LA PERCEPTION AUDITIVE DE GESTES VOCALIQUES ANTICIPATOIRES

Béatrice VAXELAIRE, Véronique FERBACH-HECKER & Rudolph SOCK
*Institut de Phonétique de Strasbourg • EA 3403 • Université Marc Bloch
22, rue Descartes
67084 Strasbourg (France)
vaxelair@umb.u-strasbg.fr

ABSTRACT

This research, based on X-ray data, examines the relationship between anticipatory labial and lingual gestures and the auditory perception of an upcoming rounded vowel in French Vowel-Consonant-Vowel sequences (V1CV2). V1 is always vowel [a] and V2 vowel [u]; C is either [t] or [k]. The contribution of anticipatory coarticulation to the perception of the rounded element is examined on both the motor (articulatory) and acoustic levels. The robustness of the temporal extent of the perceptual effects is also evaluated under increased speaking rate. The paradigm consists in generating speech samples by representative speakers, then segments are « gated-out » and listeners are asked to judge what the truncated segments were.

1. INTRODUCTION

Les gestes anticipatoires – compris ici comme l'expansion ou l'extension de certains gestes à des segments adjacents ou avoisinants – sont souvent considérés comme apportant une contribution essentielle à la production de la parole. On sait aussi que les auditeurs exploitent des indices précoces liés à ces éléments anticipatoires dans la chaîne parlée (MAEDA, 1999 ; SOCK et al., 1999).

L'utilisation de tels indices, ou du décalage naturel des gestes articulatoires en avance sur le signal acoustique, a été démontrée pour le français, dans le domaine de la perception visuelle, par Cathiard *et al*. 1996. Dans le domaine de la perception de la parole, et sur le plan acoustique-auditif, les résultats dont on dispose (BENGUEREL et ADELMAN, 1976) ne prennent pas en compte la relation entre le niveau articulatoire et ses efficacités acoustiques. La présente recherche suit de très près celle conduite par LUBKER et LINDGREN (1982) sur le suédois. Elle présente des données articulatoires et acoustiques en soulevant quatre questions précises : (1) Quelle est l'extension temporelle de l'anticipation des gestes vocaliques labial et lingual à travers une consonne occlusive ? (2) Est-ce que ces gestes anticipatoires contribuent à la perception auditive précoce d'une voyelle arrondie en français ? (3) Quel est le domaine de l'effet perceptif de ces gestes ? (4) De quelle manière la variation de la vitesse d'élocution et la stratégie individuelle du locuteur pourraient-ils influencer l'extension perceptive de ces gestes anticipatoires ?

2. ANTICIPATION MOTRICE

2.1. Locuteurs, corpus et acquisition des données

Les locuteurs étaient deux adultes de langue maternelle française (A.E. et M.M.), sans antécédent pathologique du conduit vocal et possédaient une audition normale.

Le corpus est constitué des phrases suivantes : « Elle a tout faux » et « Pour accourir », qui fournissent les séquences V_1+C+V_2, où V_1 est la voyelle non-arrondie [a], C est soit [t] soit [k], et V_2 est la voyelle arrondie [u].

Des radiofilms, ainsi qu'un enregistrement simultané du signal audio des productions des locuteurs ont été obtenus sous surveillance médicale, à l'aide d'une caméra 35 mm, d'un magnétophone stéréo et d'un microphone hautement directif (BROCK, 1977).

2.2. Mesures

Des événements temporels ont été détectés sur le signal audio et des relations temporelles spécifiques entre ces événements ont permis de déterminer, dans le domaine VCV, des durées acoustiques correspondant à des gestes articulatoires ouvrants et fermants du conduit vocal.

Des paramètres de mesures ont été déterminés sur les vues de profil, à l'aide d'une grille. Les articulateurs suivis pour l'analyse du comportement anticipatoire étaient les suivants : la protrusion des lèvres (déplacement horizontal), le déplacement vertical de la lèvre inférieure, l'ouverture des lèvres (distance intéro-labiale), le déplacement de la pointe de la langue et le déplacement vertical du dos de la langue.

3. RÉSULTATS

3.1. Timing des gestes

3.1.1. Le contexte apical [atu]

Les données montrent (figure 1 ; locuteur A.E.), en vitesse d'élocution normale, que la protrusion s'installe de manière graduelle (images 1 à 4) avant l'arrivée du contact apical (image 5). Il en va de même pour le déplacement vertical de la lèvre inférieure et du dos de la langue, contribuant eux aussi à la formation de la voyelle arrondie. Ce timing des gestes est structurellement comparable à celui observé pour le locuteur M.M., bien que l'amplitude de ses gestes soit moins prononcée. L'augmentation de la vitesse d'élocution ne modifie pas le timing des articulateurs sur le plan structurel (VAXELAIRE et al., 1999).

En résumé, ces résultats montrent que les gestes liés à la voyelle arrondie sont anticipés bien avant l'apparition de l'occlusive et vont même aussi loin que dans les configurations tardives de la voyelle non arrondie [a]. Qu'advient-il des gestes labiaux (protrusion et ouverture) lorsque le geste du dos de la langue pour la formation de la voyelle [u] est aussi sollicité pour la

production d'une consonne vélaire ? En d'autres termes, est-ce que le conflit entre les gestes vocalique et consonantique, au niveau du dos de la langue, aurait une incidence sur l'extension de l'anticipation des autres structures, même si ces structures sont anatomiquement indépendantes ? L'analyse de la séquence suivante devrait nous fournir des éléments de réponse à cette interrogation.

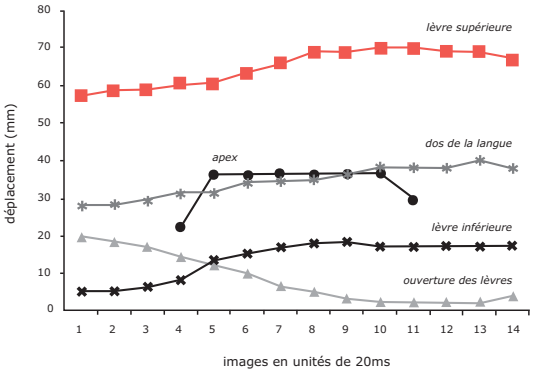

Figure 1
Analyse, image par image, de l'évolution des articulateurs lors de la production de la séquence [atu] en vitesse d'élocution normale ; locuteur A.E.

3.1.2. Le contexte vélaire [aku]

En vitesse d'élocution normale, pour les deux locuteurs, aussi bien la protrusion que le déplacement vertical de la lèvre inférieure et l'ouverture des lèvres varient avant le contact du dos de la langue. En vitesse d'élocution rapide, le timing des gestes des articulateurs est globalement le même, les valeurs de déplacement articulatoire étant relativement moins remarquables, dans ce contexte prosodique aussi, pour le locuteur M.M. En général, on n'observe aucune stratégie compensatoire : il n'y a ni extension temporelle de l'activité anticipatoire labiale dans le temps, ni augmentation de son amplitude, lorsque le geste vocalique du dos de la langue est en conflit avec le geste consonantique (WOOD, 1996).

La section suivante tente : (1) d'établir des relations coarticulatoires sensori-motrices ; (2) de déterminer les effets perceptifs auditifs et l'extension de cette anticipation auditive.

4. PERCEPTION AUDITIVE

4.1. Troncation et élaboration de la bande sonore

Les troncations des signaux acoustiques ont été effectuées sur tous les signaux numérisés, pour les vitesses d'élocution normale et rapide. Le dernier point de troncation coïncide avec l'apparition d'une structure formantique clairement définie de la voyelle labialisée. C'est à partir de ce point de troncation extrême que les autres points de troncation (dorénavant PT) ont été déterminés, vers la consonne précédente, en pas de 10 ms (supérieur aux 50 images/secondes disponibles), jusqu'au début de la plosion-friction de l'occlusive. Ainsi, PT1 est soit à 40 ms (pour M.M.), soit à 60 ms (pour A.E.) du début acoustique de la voyelle arrondie, suivant la durée de l'intervalle entre l'explosion acoustique et l'apparition de la structure formantique stable (VOT de KLATT , 1975).

Pour les phrases « Elle a tout faux » ou « Pour accourir », chaque séquence tronquée comprenait la séquence [elat...] ou [puRak..], plus un taux croissant d'information acoustique contenue dans l'intervalle plosion-friction de l'occlusive, qui précède l'apparition de la voyelle arrondie [u]. La bande sonore contenait 48 stimuli tronqués, disposés en ordre aléatoire (24 dans chaque vitesse d'élocution), qui ont été entendus par les auditeurs. Un bip, servant à alerter les auditeurs de l'imminence d'un stimulus, a précédé chaque séquence de 1,4 secondes. L'intervalle interstimuli était de 4 secondes, avec une pause de 10 secondes après chaque lot de 5 ou 7 stimuli. La bande sonore adébuté avec une liste d'entraînement, suivie des 48 stimuli de l'expérience.

4. 2. Jugements des auditeurs

18 adultes, tous de langue maternelle française, ont servi de sujets pour cette expérience de perception auditive. Ils étaient tous naïfs par rapport au but de l'expérience, et ne présentaient aucun problème d'audition ou de production de la parole. Les tests se sont déroulés dans une salle insonorisée de l'Institut de Phonétique de Strasbourg. Un magnétophone MARANTZ PM D222 a servi pour reproduire les stimuli. La bande sonore a été entendue séparément par les 18 auditeurs, chacun muni d'un casque BEYER DT 770.

Il a été dit aux sujets qu'ils allaient entendre l'une des phrases tronquées suivantes : (1) « Elle a tes faux » (2) « Elle a tout faux » (3) « Elle a ta faux » ; soit trois phrases qui fournissent les séquences [ate], [atu] et [ata] respectivement, pour le contexte apical. En ce qui concerne le contexte vélaire, les auditeurs devaient être attentifs à ces trois phrases tronquées : (1) « Pour acquérir » (2) Pour accourir » (3) « Pour accabler » qui nous livrent les séquences [ake], [aku] et [aka]. Cependant, dans cette expérience précise, seule la phrase numéro 2 dans chaque contexte a été effectivement présentée aux auditeurs, les deux autres servant ici uniquement de distracteurs. Signalons, toutefois, que les résultats obtenus dans une expérience similaire n'ont indiqué aucun changement significatif lorsque les deux phrases tronquées avaient été réellement livrées aux auditeurs (HECKER, 1998).

Des feuilles de réponse leur ont été fournies et durant les 4 secondes d'intervalle entre les stimuli, les auditeurs devaient remplir deux tâches pour chacun des 48 stimuli : (1) marquer à l'aide d'une croix laquelle des trois voyelles [e,a,u] ils pensaient avoir entendu et (2) attribuer un poids de certitude ou de confiance à leur réponse, un poids qui pouvait varier dans une échelle allant de « 1 » à « 5 », où « 1 » indiquait peu de confiance dans leur choix et « 5 » indiquant une certitude absolue.

4.3. Effets perceptifs des gestes anticipatoires : généralités

La valeur moyenne du seuil de confiance, pour les 48 stimuli et les 18 auditeurs (soit 864 réponses en tout) était

de 3 avec un écart type de 1, ce qui nous montre que la plupart des jugements avaient été mis avec un taux de confiance satisfaisant dans les deux vitesses d'élocution. De plus, le pourcentage de réponses correctes est hautement corrélé avec le seuil de confiance (r=0.83 en vitesse d'élocution normale et r=0.86 en vitesse d'élocution rapide), ce qui indique que les sujets étaient confiants tout en étant performants, par rapport à la tâche d'identification qu'ils devaient remplir.

La sévérité des auditeurs dans l'attribution des poids sur l'échelle subjective à 5 poids est manifeste, puisqu'un poids de «2 » révélait des taux de pourcentages d'identification correcte très élevés dans les deux vitesses d'élocution (supérieure, en moyenne, à 80% en vitesse d'élocution normale et à 60% en vitesse d'élocution rapide ; les valeurs intra-conditions sont aussi de cet ordre). En effet, la courbe atteint des valeurs plafond entre les poids « 3 » et « 5 ». Cependant, les sujets avaient tendance à être globalement plus confiants en vitesse d'élocution normale ; mais ceci n'est qu'une tendance. Le seuil de confiance diminue progressivement à mesure que l'on s'éloigne de la voyelle arrondie, *i.e.* à mesure que l'information sensorielle disponible diminue (r=-0.89 en vitesse d'élocution normale et r=-0.97 en vitesse d'élocution rapide).

Cependant, il ne s'agit nullement d'un modèle linéaire mais plutôt d'une hyperbole, étant donné que la réduction du seuil de confiance devient moins sensible à mesure que les points de troncation sont temporellement distants du début acoustique de la voyelle. On retrouve ici aussi le phénomène de l'utilisation d'un timbre vocalique – soit celui du [e] – en réponse « poubelle », lorsque les auditeurs sont incertains par rapport à l'identité de la voyelle tronquée, puisqu'au premier point de troncation, le pourcentage de réponses attribué à cette voyelle [e] était remarquablement plus élevé que celui attribué à la voyelle [a].

4.3.1. Les effets perceptifs de l'anticipation dans le contexte apical [atu]

Les résultats obtenus pour [atu], et donnés dans la figure 2 (locuteur A.E.), montrent que les auditeurs parviennent à identifier une voyelle, même lorsque celle-ci avait été tronquée du signal acoustique. Le pourcentage de réponses correctes est corrélé avec le point de troncation ou la distance de la voyelle : il est élevé lorsqu'on est proche de la voyelle arrondie mais diminue brutalement à partir d'une certaine distance de cette voyelle cible (r=0.73 et r=0.88 en vitesses d'élocution normale et rapide respectivement).

Notons qu'en vitesse d'élocution normale les scores sont très élevés, jusqu'à 50 ms de la voyelle (entre 89% et 100% de réponses correctes), mais chutent remarquablement (à 16%) dès que le point de troncation devient temporellement plus éloigné, soit à partir de PT 2. En vitesse d'élocution rapide, les scores, très élevés près de la voyelle (entre 83% et 100%), chutent brutalement (6% à 28%) au-delà de 40 ms de la voyelle cible, soit à partir de PT 3. Pour le locuteur M.M., le pourcentage de réponses correctes est également corrélé avec le point de troncation (r=0.95 et r=0.98 en vitesses d'élocution normale et rapide respectivement) : il reste élevé seulement

jusqu'à 10 ms (soit au PT 4) de la voyelle arrondie (entre 83% et 100%) dans les deux vitesses d'élocution, mais baisse sensiblement au-delà de cette date (à moins de 50%).

4.3.2. Les effets perceptifs de l'anticipation dans le contexte vélaire [aku]

Dans ce contexte aussi, le pourcentage de réponses correctes est hautement corrélé avec le point de troncation, dans les deux vitesses d'élocution (r=0.87 et r=0.93 en normale et rapide respectivement pour A.E. ; r=0.95 et r=0.81 en normale et rapide respectivement pour M.M.).

Les données montrent que les auditeurs parviennent à identifier la voyelle tronquée jusqu'à 60 ms (PT 1) de l'apparition de la structure formantique (avec 83% à 100% de réponses correctes), en vitesse d'élocution normale. En vitesse d'élocution rapide l'identification se fait jusqu'au PT 2, soit à 50 ms de la voyelle cible (avec 67% à 100% de réponses correctes). Pour ce qui concerne le locuteur M.M., l'identification est correcte à partir de PT 3, soit à 20 ms de la voyelle en vitesse d'élocution normale (avec 83% à 94% de réponses correctes) ; elle se fait plus tôt en contexte rapide, au PT 2, qui est à 30 ms de la voyelle (avec 83% à 95% de réponses correctes).

De manière générale, il semble que l'identification de la voyelle arrondie [u] se fasse plus précocement dans ce contexte vélaire, par rapport au contexte apical précédent et ceci quels que soient le locuteur et la vitesse d'élocution.

4.4. Vitesse d'élocution et procédure de normalisation

Sachant que l'augmentation de la vitesse d'élocution avait effectivement provoqué la compression des durées absolues au niveau acoustique et une accélération des gestes sur le plan articulatoire, il était nécessaire de contrôler le comportement des auditeurs, dans les deux vitesses d'élocution, en ramenant les mesures à des valeurs relatives. Nous avons, en conséquence, *normalisé* les données en calculant le pourcentage de temps pris par chaque PT (distance entre PT*x* et le début de la voyelle arrondie) dans l'intervalle acoustique obstruant (le domaine de l'extension du geste). Les résultats révèlent un scénario comparable à celui obtenu avec les données absolues. On peut donc affirmer que le maintien des différences dans les deux conditions prosodiques indique, même lorsque les données sont normalisées, qu'il s'agit de comportements auditifs reposant sur les relations sensori-motrices spécifiques à chaque vitesse d'élocution, plutôt que de simples effets provoqués par la compression de durées absolues.

5. RELATIONS SENSORI-MOTRICES

En résumé, ces résultats (pourcentages de réponses correctes et seuils de confiance) montrent que le geste d'arrondissement, ainsi que celui de la constriction de la

voyelle labialisée traversent la consonne intervocalique et peuvent même atteindre les dernières configurations du [a]. Cependant, il s'agit là de l'extension anticipatoire de gestes, certes visibles mais pas audibles, puisque la partie auditivement efficace de ces gestes se situe après le relâchement de l'occlusion, en direction de la voyelle labialisée.

La perception anticipatoire de la voyelle [u] est plus précoce dans le contexte vélaire. On s'explique ce phénomène par rapport au relâchement de l'occlusion (apicale ou vélaire). Sachant que ce relâchement arrive plus tôt dans le contexte vélaire par rapport au contexte apical, la perception de la voyelle subséquente se fait, en conséquence plus précocement. La formation de la consonne vélaire (contact – tenue – relâchement) dure moins longtemps que celle de la consonne apicale, en quelque sorte pour permettre l'émergence de la voyelle [u], avec qui elle partage le même lieu d'articulation. Soulignons qu'au relâchement des occlusions, la protrusion labiale est toujours à sa valeur maximale, ce qui renforce l'efficacité anticipatoire auditive de la voyelle.

La réaction des auditeurs aux productions de nos deux locuteurs semble dépendre d'une différence de stratégie anticipatoire. Si, pour A.E., la perception de la voyelle peut se faire dès le relâchement, avec une protrusion maximale, la combinaison de ces deux gestes ne suffit pas pour une perception anticipatoire chez M.M. Pour ce dernier, il faudrait de surcroît leur associer une ouverture minimale des lèvres, un geste qui n'arrive que quelques millisecondes plus tard, par rapport au pic de protrusion et au relâchement. Ainsi, la perception anticipatoire de la voyelle se fait toujours plus tardivement chez M.M. que chez A.E. Étant donné que le timing relatif reste comparable dans les deux vitesses d'élocution, on ne trouve pas de comportement cohérent, quant à la perception auditive anticipatoire suivant la variation de la condition prosodique. Encore une fois, tout dépend de l'arrivée de l'événement articulatoire critique : le relâchement.

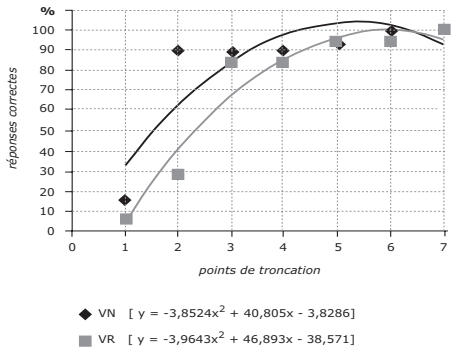

figure 2
Valeurs moyennes de réponses correctes et fonctions d'identification en vitesses d'élocution normale et rapide, par rapport aux points de troncation ; locuteur A.E.

REMERCIEMENTS

Cette recherche a été financée par le Programme « Cognitique » du Ministère de la Recherche (ACT 1b 2001 – 2003).

BIBLIOGRAPHIE

BENGUEREL A.P. ADELMAN S. (1976) Perception of coarticulated lip rounding. *Phonetica 33*, 113-126.

BROCK G. (1977) Méthode de synchronisation graphique image/son pour l'exploitation des films radiologiques. Présentation de l'appareillage réalisé à l'Institut de Phonétique de Strasbourg. Travaux de l'Institut de Phonétique de Strasbourg 9, 221-232.

CATHIARD M.-A. LALLOUACHE T. ABRY C. (1996) Does movement on the lips mean movement in the mind? In D. Stork and M. Hennecke (Eds.), Speechreading by Humans and Machines, NATO ASI Series 150, 211-219.

HECKER V. (1998) Contribution à l'étude de la perception de l'anticipation labiale. Données cinématiques et acoustiques pour le français. *Mémoire de D.E.A. - Institut de Phonétique de Strasbourg*, Université des Sciences Humaines de Strasbourg, 365 p.

KLATT D. (1975) Voice onset time, frication and aspiration in word-initial consonant clusters. Journal of Speech and Hearing Research 18, 686-706.

LUBKER J.F. LINDGREN R. (1982) The perceptual effects of anticipatory coarticulation. In P. Hurme (Ed.), *Papers in Speech Research*, Institute of Finnish Language and Communication, University of Jyväskylä, 252-271.

MAEDA S. (1999) Labialization during /k/ followed by a rounded vowel is not anticipation but the auditorily required articulation. *14 th Int. Congr. of Phonet. Sciences*, San Francisco, Vol. 1, 41-44.

SOCK R. HECKER V. CATHIARD M.-A. (1999) The perceptual effects of anticipatory labial activity in French. 14 th Int. Congr. of Phonet. Sciences, San Francisco, Vol. 5, 2057-2060.

VAXELAIRE B. SOCK R. BONNOT J.-F. KELLER D. (1999) Anticipatory labial activity in the production of French rounded vowels. *14th Int. Congr. of Phonet. Sciences*, San Francisco, Vol. 1, 52-56.

WOOD S. (1996) Assimilation or coarticulation ? Evidence from the coordination of tongue gesture for the palatalization of Bulgarian alveolar stops Journal of Phonetics 24, 139-16.