

## SEGMENTATION ARTICULATOIRE ET ACOUSTIQUE DES DIPHTONGUES ALLEMANDES

SIBYLLE VATER

A la différence des recherches consacrées aux voyelles homogènes, l'analyse phonétique des diphtongues a été longtemps entravée par les limites de perception imposées à notre oreille.

Nous considérons ici seulement quelques obstacles relatifs au temps. Notre perception d'un phénomène sonore dépend à la fois de la constante d'intégration auditive (de l'ordre de 50 millisecondes [ms]) et du temps que l'oreille nécessite pour détecter la structure formantielle d'un son, c'est-à-dire pour reconnaître le timbre (environ 30 ms). La durée totale qui résulte de ces deux valeurs temporelles ( $\sqrt{50^2 + 30^2}$ ) représente approximativement 60 ms (cf. Winckel 1960:99, d'après M. Joos). Ceci est dix fois le temps qu'une voyelle exige pour se constituer, et c'est environ le quart de la durée moyenne d'une diphtongue tonique allemande (240 ms).

D'autre part, la qualité sonore qui forme la première phase des diphtongues en question est plus étendue et plus intense que pour les timbres décroissants. De plus, les formants 1 et 2 du timbre initial se situent directement dans la zone de 800 à 1000 cps ou non loin d'elle (structure compacte) à la différence de ceux qui surgissent seulement au cours de la diphtongue. Or, dans cette zone, l'oreille est particulièrement sensible à des changements de fréquence et d'intensité (Winckel 1960:89). Ces avantages réunis permettent de conclure que la qualité sonore initiale dont nous venons de parler possède une rémanence auditive supérieure à 50 ou 60 ms. Celle-ci affecte, en premier lieu, la phase diphtongale qui est comprise entre le 110e et le 160e milliseconde environ (sur une durée totale de 240 ms) et pendant laquelle l'articulation et, par conséquent, les formants se modifient continuellement; autrement dit, couvant sous l'effet de la rémanence, les nouveaux timbres ne parviennent pas à se maintenir, mais ils chevauchent tout en allant decrescendo.

La rémanence auditive est un bien pour la cohérence du complexe diphtongal. Néanmoins, celle-là gêne le phonéticien désireux de connaître la structure phonique réelle sous-jacente. A cette fin, deux genres d'analyse ont été coordonnés: (1) la décomposition articulatoire obtenue par la radiocinématographie, et (2) la segmentation acoustique.

### 1. L'ANALYSE DE L'ARTICULATION AVEC LE CONCOURS DE LA RADIOCINÉMATOGRAPHIE

Lors de la réalisation de nos radiofilms, le temps de pose a été de 1 ms par image (1 ms/i), indépendamment du débit auquel les prises de vue se succèdent. Si, par exemple, une séquence de film a retenu 70 images par seconde (70 i/s), il faut compter 1/70 s, c'est-à-dire 14 ms, pour passer d'une image à l'autre. Or, l'image même — l'équivalent du temps de pose — correspond seulement à la quatorzième partie de cette durée. Les autres treize millièmes de seconde qui se sont écoulés jusqu'à la prise suivante n'ont pas été exploités par la caméra bien que l'articulation ait continué pendant cet intervalle. Plus le nombre des images est petit par unité de temps, plus les phases d'articulation non filmées s'allongent. Ainsi, d'une diphtongue mesurant 280 ms, un film de 70 i/s ne retient que 20 images équidistantes d'un milliseconde chacune. Ceci signifie que, pendant 260 ms (presque 93 % de la durée totale), l'articulation n'est pas rendue visible. Ces réflexions devraient décourager le phonéticien. Or, il n'en est rien. Nous possédons deux invariables qui permettent d'interpoler aisément les phases d'articulation non filmées. Ce sont, d'une part, les images d'une milliseconde que nous obtenons réellement, d'autre part, la durée pratiquement constante des intervalles.

### 2. LA SEGMENTATION ACOUSTIQUE

Elle permet, d'un côté, l'écoute isolée de fragments diphtongaux, de l'autre, l'établissement de spectres dont les segments correspondants ne dépassent pas 5 ms, c'est-à-dire une vibration double de voix féminine. Une durée plus brève ne serait plus utile. Les phases diphtongales qui présentent un intérêt particulier peuvent donc être analysées par des tranches successives de 5 ms sans qu'un instant échappe à ce traitement.

L'analyse acoustique est à la fois précise et continue; en revanche, la radiocinématographie fournit des informations détaillées mais discontinues. Si la segmentation acoustique et l'articulatoire sont parfaitement synchronisées, nous constatons que, sur trois fragments acoustiques successifs de durée minimale, un seul correspond à une image filmée. Cependant, l'examen minutieux de chaque période ne s'impose que rarement même lorsque les changements formantiels sont très rapides.

La segmentation acoustique et articulatoire aboutit à des résultats juxtaposés. Ce pointillisme n'a de sens que si nous rattachons les valeurs obtenues à une synthèse. Cette dernière est possible grâce à un système dont les paramètres sont les fréquences et le temps. Ainsi, nous calculons le taux de variation de la fréquence et l'inflexion des formants aux instants les plus caractéristiques.

A titre d'exemple, nous relevons les expériences suivantes: (1) traitement acoustique et radiocinématographique complet de la diphtongue [a<sup>e</sup>] dans le mot *feig*, (2) ampu-

tations pratiquées dans le cœur de la diphtongue [a<sup>e</sup>] du mot *seit*, (3) synthèse de [a<sup>e</sup>] à partir des mots *sah*t (prononcé [za:t] avec forte intensité) et *seht* (à faible intensité).

*Université de Genève*  
*Suisse*

#### BIBLIOGRAPHIE

Brichler-Labaeye, C.

1970 *Les voyelles françaises, mouvements et positions articulatoires à la lumière de la radiocinématographie* (Paris, Klincksieck).

Simon, P.

1967 *Les consonnes françaises, mouvements et positions articulatoires à la lumière de la radiocinématographie* (Paris, Klincksieck).

Winckel, F.

1960 *Phänomene des musikalischen Hörens* (Berlin, Hesse).