des séquences [is + Vlab], où Vlab peut être un [y], un [u], un [ø] ou un [o], se proposera donc de vérifier si toute labialité perçue sera accompagnée ou non d'une identification catégorielle des voyelles, ou si les auditeurs se limiteraient à ne percevoir que les propriétés acoustiques de la labialité vocalique présentes dans la fricative précédente [s], sans pour autant réussir à affiner leurs décisions en termes de discrimination des voyelles labialisées. Nous tâcherons de rationaliser nos données dans le cadre des résultats obtenus précédemment utilisant ce même paradigme du *gating*.

2. ACQUISITION DES DONNEES

2.1 Locuteurs et corpus

Quatre phrases porteuses, comprenant la séquence [is +Vlab] ont été prononcées par un locuteur, VT et une locutrice, BV en deux vitesses d'élocution, normale et rapide. La variation de la vitesse d'élocution permet d'évaluer toute modification éventuelle de l'organisation temporelle du signal, ainsi que la robustesse des résultats perceptifs en fonction du changement de la condition prosodique.

Le corpus, qui a été enregistré dans la chambre insonorisée de l'Institut de Phonétique de Strasbourg, comprend les quatre phrases suivantes :

1. C'est issu ça.

2. C'est Tissou ça.
3. C'est Tissot ça.
4. C'est Tisseut ça.

La première phrase permet d'étudier la séquence [isy], la deuxième la séquence [isu], la troisième la séquence [iso] et la dernière, la séquence [isø].

Nous avons ensuite enregistré, sur un CD, les quatre phrases tronquées, toutes les 20 ms (dates de troncation), en partant du début de la structure formantiquement stable des voyelles arrondies, jusqu'à ce que nous arrivions à un niveau suffisamment éloigné de cette même voyelle arrondie pour ne plus percevoir auditivement l'effet de la présence d'un élément arrondi subséquent. Les séquences tronquées ont été ensuite disposées dans un ordre aléatoire pour effectuer le test de perception.

2.2 Test de perception

12 auditeurs adultes, hommes et femmes, de langue maternelle française et âgés de 18 à 26 ans se sont portés sujets pour le test. Ils étaient tous naïfs par rapport au but de l'expérience et ne présentaient aucun trouble d'audition ou de production de la parole. Les tests se sont déroulés à l'Institut de Phonétique de Strasbourg. Un magnétophone MARANTZ PM D 222 a été utilisé pour l'écoute des stimuli. La bande sonore a été entendue séparément par les 12 auditeurs, chacun muni d'un casque BEYER DT 770.

Lors du test, les sujets ont été avertis qu'ils allaient entendre l'une des phrases suivantes :

1. C'est issu ça
2. C'est Tissou ça
3. C'est Tissot ça
4. C'est Tisseut ça
5. C'est Tissa ça
6. C'est Tissé ça

Soit six phrases dans lesquelles les séquences [isy], [isu], [iso], [isø] [isa] et [ise] étaient incluses.

En réalité, dans cette expérience précise, les phrases 5 et 6 n'ont jamais été présentées, étant donné qu'elles servaient simplement de distracteurs pour les auditeurs. Signalons toutefois que les résultats obtenus dans des expériences similaires n'ont indiqué aucun changement significatif lorsque les deux phrases tronquées, comportant des voyelles non-arrondies, avaient été réellement livrées aux auditeurs (voir, par ex., FERBACH-HECKER, 2002 [2], pour plus de détails).

Chaque échantillon tronqué consistait en l'amorce [sɛtis...], suivie de la voyelle tronquée, avec de plus en plus d'informations acoustiques/phonétiques contenues dans l'intervalle fricatif de la consonne constrictive précédant la voyelle arrondie. Ainsi, la dernière date correspondait toujours à l'établissement de la structure formantique clairement définie de la voyelle arrondie. Un signal sonore (bip acoustique) précédait de 1,4 seconde chaque échantillon. Ce bip servait à prévenir l'auditeur de l'imminence d'un stimulus. L'intervalle inter-stimuli était de 4 secondes. Les échantillons ont été présentés dans un ordre aléatoire. Une courte pause de 10 secondes a été intercalée entre chaque vitesse d'élocution.

Des feuilles de réponses ont été fournies aux auditeurs. Durant les 4 secondes d'intervalle entre les stimuli, ceux-ci devaient remplir deux tâches pour chaque stimulus : (1) marquer à l'aide d'une croix laquelle des voyelles [e, a] ou Vlab [y, u, o, ø] ils pensaient avoir entendu et (2) attribuer un poids de certitude ou de confiance à leur réponse, un poids qui pouvait varier sur une échelle subjective allant de « 1 » à « 5 », où « 1 » indiquait une faible confiance dans leur choix et « 5 » une certitude absolue.

Lors de cette étude, la date la plus éloignée de la voyelle cible, à laquelle 60% environ des réponses sont correctes, est considérée comme le seuil à partir duquel cette voyelle commence à être perçue.

3. RÉSULTATS ET DISCUSSION

La valeur moyenne du seuil de confiance est relativement élevée (2,31, avec un écart-type de 0,64) pour les 69 stimuli entendus par les 12 auditeurs (soit 828 réponses). Ce résultat indique que les réponses ont été données avec un niveau de confiance satisfaisant dans les deux vitesses d'élocution. De même, le pourcentage de réponses correctes est fortement corrélé avec le niveau de confiance (r=0,97), ce qui signifie que les auditeurs étaient performants et cohérents dans les tâches d'identification.

3.1 Perception anticipatoire des voyelles labialisées

La figure 1 montre que, pour le locuteur VT en vitesse d'élocution normale, le [y] est perçu de manière anticipatoire et catégorielle à partir du point de troncation n°3, situé à 60 ms du début acoustique de la voyelle, le taux de réponses correctes étant de 92% à cette date. La légère baisse du pourcentage de réponses correctes aux dates 5 (83 %) et 6 (75%), soit respectivement à 20 ms du début acoustique de la voyelle et au début celle-ci, laisse penser que les réponses à ces dates sont légèrement perturbées par la présence d'autres voyelles labialisées concurrentes à identifier. Notons cependant qu'il ne s'agit là que d'une tendance. Nous y reviendrons plus loin.

Pour ce qui est de la voyelle [u], on constate sur la figure 1 que le pourcentage de réponses correctes commence à être remarquable à partir de la date 4, située à 40 ms de son début acoustique, puisque l'identification catégorielle se fait alors à 92%. Signalons, encore, la légère baisse des scores d'identification catégorielle de la voyelle, après la date 4. Cela pourrait être dû aux effets distracteurs des autres voyelles labialisées.

Quant à la voyelle [o], elle ne commence à être perçue de manière catégorielle, à 91%, qu'à la date 6 (figure 1), qui correspond à son début acoustique. Il en va de même pour le [ø] qui n'est identifié significativement (58%) qu'à la date 6, soit à son début acoustique (figure 1).

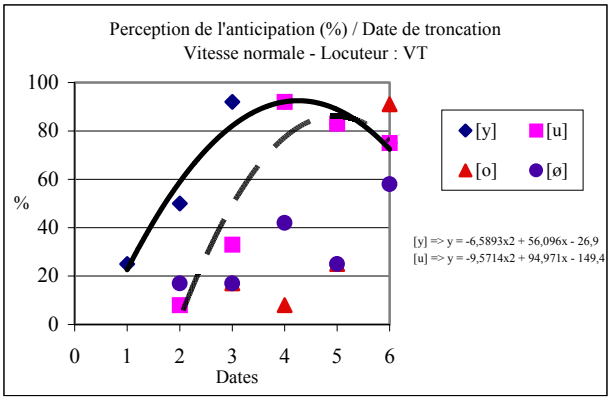

Figure 1 : Pourcentage de réponses correctes (en abscisse) par rapport au point de troncation (en ordonnée). La courbe de tendance du [y] est celle qui est continue, celle du [u] est discontinue. La date 6 correspond au début de la zone stable de la structure formantique des voyelles labialisées. Locuteur VT en vitesse d'élocution normale.

En replaçant ces résultats dans le cadre des données acoustiques analysées dans de précédentes études (voir, par ex., FERBACH-HECKER [2] ; HIRSCH *et al.* [3] ; ROY *et al.* [5]) on peut tirer les enseignements suivants : 1) la labialité du [y] commence, en général, à être perçue entre 100 ms et 80 ms du début de la structure formantique vocalique ; 2) sur le plan acoustique, cette date de perception auditive d'un élément labialisé dans la phase fricative correspond à une inflexion remarquable de la limite inférieure du bruit de friction du [s] ; 3) la perception catégorielle de la voyelle devient possible seulement à 60 ms de l'établissement de la structure formantique stable, lorsque la valeur fréquentielle de la limite inférieure du bruit de friction est proche ou égale à celle du F3 du [y].

La même analyse s'applique aussi au contexte [u] : 1) la labialité de cette voyelle commence à être perçue entre 60 ms et 80 ms de son début acoustique (HIRSCH *et al.*, [3]) ; 2) son identification catégorielle ne devient néanmoins possible que lorsque la valeur fréquentielle de la limite inférieure du bruit, après une inflexion prononcée, avoisine celle du F3 de cette voyelle.

Les données dont nous disposons pour le [o] montrent que sa labialité pouvait être perçue entre 40 ms et 20 ms de son apparition acoustique. Toutefois, nous avons vu précédemment que toute identification catégorielle ne pouvait se faire avant son émergence acoustique même (91% de réponses correctes). Cette remarque est également valable pour la voyelle [ø], pour laquelle aucune identification anticipatoire n'est observée.

L'augmentation de la vitesse d'élocution ne change de façon significative ni le timing de la séquence [is + Vlab], ni le comportement des auditeurs. A noter encore que ces résultats sont structurellement comparables pour la locutrice BV.

Cette première partie de l'étude a donc permis de constater que les voyelles [y] et [u] pouvaient être identifiées de manière catégorielle avant leur début acoustique, et ce malgré la présence d'autres voyelles labialisées dans le corpus du test de perception. En ce qui concerne les voyelles [ø] et [o], elles n'ont pu être correctement reconnues par les auditeurs qu'à leur début acoustique. Dans tous les cas de figure, la labialisation des voyelles est détectable de manière anticipatoire, quel que soit le locuteur et quelle que soit la vitesse d'élocution.

Nous avons vu qu'une difficulté majeure surgit alors lorsqu'il s'agit d'effectuer une tâche de discrimination entre plusieurs voyelles labialisées. Il est judicieux de comprendre la nature de cette difficulté et de procéder à l'analyse des erreurs commises lorsque la tâche de discrimination comporte des candidats concurrentiels significatifs.

3.2 Analyse des erreurs

Signalons, au préalable, que les voyelles non labialisées (qu'elles soient réellement incluses ou non dans les tests de perception) ne présentent aucun effet perturbateur significatif dans la tâche de discrimination des catégories vocaliques labialisées. Au mieux, elles sont choisies à des points de troncation éloignés de l'émergence de la voyelle labialisée, cible à identifier. Par conséquent, les analyses ici ne les retiendront pas dans le calcul des pourcentages de réponses correctes.

Lorsqu'il s'agit d'identifier le [y] (Figure 2), les réponses conflictuelles proviennent essentiellement du [u] et du [ø]. Les scores restent, toutefois, faibles étant donné que le [y] est identifié avec un pourcentage élevé dès le troisième point de troncation (à 92%).

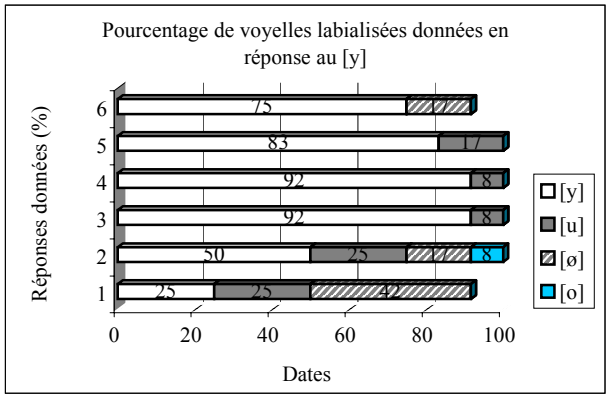

Figure 2 : Total (en pourcentage) des voyelles labialisées données en réponse lorsqu'un [y] était tronqué, aux différents points de troncation. La date 6 correspond au début acoustique de la voyelle dont on s'éloigne de 20 ms, à chaque autre date de troncation.

En ce qui concerne l'identification du [u] (Figure 3), les réponses perturbatrices données à la place de cette voyelle sont le [y] et le [o], mais avec des scores assez faibles, ici aussi.

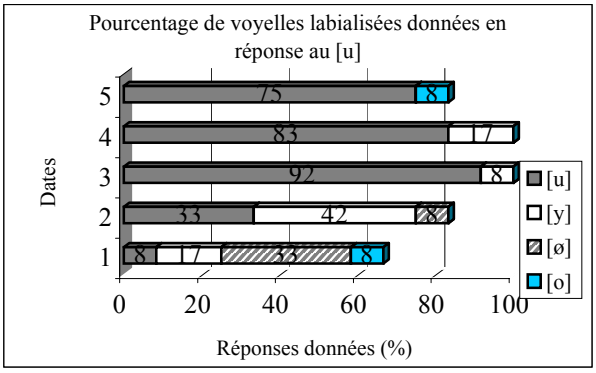

Figure 3 : Total (en pourcentage) des voyelles labialisées données en réponse lorsqu'un [u] était tronqué, aux différents points de troncation. La date 6 correspond au début acoustique de la voyelle dont on s'éloigne de 20 ms, à chaque autre date de troncation.

L'impossibilité d'identifier le [o] de manière anticipatoire est provoquée surtout par les désignations concurrentielles du [u] (à 25% pour les deux voyelles au point de troncation 3), suivi du [ø] (à 17%).

Pour ce qui concerne le [ø], toute difficulté d'identification catégorielle précoce semble être liée à la présence du [y] et du [o] dans le corpus ; ils sont désignés tous les trois à égalité, avec un score de 25% à 20 ms de la voyelle cible [ø].

En résumé, les deux voyelles de petite aperture qui peuvent être identifiées de manière anticipatoire le sont sans véritables candidats concurrentiels. On peut plutôt parler d'éléments perturbateurs à leur identification catégorielle, baissant légèrement le taux de leur reconnaissance. Ainsi, lorsqu'il s'agit de reconnaître le [y], c'est le [u] et le [ø] qui apparaissent également dans les réponses, la présence du [ø], bien qu'avec un score faible, persistant jusqu'au dernier point de troncation. Pour l'identification du [u], on voit apparaître le [y] et le [o], le [ø] se maintenant jusqu'à l'établissement de la structure formantique de la voyelle cible. Dans les deux cas, c'est la voyelle labialisée, ayant le même lieu d'articulation que la voyelle cible à identifier, qui sera le distracteur le plus tenace, même si son score reste peu élevé.

CONCLUSIONS

La tâche de détection de la labialité se fait de manière anticipatoire pour toutes les voyelles labialisées étudiées. L'identification de ce trait est possible dès le début d'une flexion sensible de la trajectoire de la limite inférieure du bruit de friction de la constrictive, en direction de la voyelle labialisée cible. Cependant, la catégorisation anticipatoire de la voyelle, lorsque celle-ci est possible, ne peut intervenir que lorsque la valeur fréquentielle de cette limite inférieure se rapproche de celle du F3 de la voyelle tronquée. Ainsi, les deux voyelles du français dites de petite aperture, à savoir le [y] et le [u], peuvent être reconnues de manière catégorielle bien avant leur début acoustique. Pour ce qui est des voyelles [ø] et [o], les auditeurs sont dans l'incapacité de les identifier avant leur début acoustique. En effet, la limite inférieure du bruit de friction, pour ces deux voyelles, reste stable jusqu'à l'apparition de leur structure formantique.

Pris dans leur ensemble, nos données sur la perception des gestes anticipatoires, en général, nous livrent les renseignements suivants : 1) le début de la protrusion provoque la descente de la limite inférieure du bruit de friction ; 2) la détection auditive de toute labialité se fait à partir de l'événement cinématique « pic de vitesse », correspondant sur le signal acoustique à une flexion remarquable de la limite inférieure du bruit de friction ; 3) l'identification anticipatoire des voyelles [y] et [u] devient possible lorsque le pic de protrusion est obtenu, avec une ouverture minimale de l'aire aux lèvres et avec la valeur fréquentielle de la limite inférieure du bruit de friction avoisinant celle du F3 de la voyelle protruse.

Après avoir mis au jour ces relations sensori-motrices, il nous reste, en perspective, à les compléter par des données visuelles pour connaître la contribution de la dimension anticipatoire visuelle à la récupération de l'élément tronqué (ROY, 2004 [en préparation]).

REMERCIEMENTS

Cette recherche a été financée par le Programme de Recherche "Cognitique ACT1B" 2001-2003, soutenu par le Ministère de la Recherche et des Nouvelles Technologies.

BIBLIOGRAPHIE

[1] AP. Benguerel S. Adelman. Perception of coarticulated lip rounding. *Phonetica,* volume 33, pages 113-126, 1976.

[2] V. Ferbach-Hecker. La perception auditive de l'anticipation des gestes vocaliques en français. *Thèse Nouveau Régime*, 2002.

[3] F. Hirsch R. Sock P.Y. Connan G. Brock. Auditory effects of anticipatory rounding in relation with vowel height in French. In *15th I.C.Ph.S.,*pages 1445-1448, Barcelone, 3-9 août 2003.

[4] J. Lubker R. Lindgren. The perceptual effects of anticipatory coarticulation. P.Hurme (Ed.), In *Speech Research*, pages 252-271, 1982.

[5] J.P. Roy R. Sock B. Vaxelaire F. Hirsch V. Ferbach-Hecker. Auditory Effects of anticipatory and carryover Coarticulation. X-Ray and acoustic data. *I.S.S.P.*, Sydney, 8-10 décembre 2003.

[6] J.L. Schwartz. Perception de la parole : des représentations sensori-motrices à l'émergence des systèmes linguistiques. Ecole thématique. In *Fondements et Perspectives en Traitement Automatique de la Parole*, *Ed. Henri MELONI* Université d'Avignon et des Pays du Vaucluse, pages 9-22, 1995.

[7] B. Vaxelaire R. Sock A. Asci V. Ferbach-Hecker J.P. Roy. Audible and inaudible anticipatory gestures in French. In *15th I.C.Ph.S*, pages 447-450 Barcelone, 1-8 août 2003.