

CARACTERISTIQUES ACOUSTIQUES DE LA PREMIERE CONSONNE
DANS UN GROUPE DE DEUX OCCLUSIVES

Alain MARCHAL et Anne FOTI

U.A. 261, CNRS, Institut de Phonétique,
29, av. Robert-Schuman, 13621 Aix-en-Provence, France

INTRODUCTION

L'étude détaillée de l'organisation des appuis linguo-palatins dans des groupes d'occlusives observée à l'aide de l'électropalatographie (1) nous a permis de mettre en évidence des phénomènes de :

- coproduction consonnantique
- coproduction vocalique
- double occlusion
- relâchement de C1 caractéristique de la production de "clics".

Nous nous attacherons dans cette communication à décrire les caractéristiques acoustiques de la première consonne dans un groupe d'occlusives. Nous essayerons de voir s'il est possible de reconnaître automatiquement C1 à partir des informations spectrales ou temporelles apportées par le bruit de relâchement.

Corpus et méthode

Les données proviennent de 3 corpus différents qui ont été répétés de 3 à 6 fois par 3 locuteurs. Il s'agit d'un ensemble de 800

phrases naturelles de 4 à 6 syllabes où les groupes d'occlusives formées par les consonnes /t, d, k, g/ apparaissent dans plusieurs contextes syntaxiques, précédées et suivies des voyelles /i, a, u/. Le débit rapide favorisait les phénomènes de coarticulation.

La segmentation des événements articulatoires a été opérée à partir de la séquence des appuis linguo-palatins relevés par le système de palatographie dynamique de Montréal (2). Nous avons délimité les événements suivants :

- 1 - Début de la préparation de C1; de C2.
- 2 - Implosion de C1; de C2.
- 3 - Tenue de C1; de C2.
- 4 - Relâchement de C1; de C2.

Nous avons examiné l'évolution correspondante du signal acoustique. Le son synchronisé a été digitalisé et analysé. Compte tenu de la sensibilité des extracteurs de maxima fréquentiels au bruit (conditions d'enregistrements, rapport signal sur bruit, nombre de points d'analyse, gain, fenêtrage, nombre de paramètres, ...), nous avons jugé dans la perspective d'une application en

Se 37.4.1

reconnaissance de la parole qu'il fallait

- 1 - utiliser des procédures automatiques
- 2 - admettre la possibilité d'erreurs de détection
- 3 - comparer les résultats d'analyse provenant de systèmes différents pour éviter les artefacts induits par une méthode et une situation expérimentale donnée.

Nous avons relevé les données acoustiques suivantes :

les valeurs des trois premiers maxima de fréquence pour les voyelles et les consonnes et leur amplitude relative dans les parties stables et dans les transitions, les durées des tenues de C1 et de C2, le VOT, et la durée de bruit de constriction, soit un ensemble de 39 mesures pour chaque suite V1C1C2V2.

Nous avons utilisé pour l'analyse : 1) le logiciel de traitement de signal de parole "SIGNEX" implanté à l'Institut de Phonétique d'Aix-en-Provence (3), 2) le logiciel commercial "ILS" de SIGNAL TECHNOLOGY, 3) un banc de filtres numériques (4).

Résultats et discussion

Vu l'importance du corpus et le très grand nombre de mesures à effectuer, le travail d'analyse et d'interprétation des données se poursuit encore actuellement. L'état de la recherche est le suivant : le signal acoustique de tous les enregistrements a été segmenté et les 8000 marqueurs corres-

pondant aux événements articulatoires ont été placés. L'analyse acoustique complète a porté pour deux locuteurs sur les phénomènes temporels et sur la structure acoustique du bruit de relâchement pour /t, k/. Nous avons interprété les résultats en fonction de la possibilité d'écrire des règles de production simples et sûres.

Un invariant spectral ?

La thèse voulant qu'un invariant spectral permette d'identifier la consonne occlusive a été examinée (5). Lorsqu'on calcule les valeurs moyennes des trois premiers pics depuis les basses fréquences indépendamment de leur amplitude relative, tous corpus confondus, on trouve

/t/ : 1304 Hz;	/k/ : 700 Hz
2887 Hz	1985 Hz
4416 Hz	3806 Hz

si l'on considère seulement les valeurs absolues, il semblerait possible de distinguer ainsi les deux occlusives; il apparaît toutefois que les valeurs d'écart-type sont trop importantes; respectivement pour /t/ : 926 Hz; 1320 Hz; 1670 Hz et pour /k/ : 547 Hz; 904 Hz; 1569 Hz. Celles-ci font apparaître une zone de recouvrement importante entre les maxima 1 et 2 et entre les maxima 2 et 3.

On pourrait objecter que l'amplitude relative des zones de bruit est essentielle pour les consonnes et que les valeurs de fréquence doivent être ordonnées en fonc-

Se 37.4.2

tion de ce paramètre.

Nous avons remarqué que le spectre de bruit de relâchement de C1 dans un groupe d'occlusives était caractérisé par une amplitude et par une dynamique faibles. Les valeurs de maxima d'énergie par ordre décroissant et les écarts-type font apparaître une zone de recouvrement moins grande pour /t/ et /k/ mais encore trop importante pour pouvoir distinguer avec confiance les deux consonnes.

/t/		/k/	
moy.	écart	moy.	écart
Max 1 : 1983 Hz	1607 Hz	996 Hz	787 Hz
Max 2 : 2873 Hz	1716 Hz	1860 Hz	1157 Hz
Max 3 : 3776 Hz	1789 Hz	3623 Hz	1714 Hz

Nous n'avons pas pu identifier un ou plusieurs facteurs stables permettant de réduire la dispersion autour de la valeur centrale. L'influence de V1 et V2 ne suffit pas à expliquer la variabilité. La méthode d'extraction des maxima ne peut non plus être invoquée car la dispersion se retrouve à peu de choses près identique dans les trois analyses.

Un invariant temporel ?

La physiologie de l'articulation (Hardcastle 1976) nous enseigne que la masse linguale mise en mouvement et que l'activité musculaire requise pour la production de /t/ et /k/ sont très différentes. On pour-

rait alors faire l'hypothèse qu'une cible peut être atteinte plus rapidement que l'autre et que cette différence cinétique, due à des contraintes fonctionnelles pourrait se retrouver dans le signal acoustique. Nous avons examiné les durées des tenues seules et les durées des bruits de relâchement. Il est apparu assez rapidement que la durée du bruit de relâchement varierait considérablement et que dans un corpus important les différences phonémiques intrinsèques /k>t/ étaient annulées par des variables mal contrôlées (conditions d'enregistrement, proximité du micro, gain..., conditions phonétiques : débit, ...).

Nous avons alors concentré notre attention sur les tenues silencieuses; soit l'intervalle séparant l'implosion de C1 de l'apparition de la trace du bruit de relâchement sur les tracés acoustiques. Nous avons aussi relevé la tenue silencieuse totale de C1 et C2.

Les résultats complets en ms apparaissent dans le tableau suivant :

	/t/	/tk/	/k/	/kt/
	C1	C1C2	C1	C1C2
Moy. :	75	177	75	155
Ecart :	12	27	14	29

On note que les moyennes des durées des tenues silencieuses de /t/ et /k/ sont très comparables. Ce qui importe toutefois, c'est que le rapport de la durée de C1 sur la durée totale de C1 et C2 est significative-

ment différent.

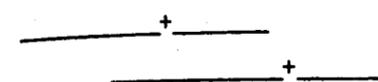


schéma des durées relatives des tenues de C1 et C2

On observe la même tendance pour l'enchaînement de /d/ à /g/ ou de /g/ à /d/. Notre observation sur le français rejoint les résultats de Hardcastle & Roach (1977) (6) sur l'anglais; ceci semble bien indiquer que ce phénomène temporel est lié à une contrainte d'ordre physiologique de bas niveau. Il pourrait s'agir d'une boucle tactile déclenchant le mouvement vers la deuxième occlusive dès que l'information tactile appropriée est transmise par le contact linguo-palatal de la première occlusive. Dans le cas d'un enchaînement "alvéo-dentale---vélaire", l'information pourrait être envoyée par des mécano-récepteurs de la langue quand elle fait contact sur les côtés du palais et sur les alvéoles. A partir de cette position, tout ce qui est nécessaire pour la production d'un /k/ est une contraction d'un muscle intrinsèque de la langue : le longitudinal inférieur. Un tel mouvement peut se produire relativement rapidement parce que les muscles intrinsèques ont des temps de contraction rapides. Dans le cas d'un enchaînement "vélaire---alvéo-dental", la situation n'est pas aussi simple et plusieurs gestes deviennent anta-

gonistes. Le mouvement vers le /t/ implique un mouvement vers l'avant et vers le haut. Celui-ci est dû essentiellement à l'activité du génioglosse, muscle extrinsèque lent qui provoque une remise en position de tout le corps de la langue. L'élévation de l'apex étant produite par la contraction du longitudinal supérieur et du muscle transverse. Il n'est donc pas étonnant que cette différence fonctionnelle se retrouve dans les données temporelles.

CONCLUSION

Dans le cas de l'enchaînement de deux occlusives, il ne semble pas possible d'identifier un invariant spectral permettant la reconnaissance de la première consonne. L'organisation des données temporelles et notamment le rapport des tenues silencieuses de /t/ et /k/ mettent en évidence une contrainte physiologique universelle qui pourrait être utilisée en reconnaissance automatique de la parole. Le groupe /tk/ est plus long que le groupe /kt/ et surtout le rapport de la durée de la tenue silencieuse de C1 sur la durée de la tenue silencieuse totale est environ de 0,5 lorsque C1 est un /k/ et s'abaisse autour de 0,4 lorsque C1 est un /t/.

REFERENCES BIBLIOGRAPHIQUES

- (1) Marchal, A. (1985), "L'électropalatographie : contribution à l'étude de la coarticulation dans les groupes d'occlusives", th. de doct. d'Etat, Nancy.
- (2) Marchal, A. (1984), "Le système d'électropalatographie de Montréal : contribution à l'étude des occlusives du français", *Trav. Inst. Phon. Aix*, 9 : 267-341.
- (3) Espesser, R. & Nishinuma, Y. (1984), "Traitement de signal sous le système UNIX : description des commandes", Institut de Phonétique d'Aix-en-Provence.
- (4) Manceron, F. (1982), "Contribution à l'analyse spectro-temporelle du signal de parole considéré comme une unité d'impulsion acoustique", th. doct. ing. LIMSI, Paris.
- (5) Blumstein, S. & Stevens, K. (1979), "Acoustic invariance in speech production : evidence from measurement of the spectral characteristics of stop consonants", *J. Acoust. Soc. Am.*, 66 (4) : 1001-1013.
- (6) Hardcastle, W. & Roach, P. (1977), "An instrumental investigation of coarticulation in stop consonant sequences"; *Work Prog. Phonet. Lab. Univ. Reading*, 1 : 27-44.

АН
ле
(Ц
се
э
ка
и
ис
ра
ты
на
та

ме
ре
та
со
ре
де
по

мы
пр
ла
на
ци
на
ра
ра
ме
Од
це
Ос
ци
но

ГО
МО
РИ
МЕ
НЕ
РИ
В
Ч

Об
ст