
DES INDICES ACOUSTIQUES AUX TRAITS PERTINENTS

PIERRE DELATTRE*

INTRODUCTION

Les recherches sur la nature des traits pertinents par lesquels les phonèmes d'une langue donnée se distinguent les uns des autres doivent commencer au niveau acoustique (physique) plutôt qu'articulatoire, nous semble-t-il, parce que le fonctionnement des traits est un phénomène de perception, et que, du moins dans l'état présent des facilités de recherche, seuls les corrélatifs acoustiques peuvent être soumis au test perceptuel. En effet, grâce aux techniques expérimentales de synthèse établies par les Laboratoires Haskins, la manipulation de formes acoustiques appropriées, au moyen d'un synthétiseur de parole, permet d'isoler, de modifier, de combiner, chacun des éléments acoustiques (physiques) qui contribue à la perception du lieu ou du mode d'articulation d'un phonème donné, permet d'en faire varier une seule dimension à la fois et de tester à l'oreille les effets de ces variations; cette manipulation des formes synthétiques permet même d'établir quels sont les éléments acoustiques qui ne contribuent pas à la perception de tel ou tel trait pertinent ou qui n'y contribuent que dans une certaine mesure. Or il n'est pas possible d'isoler, de modifier, de disséquer, de combiner les éléments d'un trait *articulatoire* et de tester séparément à l'oreille l'effet des variations de chacun des éléments. Il est possible par contre de décrire avec un assez haut degré d'exactitude les mouvements articulatoires élémentaires qui correspondent à des indices acoustiques élémentaires une fois que ceux-ci ont été bien établis, limités et spécifiés par les tests perceptuels. Si donc on ne parvient pas à isoler et spécifier les corrélatifs articulatoires des traits pertinents des phonèmes directement, on peut du moins le faire indirectement par l'intermédiaire des indices acoustiques obtenus par les tests de perception.

Pour illustrer l'état présent des connaissances acoustiques sur la parole, et surtout les avancées récentes dans le domaine des modes d'articulation (les lieux d'articulation ont déjà fait l'objet de nombreuses études), nous nous proposons de présenter ici le tableau complet de la structure acoustique des consonnes intervocaliques d'une langue, en tant que système. Nous avons choisi le français plutôt que l'anglais (pourtant mieux étudié) parce que les problèmes d'acoustique qu'on y rencontre sont d'une nature plus universelle.

* Université de Californie (Santa Barbara).

La Figure 1 est un tableau structural des formes spectrographiques des consonnes françaises entre voyelles [ε]. L'arrangement des formes en sept colonnes et en huit rangées est entièrement basé, non sur les traits articulatoires mais sur les traits acoustiques et le genre de perception qu'ils caractérisent. Ainsi les consonnes d'une

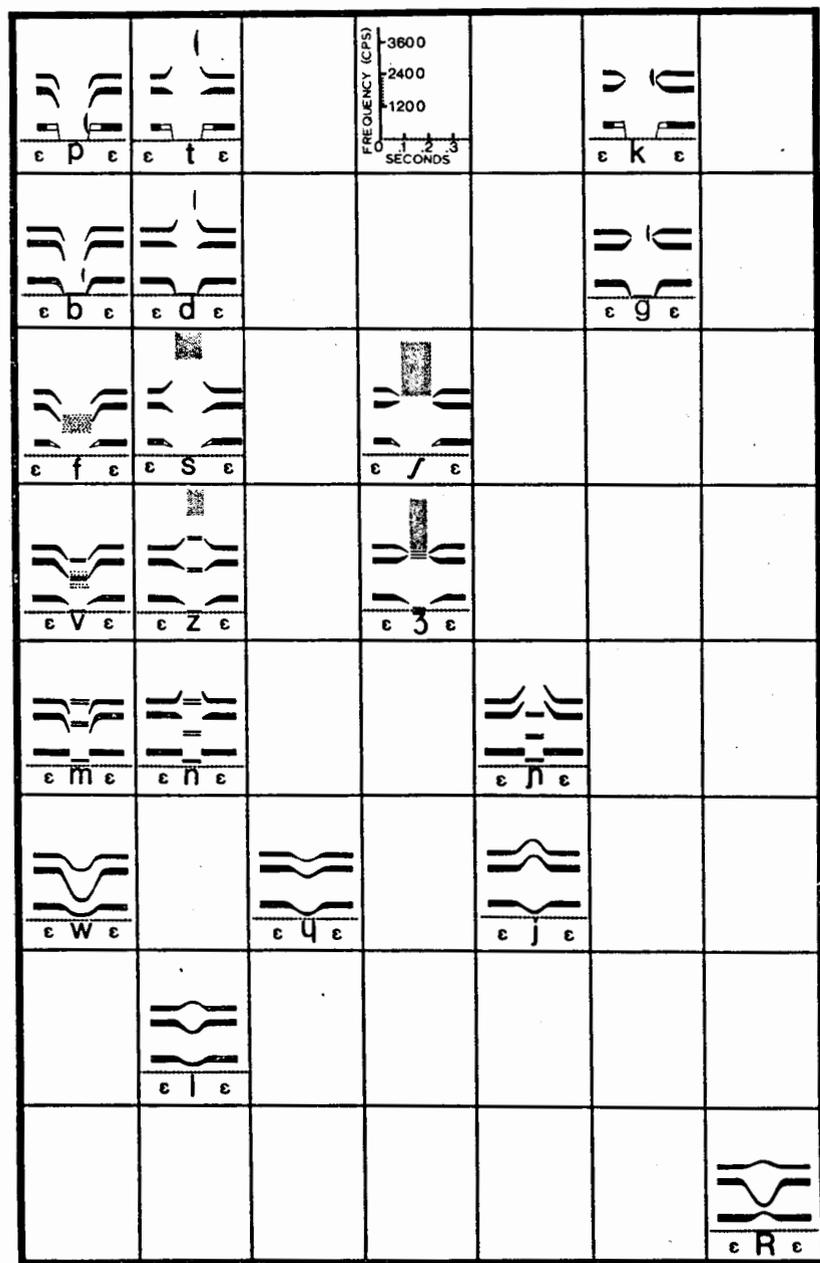


Fig. 1.

même colonne ont en commun des indices *acoustiques* qui ont quelque rapport avec la perception des lieux d'articulation, et celles d'une même rangée ont en commun des indices *acoustiques* qui ont quelque rapport avec la perception des modes d'articulation.

DÉMONSTRATION DE PAROLE ARTIFICIELLE

[Ici, les enregistrements au magnétophone des formes acoustiques de la Figure 1 ont été entendus sous trois aspects différents: a) transformées en son voisé par un synthétiseur du type „playback“ à 50 bandes de fréquence, sans variations de la fréquence fondamentale; b) transformées en son voisé par un synthétiseur du type „voback“ à 18 bandes de fréquence, avec variations de fréquence fondamentale; c) transformées en son chuchoté par le même voback n'utilisant que la source de bruit. Les voyelles chuchotées font parfaitement entendre la distinction dite „sourdes/sonores“, ce qui montre bien le poids des indices acoustiques autres que la fréquence fondamentale produite par la vibration des cordes vocales. (Ces indices seront démontrés plus loin). D'ailleurs, l'intelligibilité des syllabes chuchotées, dont les formants s'obtiennent par filtrage de bruit blanc, est tout aussi claire que celle des syllabes voisées. Cela démontre la prépondérance du „formant“ dans la perception: les sons vocaliques ne sont pas caractérisés par des harmoniques mais par des formants (bandes étroites de son filtrées par la résonance sélective du tractus vocalique), que la source au larynx soit composée de son périodique ou de son turbulent.]

LES INDICES DE LIEU D'ARTICULATION

Sur la Figure 1, on peut voir que toutes les formes spectrographiques d'une même colonne ont en commun deux indices acoustiques de transition (deuxième et troisième formants), et que de plus toutes les explosives et fricatives d'une même colonne ont en commun la fréquence du bruit, que ce soit un bruit d'explosion ou un bruit de friction.

Première colonne (en partant de la gauche)

Les transitions du deuxième formant visent toutes un locus de très basse fréquence (700 cps environ). Les transitions du troisième formant visent toutes un locus de 2000 cps environ. Les bruits, sous forme d'explosion pour (*p*) et (*b*) ou de friction pour (*f*) et (*v*), ont tous une fréquence très basse (1000 cps environ).

[On donne le nom de *locus* au point de convergence de toutes les transitions de formant qui sont reliées à la perception d'un même lieu d'articulation consonantique ou de tout autre trait pertinent, quelle que soit la voyelle qui précède ou qui suit. Il est pratique de considérer le locus comme un point virtuel placé à quelque 5 cs des têtes de transition dans le cas des explosives et des fricatives; et comme un point réel dans tous les autres cas. La théorie du locus décrit la direction que les transitions de formant doivent prendre pour contribuer à la perception d'un lieu d'articulation ou de tout autre trait pertinent consonantique dans une langue donnée. Etant fondée non

sur l'observation seule des spectrogrammes mais essentiellement sur des tests perceptuels de variables synthétiques, cette théorie permet de juger, par extension, les conditions dans lesquelles telle transition de formant ne peut pas contribuer à la perception de tel trait pertinent, donc de déterminer sur les spectrogrammes de parole naturelle quelle est la contribution perceptuelle de chaque transition de formant. Ainsi lorsque la coarticulation empêche une transition de se diriger vers son locus, comme c'est le cas pour [kɔ, ko, ku], les tests de perception indiquent clairement que cette transition ne contribue pas à la perception du trait pertinent voulu et qu'un autre indice s'est substitué à la transition inefficace. Les tests perceptuels de parole synthétique ont d'ailleurs été confirmés par des expériences de parole naturelle faites par découpage de bandes magnétiques.]

Deuxième colonne

Le locus des transitions du deuxième formant est à environ 1700 cps — compromis économique possible entre le locus de (t, d, n) à 1800 cps et celui de (s, z, l) à 1600 cps. Le locus des transitions du troisième formant est à environ 2700 cps. Les bruits sont relativement hauts: vers 4000 cps pour les explosions de (t, d), en dessus de 3500 cps pour les frictions de (s, z). Dans cette colonne, en dehors du (l), le locus du troisième formant a plus de pouvoir perceptuel que dans les autres colonnes. Le locus du troisième formant est légèrement moins élevé et moins efficace pour (l) que pour les autres consonnes de la même colonne, mais il ne semble pas nécessaire d'ouvrir une colonne particulière pour cette résonnante. Bien que latéral, un (l) français est assez antérieur pour s'identifier *acoustiquement* avec les consonnes dentales. Ce ne serait pas le cas d'un (l) anglais.

Troisième colonne

Pour la semi-voyelle (y), il a fallu ouvrir une colonne spéciale. Bien qu'arrondie du point de vue labial, son locus de deuxième formant coïncide plutôt avec celui des dentales (1700 cps), lesquelles ont généralement les lèvres écartées, et bien que palatale du point de vue lingual, son locus de troisième formant (très efficace) coïncide plutôt avec celui des alvéolaires (2200 cps). Cela nous montre que les traits articulatoires n'agissent pas sur les traits acoustiques d'une manière indépendante mais d'une manière combinée. On y trouve aussi un rappel du fait que le seul terme „labial“ ne suffirait pas à définir le commun dénominateur articulatoire de la première colonne. La cinéradiographie montre en effet que les consonnes de la première colonne ont toutes, en plus de leur labialisation, une cavité post-labiale relativement grande, soit par recul, soit par abaissement de la langue.

Quatrième colonne

Le locus des transitions du deuxième formant est à environ 2000 cps, celui des transitions du troisième formant à environ 2200 cps, et le bruit de friction à une fréquence intermédiaire, au dessus de 2000 cps. Le fait que la différence entre les

traits acoustiques de la quatrième colonne et ceux de la deuxième sont si marqués montre combien le léger recul du point d'articulation, de pré-alvéolaire à post-alvéolaire, peut être critique.

Cinquième colonne

Le locus des transitions du deuxième formant est à environ 3000 cps, et celui des transitions du troisième formant à environ 3500 cps. Nous remarquons que la consonne dite „palatalisée“ (ɲ) partage les traits acoustiques de lieu d'articulation avec la consonne palatale (j), ce que font également les autres consonnes dorso-palatalisées (tʃ, dʒ, kʲ, gʲ). Le terme „palatalisé“ ne désigne donc pas un mode d'articulation mais un lieu, et est équivalent de dorso-palatal.

Sixième colonne

Le locus des transitions du deuxième formant est à environ 3000 cps et celui des transitions du troisième formant à environ 2200 cps. Le bruit d'explosion est à une fréquence intermédiaire qui suit la tête de transition du deuxième formant et varie donc considérablement selon la fréquence du deuxième formant de la voyelle qui suit.

Un problème assez grave se présente ici du fait que plus la voyelle qui suit (k, g) est arrondie et postérieure, plus il est difficile à la transition du deuxième formant de se diriger vers le très haut locus de 3000 cps et par conséquent de contribuer à la perception du lieu d'articulation palatovélaire. Nous avons là une excellente illustration des méfaits de la coarticulation: l'arrondissement des lèvres et le recul de la langue pour une voyelle arrondie postérieure, comme (o) agrandit et ferme la cavité antérieure ce qui donne une note de résonance très basse à la tête de transition. Par bonheur, plus la voyelle est arrondie et postérieure, plus l'explosion de (k, g) a de pouvoir perceptuel. C'est donc l'indice d'explosion qui se substitue à celui de transition.

Septième colonne

De même que pour (y), il a fallu créer une colonne spéciale pour (R), la forme acoustique des deuxième et troisième formants présentant une combinaison unique — la transition du deuxième formant s'apparente à celle de (w) et celle du troisième à celle de (l).

On remarque d'emblée que les relations acousticoarticulatoires qui prévalaient dans le cas de (w) et (l) n'ont plus cours ici — la consonne (R) n'est ni labio-vélaire ni apico-dentale, elle est pharyngale et pendant son articulation les lèvres sont le plus souvent écartées. Cela nous montre une fois de plus que ce n'est pas la labialisation en soi qui fait descendre la transition du deuxième formant mais toute condition articulatoire passible de faire résonner la cavité antérieure à une note basse: ici la note de tête du deuxième formant est basse à cause de l'immensité de la cavité antérieure (elle va du haut-pharynx aux lèvres); cette note serait d'ailleurs plus

basse même pour (*R*) que pour (*w*) si la grandeur de la cavité antérieure n'était compensée par l'écartement des lèvres.

Il nous faut dire un mot du premier formant de (*R*). On voit sur la Figure 1 que son locus est très haut (environ 700 cps), ce qui indique, d'après la théorie acoustique, que la constriction de cette consonne est dans la moitié postérieure du tractus vocalique, c'est-à-dire dans le pharynx. On pourrait donc dire que le haut locus du premier formant est un indice de lieu d'articulation. Il nous paraît cependant préférable de le ranger parmi les indices de *mode* d'articulation. Deux raisons suffiront à notre argument. a) Le haut locus du premier formant se trouve dans toute la catégorie des consonnes pharyngales, que leur constriction se trouve dans le haut pharynx, ou le bas pharynx, comme c'est le cas dans les langues arabes, ou qu'elle soit intermédiaire comme en français ou en allemand. b) Partout ailleurs que pour les (*R*), la transition du premier formant est exclusivement un indice de mode — c'est en partie la hauteur du locus de premier formant qui distingue, par exemple, les semi-voyelles des latérales, ou les occlusives des fricatives. Structuralement, donc, nous préférons considérer la hauteur du locus de premier formant des (*R*) français comme un indice de mode d'articulation.

LES INDICES DE MODE D'ARTICULATION

Examinons maintenant les rangées de la Figure 1. Ici il sera question des transitions de *premier* formant plutôt que de deuxième ou de troisième; de la *durée* des bruits plutôt que de leur fréquence; du *tempo* des transitions supérieures plutôt que de leur locus; du *caractère* des joints périodiques qui résonnent pendant la tenue d'une constriction plutôt que de leur fréquence.

Première et deuxième rangées

Voyons d'abord ce que ces deux rangées ont en commun.

1. La forme et le tempo de leurs transitions peuvent s'appeler „directe-rapide“. Les transitions des explosives, comme celles des fricatives et des nasales, sont *directes* par opposition aux transitions des résonnantes (nasales, semi-voyelles et liquides) qui sont *inverses* (leur courbe change de sens, comme celle d'un *S*); elles sont rapides, comme les transitions des nasales, par opposition à celles des fricatives et des glissantes (semi-voyelles et liquides). Ainsi les caractéristiques de forme et de tempo rangent les transitions en quatre types: directe-rapide pour les explosives, directe-lent pour les fricatives, inverse-rapide pour les nasales, et inverse-lent pour les glissantes.

2. Les transitions du premier formant visent la fréquence la plus basse possible, ce qui fait dire, sans doute arbitrairement, que leur locus est à zéro. Le locus de premier formant des fricatives, des nasales et des semi-voyelles est plus haut (250 cps), celui des latérales encore plus haut (400 cps) et celui des pharyngales le plus haut de tous (700 cps).

3. Les consonnes des deux premières rangées ont du bruit dans la tenue. Elles partagent cet indice avec les fricatives, mais pas avec les résonnantes, qui, elles,

sont privées de toute turbulence. D'autre part, les explosives n'ont pas de joint bas. Elles ne partagent cette absence de joint qu'avec les fricatives, la présence du joint bas étant l'indice le plus caractéristique des résonnantes (nasales, semi-voyelles et liquides). On peut définir le joint bas comme un formant de tenue, donc de faible intensité, qui relie les transitions implosives et explosives de premier formant. L'opposition (mutuellement exclusive) entre les bruits et les joints bas semble capitale parmi les indices consonantiques. Elle explique pour la première fois la différence essentielle entre les fricatives sonores et les semi-voyelles et liquides. (Étant corrélative de l'opposition transition directe/ transition inverse, elle peut être considérée comme redondante si l'on sépare, en ce qui concerne les transitions, la notion de forme de celle de tempo.)

4. Le bruit des explosives est court, par opposition à celui des fricatives qui occupe toute la tenue. C'est là évidemment l'indice principal pour distinguer les fricatives des explosives. Mais ce n'est pas le seul — le tempo de toutes les transitions et le locus des transitions de premier formant contribuent aussi à les distinguer.

5. Les quatre catégories d'indices que nous venons de décrire s'appliquent aussi bien à la seconde qu'à la première rangée de la Figure 1. Examinons maintenant les indices qui distinguent la première rangée de la seconde. Pour les explosives, les expériences de synthèse ont permis d'isoler six différents indices de cette opposition: toutes les formes spectrographiques de la première rangée ont une plus longue tenue que celles de la deuxième rangée; elles ont aussi une voyelle plus courte, une réduction (cutback) de la transition du premier formant, un bruit plus intense, une absence de fondamental, et enfin un peu d'aspiration si le sujet parlant est un aristocrate du Faubourg Saint-Germain.

Troisième et quatrième rangées

Examinons d'abord les indices acoustiques qui sont communs à ces deux rangées.

1. Leurs transitions sont directes-lentes — directes, comme celles des explosives, mais lentes, contrairement à celles des explosives.

2. Le locus du premier formant des fricatives est plus élevé que celui des explosives. On peut soit le considérer comme virtuel à environ 250 cps, soit le considérer comme réel à environ 350 cps.

3. Les fricatives ont un bruit de tenue et n'ont pas de joint bas. Nous l'avons déjà dit, c'est sur cette opposition d'indices acoustiques que repose principalement la distinction entre fricatives sonores et résonnantes.

4. Les indices acoustiques de bruit sont dits longs pour indiquer qu'ils occupent toute la durée de la tenue.

5. Ainsi les troisième et quatrième rangées ont quatre types d'indices acoustiques en commun. Voyons maintenant ce qui les oppose. Pour les fricatives, les expériences de synthèse ont permis d'isoler six indices acoustiques, dont cinq ont déjà été mentionnés pour les explosives. Les formes de la quatrième rangée ont des joints périodiques qui relient les transitions implosives aux transitions explosives de deuxième et

transition inverse correspond à une absence de bruit à l'initiation du mouvement d'ouverture (nasales) ou avant (glissantes).

Une transition rapide correspond à une occlusion orale complète pendant la tenue (explosives et nasales); une transition lente correspond à une occlusion orale incomplète pendant la tenue (fricatives et glissantes).

Un locus de premier formant à zéro correspond à une interruption complète de la colonne d'air par une constriction dans la partie antérieure du tractus vocalique.

Un locus de premier formant à 250 cps correspond à une interruption centrale (non latérale) incomplète dans la moitié antérieure du tractus.

Un locus de premier formant à 400 cps correspond à une interruption latérale incomplète dans la moitié antérieure du tractus.

Un locus de premier formant à 700 cps correspond à une constriction centrale dans la moitié postérieure du tractus vocalique, c'est-à-dire dans le pharynx.

Un bruit sans joint bas pendant la tenue ou à son terme correspond à un souffle dont la pression s'exerce derrière une constriction étroite ou une occlusion complète et qui passe, soit pendant la tenue (fricatives), soit au terme de la tenue (explosives).

Un joint bas sans bruit pendant la tenue correspond à un souffle qui traverse une constriction orale ou vélique (voile du palais abaissé), laquelle n'est pas assez étroite pour ralentir l'écoulement du souffle et permettre à la pression d'air d'augmenter derrière la constriction. La constriction ne sert en somme qu'à séparer les cavités de résonance.

Un bruit court correspond à l'ouverture brusque d'une fermeture complète derrière laquelle la pression du souffle a été en augmentant.

Un bruit long (occupant toute la tenue) correspond à un souffle qui traverse de façon ininterrompue une constriction étroite.

Des joints de tenue discontinus (consonnes nasales) correspondent à la réunion et à la séparation des organes de constriction pendant que le voile du palais est abaissé. (Les joints sont discontinus même si la consonne nasale est précédée et/ou suivie d'une voyelle *nasale*.) Les deux instants de discontinuité (aux jonctions des joints avec le terme des transitions implosives et le départ des transitions explosives) correspondent aux instants où, a) l'occlusion buccale se fermant, le souffle est brusquement dérivé vers le nez, et b) l'occlusion buccale se rouvrant, le souffle est brusquement dérivé vers la bouche. A ces deux instants le système de résonance du tractus change radicalement.

Des joints hauts sans joint bas (comme pour les fricatives sonores) correspondent à un état de résonance du tractus qui permet à la seule cavité antérieure (entre la constriction et les lèvres) de résonner effectivement. L'inefficacité de la cavité postérieure vient de ce que la glotte fait double emploi: elle s'ouvre à l'arrière pour laisser passer le souffle librement et se ferme à l'avant pour vibrer périodiquement. L'ouverture à l'arrière fait communiquer le pharynx avec la trachée, ce qui abaisse et réduit sa fréquence de résonance. Cette inefficacité doit aussi s'attribuer au fait

que la constriction des fricatives — même des fricatives sonores — est trop étroite pour laisser passer les vibrations de la cavité postérieure.

Une longue tenue acoustique correspond à une longue tenue articuloire des organes au degré d'aperture de cette consonne.

Une longue voyelle précédente correspond à une longue tenue des organes au degré d'aperture de cette voyelle.

La réduction des transitions de premier formant (cutback) correspond au degré exceptionnel de pression musculaire qui s'exerce lorsque les organes se joignent ou se rapprochent étroitement pour une consonne dite sourde.

Un bruit fort correspond à un haut degré de pression derrière une constriction tendue.

L'aspiration correspond à un délai dans la mise en vibration des cordes vocales après l'ouverture brusque de l'occlusion, délai pendant lequel la glotte est ouverte et le débit de souffle est violent (sept fois plus grand que pour une voyelle chuchotée) ce qui ne permet qu'à la cavité antérieure de résonner efficacement (pour l'aspiration, comme pour le *(h)*, le premier formant vocalique est fort atténué, tandis que pour la voyelle chuchotée, il est renforcé).

SOMMAIRE DES CORRÉLATIFS ACOUSTIQUES

La Figure 2 résume les résultats des pages précédentes.

Le carré central présente le même arrangement de consonnes que le tableau structural des formes spectrographiques de la Figure 1, sauf qu'il y est réservé deux rangées aux occlusives vélares (*k*) et (*g*) parce que leurs spécifications acoustiques sont différentes selon qu'elles sont suivies de voyelles écartées (symbolisées par *k_eg_e*) ou de voyelles arrondies (symbolisées par *k_og_o*). Nous avons en effet vu plus haut que devant les voyelles arrondies-postérieures, les transitions du deuxième formant ne peuvent pas se diriger vers le locus de 3000 cps et que c'est l'explosion, au lieu de la transition, qui fait percevoir le lieu d'articulation vélaire. En somme, acoustiquement parlant, on identifie les occlusives vélares par deux types d'allophones extrêmement différents l'un de l'autre.

L'arrangement du carré de consonnes montre que l'analyse acoustique répartit les consonnes françaises en sept catégories de lieu d'articulation (rangées sur la Figure 2) et huit catégories de mode d'articulation (colonnes sur la Figure 2). Mais il en est tout autrement des indices acoustiques mêmes qui permettent l'identification linguistique de ces consonnes.

D'après la Figure 2, l'identification linguistique des consonnes utilise au moins 11 indices acoustiques de lieu (extensions du carré vers le haut) et 17 indices acoustiques de mode (extension du carré vers le bas). Les matrices de signes „plus“ et „moins“ indiquent pour chaque consonne le choix qui s'opère parmi les indices de lieu et de mode.

Les 11 indices de lieu se rangent en trois catégories: A. La direction de la transition du deuxième formant, spécifiée par un locus. B. La direction de la transition du

troisième formant, également spécifiée par un locus. C. La fréquence du bruit (seulement les explosives et les fricatives), lequel peut se réaliser par une explosion ou une friction. En français, la largeur de bande et l'intensité ne sont pas pertinentes.

Les 17 indices de mode se rangent en six catégories:

1. La forme et le tempo des transitions.
2. Le locus de la transition de premier formant.
3. La présence de bruit, contre la présence de joint bas.
4. La longueur du bruit (s'il y en a). Cela remplace l'opposition conventionnelle: interrompu/non-interrompu.
5. La continuité contre la discontinuité des joints (s'il y en a).
6. Les facteurs de durée et d'intensité, communément appelés „sourd“ contre „sonore.“

Les cases remplies des matrices révèlent les faits suivants:

a) Deux consonnes ayant le même mode d'articulation ne se distinguent jamais l'une de l'autre par moins de deux indices de lieu ou par plus de trois.

b) Deux consonnes ayant le même lieu d'articulation ne se distinguent jamais l'une de l'autre par moins de deux indices de mode ou par plus de 11 (par plus de 5 si les indices de voisement sont comptés séparément).

Les conditions (a) et (b) montrent bien la complexité des corrélatifs objectifs des traits pertinents: Il n'y a pas d'indice acoustique qui soit distinctif par lui-même, c'est-à-dire qui puisse convertir une consonne en une autre par substitution; il faut au moins deux indices et jusqu'à onze pour produire une distinction de sens. Le corrélatif acoustique d'un trait pertinent (appelons-le „indice distinctif“) se présente donc sous la forme d'un faisceau d'indices contributifs, en français, et il y a de bonnes raisons de croire que le corrélatif articulatoire d'un trait pertinent (appelons-le „articulation distinctive“) n'est pas moins complexe.

[Il a été donné en conclusion une démonstration sonore de la contribution perceptuelle à l'identification des consonnes françaises, de chacun des 11 indices de lieu d'articulation et des 17 indices de mode, au moyen de diapositives des formes spectrographiques et d'enregistrements magnétiques de parole artificielle.]

Disons pour terminer que si les traits pertinents sont des signaux perceptuels qu'on ne peut pressentir qu'indirectement à travers leurs corrélatifs acoustiques et articulatoires, et que les corrélatifs articulatoires ne peuvent être spécifiés qu'une fois accompli l'isolement des corrélatifs acoustiques, il n'est peut-être pas possible de toucher les traits pertinents de plus près qu'en arrivant à une connaissance suffisante de ce qui est distinctif dans les signaux acoustiques. C'est à quoi nous avons visé ici.¹

DISCUSSION

Achmanova:

The "vision", the functional, the *humanistic* approach Prof. Delattre has been speaking about *will* become useful and promising only when the *instrumental* studies have been, if not completed, at any rate carried far enough. I could not follow him in thinking that the minute instrumental "preliminaries" can be regarded as completed.

Tillmann:

Thank you for the many interesting details you presented. What I want to suggest is that you should restrict your results to artificial speech. Your cues and features are, perhaps, sufficient but not necessary properties of acoustic signals, perceived as speech. Nevertheless I am sure that the findings through analysis by synthesis are a good base to proceed from to the more sophisticated study of natural speech signals.

Wode:

Prof. Delattre has made it clear that he accepts a one-to-many relationship between acoustic cue and articulatory correlates. For establishing his acoustic cues he assumes a direct one-one correspondence between the auditory impression (satisfactory synthesis of the French consonants in terms of listener tests) and acoustic stimulus. This is in contradiction to much evidence that points to the fact that there is at least a many-to-one (if not a many-to-many) correspondence between acoustic cue and auditory impression. On the basis of this one would have to doubt the relevance of Prof. Delattre's conclusions concerning natural speech.

¹ The research reported herein was performed pursuant to a contract with the United States Department of Health, Education and Welfare, Office of Education.