

Coarticulation en production de parole : aspects acoustiques

René Carré, DDL, Lyon

Résumé.

En production de parole, la coarticulation est un phénomène qui intéresse tous les phonologues et les phonéticiens. En effet, ce phénomène est au centre du passage d'une représentation abstraite à une réalisation périphérique. On peut alors se poser plusieurs questions : Est-ce qu'une partie de la coarticulation est planifiée ? Au niveau périphérique, sur quelle échelle de temps observe-t-on des effets de coarticulation ? Comment peut-on les interpréter ? Comment sont perçus les effets acoustiques liés au phénomène de coarticulation ? L'explication de ce phénomène aux facettes très variées et complexes ne pourrait que s'appuyer sur une théorie générale du mécanisme de la communication parlée, théorie qui n'est pas encore à notre disposition. Notre contribution visera ici à apporter un éclairage particulier en nous appuyant le plus possible sur une approche déductive.

Dans un premier temps, nous rappellerons quelques uns des phénomènes les plus connus liés à la coarticulation, et nous développerons plus particulièrement les résultats obtenus concernant le locus equation (Lindblom, Sussman, ...). Sussman estime que les paramètres du locus permettent de caractériser la consonne ; Krull estime qu'ils donnent une mesure du degré de coarticulation !

Nous développerons ensuite une démarche déductive pour développer un système de communication d'informations à partir des caractéristiques acoustiques d'un tube. Nous faisons l'hypothèse que le système de production de parole, résultat d'une évolution, est optimal : il exploite le plus grand espace acoustique possible ; il est efficace en termes de critère de minimum d'énergie et il est le plus simple possible. C'est un système qui doit être bien adapté pour les tâches à réaliser : la réalisation doit correspondre à l'intention.

Cette approche nous permet de mettre en évidence un espace acoustique de travail qui correspond au triangle vocalique. Elle nous permet aussi de déterminer des gestes spécifiques de déformation du conduit vocal entraînant des trajectoires formantiques correspondantes dans l'espace acoustique. Ces déformations peuvent être modélisées simplement par le modèle en régions distinctives (DRM). Ces régions structurent le tube acoustique de même que les trajectoires correspondantes structurent l'espace acoustique. On note que les régions du modèle correspondent aux lieux d'articulation des voyelles et des consonnes. De telles caractéristiques obtenues sans référence à la parole sont donc intrinsèques au tube acoustique. On utilise ce modèle avec une commande de type syllabique par coproduction de gestes pour étudier les résultats des travaux sur le locus equation de Sussman avec des ensembles V1CV2 pour trois degrés de coproduction. On retrouve les données de Sussman que l'on peut interpréter en termes de degrés de coproduction.

Notre approche déductive met en cause le statut des voyelles en termes de cibles statiques à atteindre. En effet, l'approche évolutive ne peut que chercher à accroître le contraste acoustique selon des directions données (des cibles comme les angles du triangle vocalique sont inconnues). On a donc des informations de type direction de l'évolution dans l'espace acoustique.

Ce comportement nous a conduits à regarder le geste de transition V1V2 comme pouvant être caractérisé non pas par une cible V2 à atteindre mais par une direction dans l'espace acoustique et par une vitesse de transition. On peut ainsi reproduire les voyelles dans l'espace

VF1/VF2 (VF1 est la vitesse de transition de F1,...). Une telle approche est justifiée par de nombreux travaux sur le rôle de la dynamique dans la perception des voyelles (citons les expériences sur le 'silent center' de Strange). Les conséquences de cette approche sur l'analyse du phénomène de coarticulation sont abordées.

Introduction

La parole est le résultat d'une coarticulation d'éléments à définir. Pour illustrer ce terme, on peut rappeler Hockett (1955) qui comparait les phonèmes représentant un mot à des œufs de Pâques et la parole correspondante à une omelette à partir de ces œufs. Par ailleurs, Harris (1951) a noté que le signal segmenté correspondant à 'w' de 'wip' associé au signal représentant 'up', ne donne pas 'wup'. Grâce à la représentation spectrale, on a pu noter des influences réciproques d'un phonème sur ses précédents et sur ses successeurs. Fant (1973) a bien illustré les effets de coarticulation en mettant en correspondance une suite temporelle de phonèmes avec la production de parole résultante : il n'est pas possible de segmenter le signal de parole d'une manière biunivoque, à cause des influences réciproques. De telles influences conduisent à des transitions formantiques dépendant du contexte et à des phénomènes de réduction vocalique et consonantique. Ces observations conduisent à poser les questions suivantes : quelle est la taille de l'unité de coarticulation, i.e. de l'unité de dimension nécessaire et suffisante pour intégrer les influences réciproques observées ? Est-ce que certains aspects de coarticulation sont planifiés au niveau central ? De telles questions sont fondamentales et y répondre permettrait sans doute de proposer un modèle de production de parole et donc d'apporter une contribution fondamentale à une meilleure connaissance du mécanisme de communication parlée.

Nous nous proposons ici de rappeler quelques résultats d'observations du phénomène de coarticulation, puis nous développerons une approche déductive pour proposer une explication du phénomène et enfin, en poursuivant dans la logique de notre démarche déductive, nous formulerons quelques hypothèses pouvant conduire à de nouvelles recherches.

Rappels.

Avec le pattern-playback, l'équipe de chercheurs des laboratoires Haskins (Delattre, *et al.*, 1955) a montré que des synthèses d'ensembles CV, C étant une plosive, pouvaient être obtenues par des transitions du deuxième formant distinctives issues d'un point spécifique dépendant de la consonne considérée et allant vers la voyelle. La consonne est donc représentée par ce point (locus) dans l'espace acoustique et la transition vers la voyelle (coarticulation consonne-voyelle) peut être considérée comme une interpolation entre deux cibles.

L'analyse spectrographique d'ensembles V1CV2 par Öhman (1966) va plus loin dans l'étendue du phénomène de coarticulation. On observe ici que V2 influence les transitions V1C et réciproquement V1 influence CV2. D'après Öhman, la parole serait constituée d'une superposition de gestes consonantiques sur un continuum vocalique. De telles observations ont été effectuées dans des situations encore plus complexes comme sur des ensembles V1CCCCV2 par Benguerel et Cowan (1974), avec V2 voyelle labiale : l'effet des lèvres se retrouve sur V1. De même, Moll et Daniloff (1971) ont noté dans des ensembles CVN, CVVN,... N étant une consonne nasale, l'effet du vélum qui se retrouve sur C.

Kojevnikov et Chistovich (1965) proposent un mécanisme de co-production syllabique pour expliquer les différentes observations. Les réalisations d'une syllabe seraient obtenues par production en parallèle de plusieurs gestes de déformation du conduit vocal.

Certaines caractéristiques du phénomène de coarticulation ont pu être formalisées par Lindblom (1963), puis par Sussman et al (1991) en mettant en évidence l'équation du locus : dans des ensembles V1CV2, on constate que, dans le cas de F2, et pour une même consonne,

les points représentant l'onset de la transition consonantique en fonction de la fréquence du formant pour V2, se situent sur une ligne droite. L'interprétation de ces résultats a été discutée, Sussman estimant que ces droites caractérisent les consonnes et Krull (1987) estimant que la pente de ces droites donne une indication du degré de coarticulation.

Enfin, la réduction vocalique (1963) est une conséquence du phénomène de coarticulation. Le fait que les cibles vocaliques ne soient pas atteintes est interprété par Lindblom comme une difficulté à réaliser l'intention à cause de l'inertie de l'appareil vocal lors de prononciations rapides. A vitesse normale, les cibles vocaliques seraient atteintes et le degré de coarticulation serait minimal.

A ce niveau d'observations plusieurs remarques ou questions peuvent être formulées :

- il semble difficile d'admettre que notre système de communication parlée s'appuie sur des intentions qui ne seraient que rarement réalisées ; le phénomène de coarticulation ne devrait pas présenter une gêne pour le processus.
- sur quelle échelle de temps observe-t-on des effets liés à la coarticulation : sur 1 syllabe, VCV, 2 syllabes ?
- est-ce que les effets de coarticulation sont le résultat de phénomènes périphériques (par déphasages de gestes par exemple) ou bien est-ce planifié au niveau central ? Quelle est la représentation (traits, gestes,...) ? Quelle est la tâche ? Peut-il y avoir des conflits articulatoires (acoustiques) ?
- est-ce que la proposition de Kojevnikov et Chistovich de décrire la parole par une succession de syllabes représentées par des gestes co-produits est acceptable ? Cette proposition revient à mettre en cause une coarticulation réalisée par de seules interpolations entre cibles successives.
- enfin, est-ce que la pente des équations du locus représente une information sur le degré de coarticulation ?

Pour répondre (partiellement) à ces questions, nous allons développer une approche déductive en ne nous intéressant ici qu'aux caractéristiques d'un tube acoustique représentant le conduit vocal excité par une source vocale, sans prendre en compte le conduit nasal et les sources de bruit. On ne s'intéresse donc qu'aux voyelles et aux consonnes plosives.

Approche déductive

L'approche déductive en science a été très bien décrite par Lindblom (Lindblom, 1990). Il explique les systèmes vocaliques à partir des propriétés de l'appareil vocal (qui détermine un espace acoustique : le triangle des voyelles) et du système de perception (qui permet de retenir successivement les sons les plus distants perceptivement (Lindblom, 1986). Il se place donc en dehors de l'ensemble des voyelles. Nous développons ici cette approche en allant plus loin et en nous posons la question : pourquoi cette espace acoustique ? Et aussi, est-ce que cet espace est intrinsèquement structuré ?

Nous partons d'un tube de 17 cm de long défini par sa forme et par ses caractéristiques physiques (acoustiques donc formantiques). L'objectif est de le déformer pas à pas, de manière efficace (une déformation doit entraîner la plus grande variation acoustique possible), en vue d'obtenir le plus grand espace acoustique possible. Un algorithme pilotant automatiquement cette évolution a été proposé (Carré, 2004). Il permet de déterminer l'espace acoustique possible qui correspond en fait au triangle vocalique (en ce sens le conduit vocal est optimal) et des gestes de déformation privilégiés qui peuvent être représentés par le modèle en régions DRM (Mrayati, *et al.*, 1988). On aboutit ainsi à une structuration du tube et de l'espace acoustique par les trajectoires correspondantes (voir Figure 1). Les voyelles se retrouvent sur ces trajectoires mais comme sous produits (Carré, Submitted). Par ailleurs, les régions du modèle correspondent aux lieux d'articulation des consonnes. On peut donc dire que les caractéristiques phonétiques observées (trajectoires vocaliques et lieux d'articulation

des voyelles et des consonnes) sont des propriétés intrinsèques du tube acoustique. On a obtenu par déduction et automatiquement des gestes de déformation du conduit vocal pour la production de transitions vocaliques et consonantiques. Ce type de gestes est exploité en phonologie articulatoire (Browman and Goldstein, 1992). Ils décrivent un comportement essentiellement dynamique. La dynamique ne se réduit pas ici à un passage d'une cible à une autre. L'information est dans la transition caractérisée par une direction et une vitesse de déplacement dans l'espace acoustique.

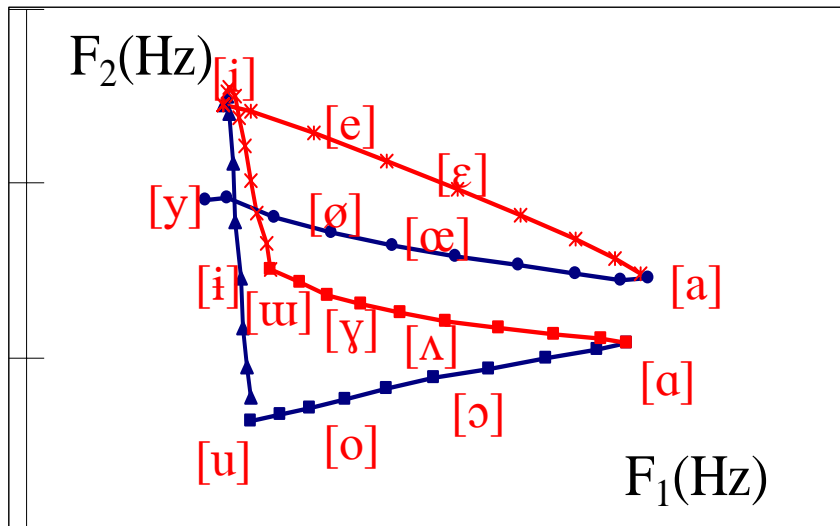


Figure 1. Trajectoires vocaliques obtenues avec le modèle DRM. Les trajectoires de couleur rouge représentent des trajectoires non labialisées ; les bleues sont labialisées.

Pour exploiter ces gestes, une commande de type syllabique peut être retenue. En effet, le rythme syllabique peut découler de la fréquence de résonance de la mâchoire. Dans (Lindblom, *et al.*, 1998), on note que la mâchoire oscille à une fréquence entre 3 et 7 Hz (périodes entre 300 et 140 ms) pour un minimum d'énergie d'excitation ce qui correspond à des durées syllabiques observées. Par ailleurs, en perception, on note que la mémoire de travail est de l'ordre de 200 ms (Carré, *et al.*, 2001). Aussi, pour rendre compte des observations faites par Öhman, on peut imaginer une coproduction syllabique de gestes selon le schéma de la figure 2.

Partant de V1, une commande syllabique abstraite C1V2 en termes de gestes est envoyée au niveau périphérique. La réalisation de chacun des gestes peut varier selon le locuteur, la langue, ... Les variations jouent sur le déphasage des gestes les uns par rapport aux autres (et par rapport à la commande issue du niveau central) et sur la durée (longueur) des gestes. Si le geste vocalique visant à produire V2 (il peut y avoir deux gestes eux-mêmes déphasés : geste de la langue et geste des lèvres) précède le geste consonantique correspondant à C1, alors les trajectoires formantiques à la fin de V1 vont être influencées par V2. Réciproquement, si le geste C1 se termine avant la fin du geste V2, les trajectoires formantiques arrivant à V2 vont être influencées par V1. Avec ce modèle, on peut expliquer les observations de Öhman. Par ailleurs, une nouvelle commande syllabique C2V3, avec une réalisation du geste V3 arrivant avant la fin du geste V2 entraîne une réduction vocalique.

Un tel modèle syllabique rend compte de phénomènes qui s'étendent sur des ensembles VICV2 si l'on accepte que, dans le mécanisme de coproduction syllabique de type CV, le geste vocalique peut précéder le geste consonantique.

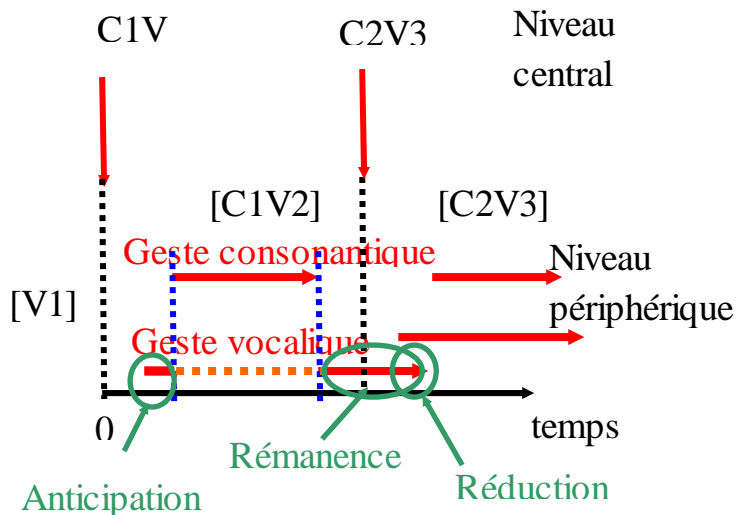


Figure 2. Coproduction syllabique de gestes vocaliques et consonantiques.

Un geste consonantique peut stopper le geste vocalique comme indiqué sur la figure par une phase en pointillés. Est-ce qu'il y a un conflit articuloacoustique ? Des gestes contradictoires correspondent à des trajectoires acoustiques contradictoires ! L'information portée par l'acoustique ne peut faire l'objet de contradictions. Les transitions formantiques d'offset et d'onset consonantique, en cas de phasage favorable des gestes peuvent contenir des informations sur V1, C et V2...

Résultats

On a pu synthétiser les ensembles V1CV2 analysés par Öhman (1966) avec le modèle DRM selon les commandes décrites figure 2. L'occlusion [b] est effectuée sur la région R8 du modèle ; l'occlusion [d] sur la région R6 ; l'occlusion [g] sur la région R5. Les gestes vocaliques et consonantiques étaient synchronisés (Carré and Chennoukh, 1995). Par ailleurs, sur les mêmes ensembles, on a étudié l'équation du locus pour différents déphasages Ph0, Ph1, Ph2, du geste vocalique sur le geste consonantique (figure 3).

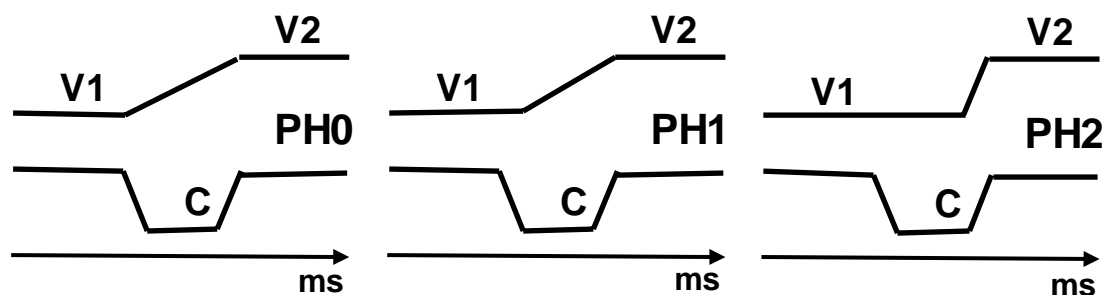


Figure 3. PH0 : le geste vocalique est synchronisé avec le geste consonantique ; PH1 : le geste vocalique commence au début de l'occlusion ; PH2 : le geste vocalique commence à la fin de l'occlusion.

Pour PH0, le geste vocalique est synchronisé avec le geste consonantique : le degré de coarticulation est maximal ; pour PH2, la transition vers V2 s'effectue pendant le relâchement de l'occlusion consonantique : le degré de coarticulation est minimal. La figure 4 représente l'ordonnée à l'origine fonction de la pente des équations du locus pour les plosives [b] (R8), [d] (R6), [g] (R5) et pour les 3 déphasages PH0, PH1, PH2. On note que plus le degré de

coarticulation est important et plus la pente est importante : on est dans la situation correspondant à PH0 c'est à dire à une stricte coproduction des gestes vocaliques et consonantiques. Le degré de coarticulation est donc fonction de la pente de l'équation du locus suivant la prédiction de Krull (1989).

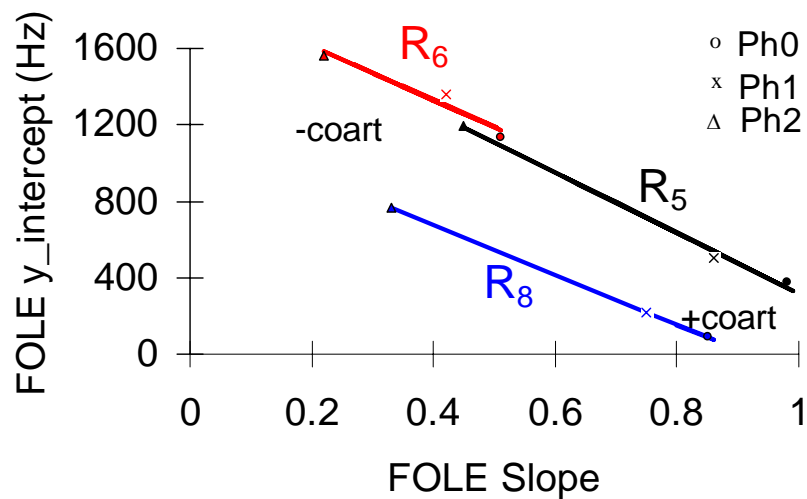


Figure 4. Ordonnée à l'origine fonction de la pente des équations du locus pour les plosives [b] (R8), [d] (R6), [g] (R5) et pour les 3 déphasages PH0, PH1, PH2.

On a ensuite confronté nos résultats de simulation avec les données de Sussman (figure 5). On observe une bonne correspondance. La plosive [d] conduit à un degré de coarticulation faible contrairement à [b], ce qui est un phénomène connu (Farnetani and Recasens, 1999). L'indépendance des gestes de la langue et des lèvres explique pourquoi le degré de coarticulation peut être grand lors de la production de la plosive [b]. Avec des données correspondant à des langues différentes (Sussman, *et al.*, 1993), l'examen de la figure 5 pourra conduire à affirmer que des langues sont plus co-articulées que d'autres,...

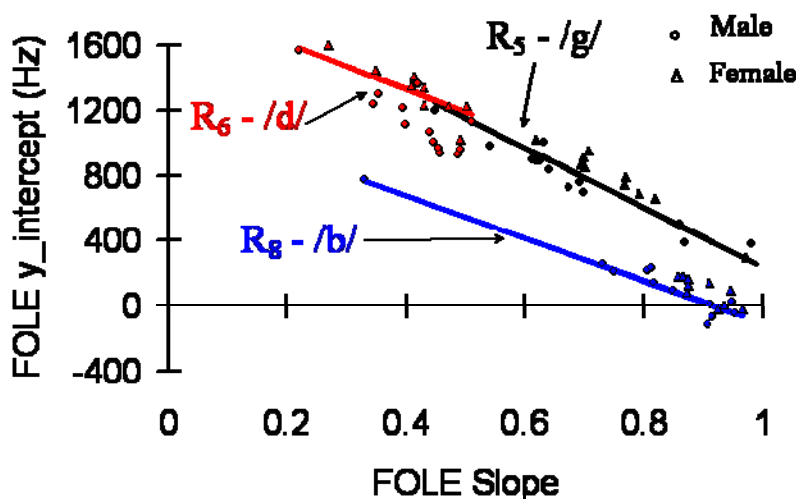


Figure 5. Ordonnée à l'origine fonction de la pente des équations du locus pour les plosives [b] (R8), [d] (R6), [g] (R5) et pour les 3 déphasages PH0, PH1, PH2. On a ajouté les données de Sussman.

Les résultats obtenus par notre démarche déductive correspondant bien aux données, on peut penser que la modélisation retenue possède intrinsèquement une part explicative du phénomène étudié ce qui nous conduit à poursuivre notre processus de déduction.

Discussions

Lors du processus de déduction, l'algorithme utilisé n'a pas pour but d'atteindre des cibles vocaliques inconnues à l'origine. Il s'agit de se déplacer dans l'espace acoustique avec une certaine direction. Sans aller jusqu'à dire que cet algorithme d'évolution, initialisé sur une forme quelconque du tube acoustique, peut simuler l'évolution du fonctionnement du conduit vocal, on peut malgré tout retenir que la direction peut être un paramètre plus important que la cible. Mais, la parole étant un phénomène temporel, une direction ne suffit pas pour caractériser une transition : il est nécessaire d'associer, par exemple, la vitesse de la transition. En fait, les vitesses de transition de F1 et de F2 déterminent aussi la direction dans le plan F1-F2. Si on prend en considération le fait que les durées de transition sont plus ou moins constantes pour un même locuteur (Kent and Moll, 1969; Gay, 1978), alors la vitesse de transition vocalique est un paramètre distinctif. La figure 6 montre les vitesses de transition de F1 pour [ai], [ae] et [aε] prononcés par un même locuteur. Ces transitions sont pratiquement alignées dans le plan F1/F2. On observe que la durée des transitions est effectivement constante (de l'ordre de 180ms). On peut donc distinguer chacune des transitions l'une par rapport aux autres par cette vitesse de transition et ceci tout au long de la transition. On a représenté, pour un même locuteur, les vitesses de transition de F1 et de F2 dans le plan Vitesse de transition de F1 / Vitesse de transition de F2 pour des productions de type [aV] (Figure 7). La vitesse de transition retenue est la vitesse maximale.

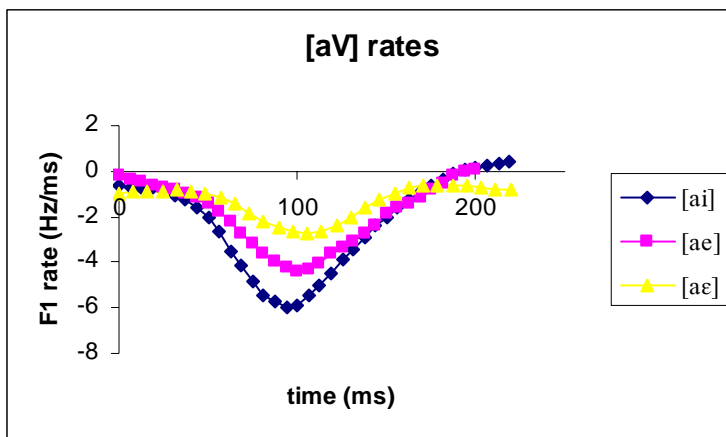


Figure 6. Vitesse de transition de F1 pour trois [aV].

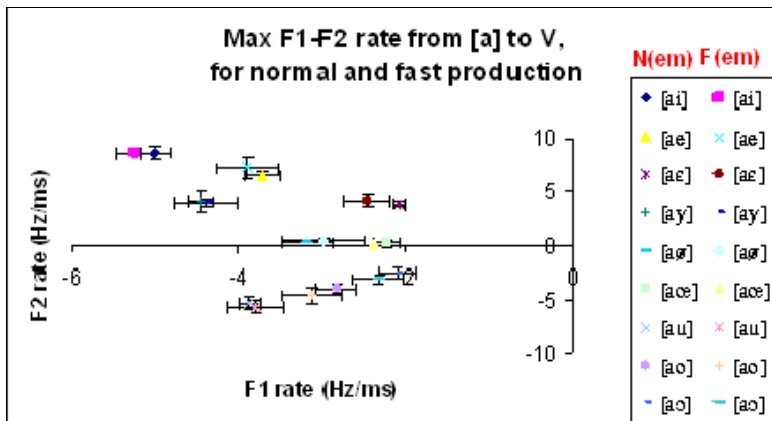


Figure 7. Représentation des transitions [aV] dans le plan Vitesse F1 / Vitesse F2, produites par un même locuteur et pour un débit normal (N) et rapide (F).

Avec une telle représentation, à partir de la voyelle de départ, les vitesses caractérisant la transition vers la voyelle suivante sont nécessaires et suffisantes pour caractériser la voyelle suivante. On a suffisamment d'information au début de la transition et tout au long de la transition pour cette caractérisation. Cette propriété explique les résultats de Strange (1983) dans des expériences de perception où la partie vocalique d'ensembles C1VC2 est extraite et remplacée par du silence ('silent center'). Les voyelles de tels ensembles restent correctement perçues et sont même mieux perçues que les parties vocaliques extraites.

Nous avons mis en œuvre une expérience de perception à partir de transitions extérieures au triangle vocalique, ayant des propriétés dynamiques de transitions de type V1V2 (même vitesse de transition pour F1 et F2) (Carré, Submitted). Ces transitions sont perçues comme les V1V2 correspondantes et, par exemple, selon la dynamique, un même point dans l'espace acoustique peut être perçu comme /a/, ou /u/ ou /o/. Ces résultats montrent que notre système de perception doit être sensible aux dérivées spectrales et, en particulier, aux vitesses de variation de phase des composantes (Divenyi, 2005). Dans une telle approche, s'il n'est pas nécessaire de connaître la cible finale de la transition, il est en revanche essentiel de connaître son point de départ. Sous quelle forme ? Phonétique ? Phonologique ? L'étude effectuée à partir de transitions situées en dehors du triangle vocalique semblerait montrer que le point de départ doit être une donnée phonologique.

Cette approche dynamique conduit à repenser fondamentalement le phénomène de production de parole (Carré, *et al.*, 2007). Dans une perspective dynamique, les voyelles et les consonnes ont des statuts assez proches. Les vitesses de variations formantiques sont porteuses d'information sur les voyelles et les consonnes et sont peut-être plus invariantes que les valeurs formantiques des max, min, onsets, offsets, ... si on compare hommes, femmes et enfants. Dans ce cas, les valeurs formantiques caractérisant classiquement les voyelles ne seraient pas 'intrinsèques' mais 'extrinsèques', les caractéristiques intrinsèques définissant la dynamique. Dans une telle approche, on ne peut que souscrire à la position de Fowler donnant au temps une dimension intrinsèque pour définir les éléments du code (Fowler, 1980). Par ailleurs, la réduction vocalique n'implique plus un calcul d'overshoot perceptif (Lindblom and Studdert-Kennedy, 1967).

Dans une perspective dynamique, les effets significatifs de coarticulation se déplaceraient donc des cibles vers les vitesses de transition, ce qui demande une exploitation des résultats d'études spécifiques plutôt rares dans la littérature d'autant plus que les techniques d'analyse ne sont pas adaptées pour mesurer correctement les vitesses de transition.

Bibliographie

- Benguerel, A. and Cowan, H., 1974. Coarticulation of upper lip protrusion in French. *Phonetica* 30, 41-55.
- Browman, C. P. and Goldstein, L., 1992. Articulatory phonology: An overview. *Phonetica* 49, 155-180.
- Carré, R., 2004. From acoustic tube to speech production. *Speech Communication* 42, 227-240.
- Carré, R., Submitted. Dynamic properties of an acoustic tube: Prediction of vowel systems.
- Carré, R., Submitted. Signal dynamics in the production and perception of vowels. In I. Chitoran, C. Coupé, E. Marsico and F. Pellegrino, (Eds.), *Approaches to Phonological Complexity*, Mouton de Gruyter, Berlin, New York.
- Carré, R., Ainsworth, W. A., Jospa, P., Maeda, S. and Padeloup, V., 2001. Perception of vowel-to-vowel transitions with different formant trajectories. *Phonetica* 58, 163-178.
- Carré, R. and Chennoukh, S., 1995. Vowel-consonant-vowel modeling by superposition of consonant closure on vowel-to-vowel gesture. *J. of Phonetics* 23, 231-241.
- Carré, R., Pellegrino, F. and Divenyi, P., 2007. Speech dynamics: epistemological aspects. In: *Proc. of the ICPHS, Saarbrücken*, pp. 569-572.
- Delattre, P. C., Liberman, A. M. and Cooper, F. S., 1955. Acoustic loci and transitional cues for consonants. *J. Acoust. Soc. Am.* 27, 769-773.
- Divenyi, P. L., 2005. Frequency change velocity detector: A bird or a red herring? In D. Pressnitzer, A. Cheveigné and S. McAdams, (Eds.), *Auditory Signal Processing: Physiology, Psychology and Models*, Springer-Verlag, New York, pp. 176-184.
- Fant, G., 1973. *Speech sounds and features*. MIT Press, Cambridge, MA.
- Farnetani, E. and Recasens, D., 1999. Coarticulation models in recent speech production theories. In W. J. Hardcastle and N. Hewlett, (Eds.), *Coarticulation Theory, Data and Techniques*, University Press, Cambridge.
- Fowler, C., 1980. Coarticulation and theories of extrinsic timing. *J. of Phonetics* 8, 113-133.
- Gay, T., 1978. Effect of speaking rate on vowel formant movements. *J. Acoust. Soc. Am.* 63, 223-230.
- Harris, C., 1951. A study of the building blocks in speech. *J. Acoust. Soc. Am.* 53, 962-969.
- Hockett, C., 1955. *Manuel of phonology*. In (Eds.), *Publications in Anthropology and Linguistics*, Indiana University Press, Bloomington.
- Kent, R. D. and Moll, K. L., 1969. Vocal-tract characteristics of the stop cognates. *J. Acoust. Soc. Am.* 46, 1549-1555.
- Kozhevnikov, V. A. and Chistovich, L. A. (1965) "Speech, articulation, and perception," JPRS-30543. NTIS, US Dept. of Commerce.
- Krull, D., 1987. Second formant locus patterns as a measure of consonant-vowel coarticulation. *Phonetic Experimental Research at the Institute of Linguistics, University of Stockholm (PERILUS) V*, 43-61.
- Krull, D., 1989. Second formant locus pattern and consonant-vowel coarticulation in spontaneous speech. *Phonetic Experimental Research at the Institute of Linguistics, University of Stockholm (PERILUS) X*, 87-108.
- Lindblom, B. (1963) "On vowel reduction," Report N° 29. Stockholm: The Royal Institute of Technology, Speech Transmission Laboratory.
- Lindblom, B., 1963. Spectrographic study of vowel reduction. *J. Acoust. Soc. Am.* 35, 1773-1781.
- Lindblom, B., 1986. Phonetic Universal in Vowel Systems. In J. J. Ohala and J. J. Jaeger, (Eds.), *Experimental Phonology*, Academic Press, Orlando, pp. 13-43.
- Lindblom, B., 1990. Phonetic contents in phonology. *Phonetic Experimental Research at the Institute of Linguistics, University of Stockholm (PERILUS) XI*, 101-118.

- Lindblom, B., Davis, J. H., Brownlee, S. A., Moon, S. J. and Simpson, Z., 1998. Energetics in phonetics: A preliminary look. In: Proc. of LP'98, Columbus, pp. 401-415.
- Lindblom, B. and Studdert-Kennedy, M., 1967. On the role of formant transitions in vowel perception. *J. Acoust. Soc. Am.* 42, 830-843.
- Moll, K. and Daniloff, R., 1971. Investigation of the timing of velar movements during speech. *J. Acoust. Soc. Am.* 50, 678-684.
- Mrayati, M., Carré, R. and Guérin, B., 1988. Distinctive regions and modes: A new theory of speech production. *Speech Communication* 7, 257-286.
- Öhman, S., 1966. Coarticulation in VCV utterances: spectrographic measurements. *J. Acoust. Soc. Am.* 39, 151-168.
- Strange, W., Jenkins, J. J. and Johnson, T. L., 1983. Dynamic specification of coarticulated vowel. *J. Acoust. Soc. Am.* 74, 695-705.
- Sussman, H. M., Hoemeke, K. and Ahmed, F., 1993. A cross-linguistic investigation of locus equations as a relationally invariant descriptor for place of articulation. *J. Acoust. Soc. Am.* 94, 1256-1268.
- Sussman, H. M., McCaffrey, H. A. and Matthews, S. A., 1991. An investigation of locus equations as a source of relational invariance for stop place categorization. *J. Acoust. Soc. Am.* 90, 1309-1325.