

LE JEU DES TRANSITIONS DE FORMANTS ET LA PERCEPTION DES CONSONNES

PIERRE DELATTRE

INTRODUCTION

Entre la tenue d'une consonne et la tenue de la voyelle suivante, c'est-à-dire entre la phase fermée et la phase ouverte d'une syllabe du type [ba], il se produit un mouvement articulatoire ouvrant combiné à un déplacement complexe des organes. Ce mouvement physiologique est reflété sur les spectrogrammes acoustiques par des changements de fréquence généralement rapides et continus dans les *formants*, c'est-à-dire dans les concentrations d'énergie acoustique qui correspondent aux fréquences des cavités du pavillon. Tout comme les notes de résonance du système des cavités changent continuellement pendant le déplacement des organes, de même les formants acoustiques changent continuellement de fréquence. C'est à ces changements de fréquence des formants que la terminologie acoustique a donné le nom de *transitions*.

Deux faits fondamentaux ressortent des recherches acoustiques des quelque douze dernières années. (1) Ce sont les trois premiers formants, les trois formants les plus bas, qui contiennent l'essentiel des traits linguistiques. Les formants supérieurs contribuent surtout à caractériser la voix individuelle et ont un rôle linguistique très limité dans l'intelligibilité totale. (2) Les transitions de formants jouent, dans la perception de la parole, un rôle autrement plus important que ne le laisserait entendre le choix peu heureux du terme "transition". Au lieu d'être une phase secondaire, ou négligeable, comme on l'a longtemps cru, les transitions sont à la clef même de la perception de la consonne. L'analyse spectrographique a fait soupçonner leur importance. La synthèse spectrographique de la parole l'a confirmée, et elle a permis d'établir avec rigueur leurs dimensions acoustiques de durée, de fréquence et d'intensité pendant la continuité dynamique du changement complexe causé par le mouvement des organes.

Il est bon de mentionner pourtant que la part que prennent les transitions dans la perception des consonnes varie d'une consonne à l'autre. Ainsi elle est plus grande pour les sonores que pour les sourdes, pour les occlusives labiales que pour les palatales, pour les fricatives à turbulence effacée, comme [f], [θ], que pour celles à turbulence stridente comme [s], [ʃ].

Nous nous limiterons ici à la question des transitions acoustiques. Nous négligerons donc les traits acoustiques qui sont reliés à la phase fermée des consonnes, et

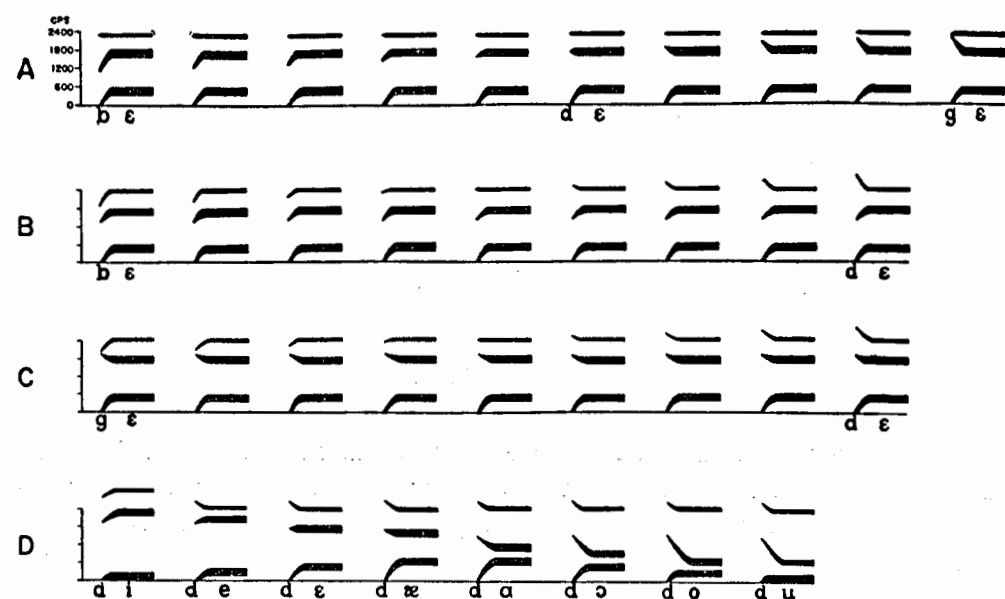


Fig. 1. Schémas spectrographiques indiquant l'effet auditif qui résulte de changements dans les transitions du deuxième et du troisième formant pour la perception du lieu d'articulation des consonnes occlusives. Les schémas sont faits pour être transformés en son par *Playback*, un des synthétiseurs de parole artificielle des Laboratoires Haskins, New York. Le premier formant a partout une forme appropriée à la synthèse spectrographique de plosives sonores. A. Les transitions du deuxième formant varient. Celles du troisième sont fixes dans une position neutre. B. Les transitions du troisième formant varient. Celles du deuxième sont fixes dans une position intermédiaire entre [b] et [d]. C. Les transitions du troisième formant varient. Celles du deuxième sont fixes dans une position intermédiaire entre [g] et [d]. D. Les transitions des deuxième et troisième formants varient sans cesser de viser leurs Locus dentals respectifs.

nous ne mentionnerons qu'en passant les corrélations entre le domaine acoustique et le domaine articulaire.

Il sera utile de distinguer entre les caractères des transitions qui contribuent à la perception du *lieu* d'articulation d'une consonne (lieu labial, dental, alvéolaire, palatal, vélaire, pharyngal, etc.) et ceux qui contribuent à son *mode* d'articulation (mode occlusif, constrictif, fermé, ouvert, oral, nasal, sourd, sonore, dur, palatalisé, pur, affriqué, simple, complexe, etc.).

Nous considérerons ensemble les transitions des deuxième et troisième formants d'une part, et les transitions du premier formant d'autre part, celui-ci ayant une fonction assez différente des deux autres.

LES TRANSITIONS DES DEUXIÈME ET TROISIÈME FORMANTS

Dans ce tableau d'ensemble du jeu des transitions dans la perception des consonnes, commençons par les faits les plus simples.

La figure 1-A démontre le rôle du 2e formant dans la perception du lieu d'articulation des occlusives sonores. Comme vous le savez, la parole artificielle permet d'isoler à tour de rôle chacun des divers facteurs de la perception d'un phonème. Ici c'est la transition du 2e formant qui est isolée. Les trois formants de la voyelle [e] restent fixes. La transition du premier formant est fixe dans la position qui convient à toutes les occlusives sonores orales. La transition du 3e formant est fixe en une position neutre intermédiaire entre celles de [b], [d] et [g]. Seule la transition du 2e formant change. Elle varie de -5 à +4 en 10 degrés de 120 cycles. Lorsque cette série de 10 schémas spectrographiques passe sous les faisceaux de lumière de la machine à synthétiser la parole, le son change légèrement à chaque nouveau schéma, commençant à [bε], passant par [dε], et aboutissant à [gε]. Pour l'oreille, seul le lieu d'articulation change; il est donc relié à la direction de la transition du 2e formant.

(son)

La figure 1-B démontre le rôle du 3e formant dans la perception du lieu d'articulation. La voyelle est toujours [e]. La transition du premier formant est fixe dans la position des occlusives orales sonores; celle du 2e formant est fixe dans une position neutre intermédiaire entre [b] et [d]. La position du 3e formant, en variant de -4 à +4, fait graduellement passer le son de [b] à [d].

(son)

La figure 1-C montre de la même manière le passage de [g] à [d] sous l'effet des mêmes variations du 3e formant. Cette fois la position fixe du 2e formant est intermédiaire entre [d] et [g].

(son)

Ainsi la perception du lieu d'articulation dépend de la direction du 3e formant aussi bien que de celle du 2e formant.

Si maintenant la voyelle change, on ne peut conserver la perception du même lieu d'articulation qu'en changeant tout au moins la direction du 2e formant. On a donc pour chaque lieu d'articulation autant de formants différents que de voyelles. Mais on s'aperçoit bientôt que toutes les transitions qui font entendre un même lieu d'articulation visent un même point. Dans la figure 1-D toutes les transitions font entendre un [d] et toutes visent un point dont la fréquence est à quelque 1800 cycles et l'éloignement quelque 5cs. Bien que très différentes les unes des autres, ces 8 transitions ont donc un invariant en commun, ce point qu'elles visent et auquel on a donné le nom de Locus dental.

La figure 2 résume la fonction des transitions de 2e formant pour les occlusives orales-sonores dans la perception du lieu d'articulation. Les 2e formants des labiales visent un point de fréquence assez basse, et ceux des dento-alvéolaires un point de fréquence moyenne. Quant aux palato-vélaire, elles fonctionnent différemment selon qu'elles sont suivies de voyelles écartées ou de voyelles arrondies. Devant les voyelles écartées, les 2e transitions des vélaire visent un point de fréquence élevée qui est leur Locus. Mais devant les voyelles arrondies les 2e transitions ne visent

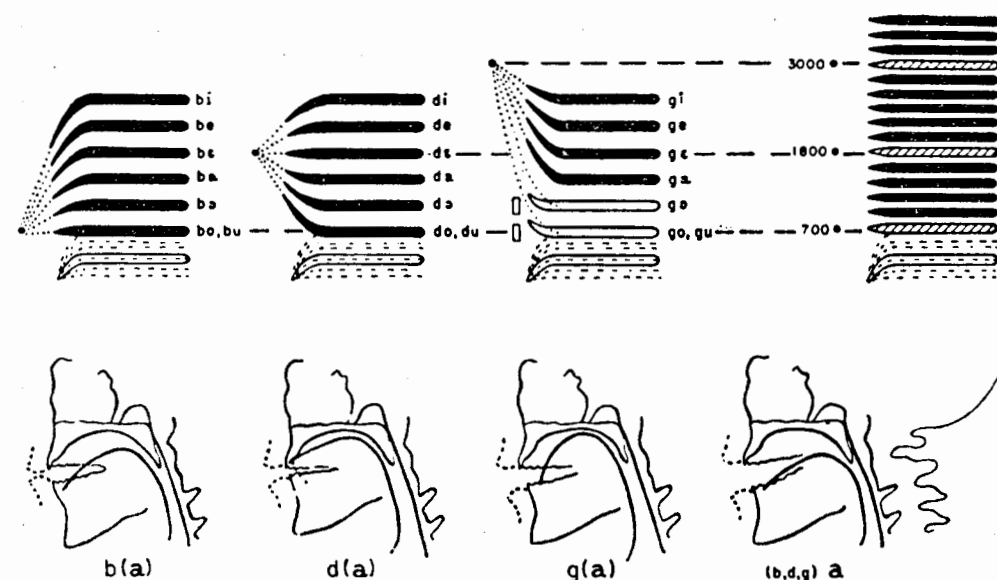


Fig. 2. Les Locus de deuxième formant pour la synthèse des occlusives. On voit, de gauche à droite, la direction des transitions du deuxième formant par rapport au Locus labial, au Locus dental, et au Locus vélaire devant voyelles écartées; devant voyelles arrondies les transitions des vélares ne visent pas leur Locus. A l'extrême droite, la série des transitions zéro (formants droits) par lesquelles les Locus ont pu être tout d'abord déterminés.

plus ce point — elles sont à peu près droites (transition zéro). La perception du lieu d'articulation dépend alors, non plus de la transition de 2e formant mais du bruit d'explosion qui précède cette transition.

Les transitions de 3e formant sont beaucoup plus simples parce que les 3e formants eux-mêmes changent assez peu pour qu'on puisse les considérer comme fixes pour toutes les voyelles sauf [i]. Pour les labiales, les 3e transitions sont fortement négatives, parallèlement aux 2e transitions. Pour les vélares elles sont négatives inversement aux 2e transitions. Pour les dentales, elles visent un point d'une fréquence d'environ 2700 cycles, autrement dit, elles sont positives devant toutes les voyelles sauf [i] et négatives devant [i] parallèlement à la 2e transition.

Il est bon de rappeler ici que dans l'investigation des Locus de 2e transitions, on n'a pas seulement procédé par formants courbés, mais aussi par formants droits (transition zéro) comme indiqué à droite de la figure 3 combinant un premier formant d'occlusive orale avec un 2e formant droit. Une fois déterminée la fréquence du formant droit faisant entendre le plus clairement tel lieu d'articulation, on a dessiné les courbes complètes qui joignaient le Locus ainsi trouvé aux 2e formants de toutes les voyelles possibles et on a procédé par coupes successives pour trouver que le temps qui sépare le plus favorablement le Locus du début de la transition est en moyenne de 5 cs.

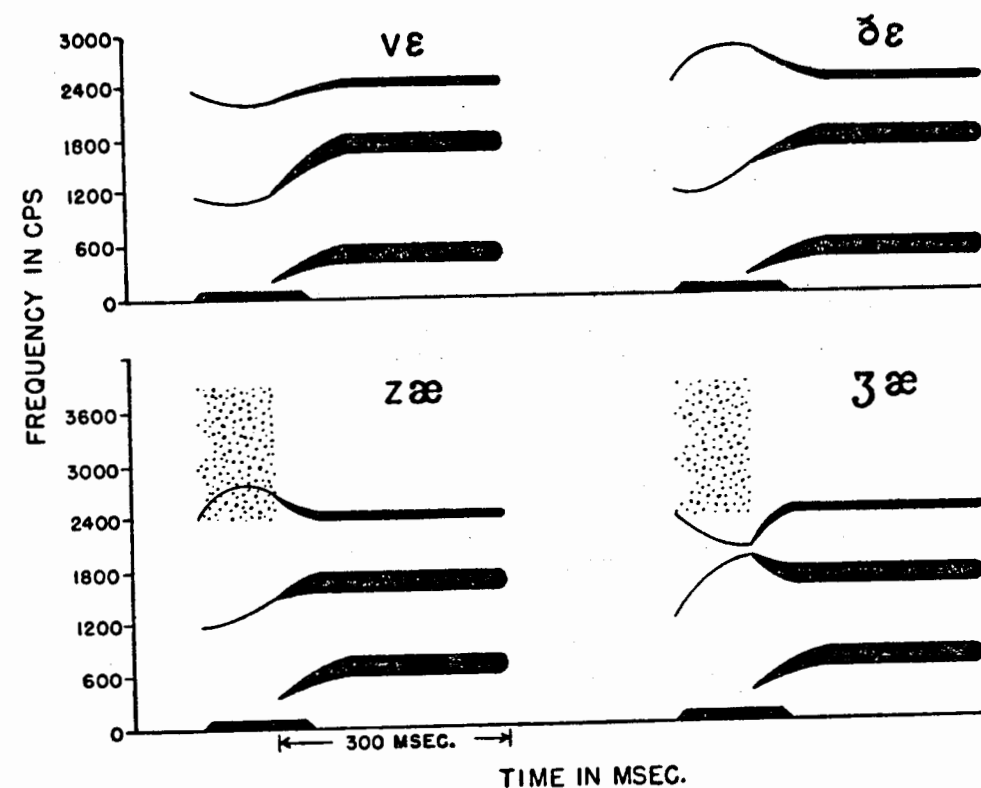


Fig. 3. Schémas typiques pour la synthèse spectrographique des fricatives sonores.

En résumé, c'est donc de la direction des transitions par rapport à un Locus que dépend la perception du lieu d'articulation consonantique.

Les profils articulatoires de la figure 2, pris dans des cinéradiogrammes des syllabes [ba], [da], [ga] permettent d'observer la corrélation entre les variations de fréquence des transitions et les variations de volume et d'ouverture des cavités du pavillon.

Pour les 2e et 3e formants, les transitions des occlusives dites sonores [b], [d], [g], sont à peu de chose près celles des occlusives dites sourdes [p], [t], [k], ou aspirées [ph], [th], [kh], et celles des occlusives nasales [m], [n], [ŋ], sauf que pour les aspirées la première portion de la transition est composée d'ondes turbulentes (bruit) au lieu d'ondes périodiques (son harmonique).

Ce qui a facilité l'étude des transitions pour les occlusives sonores, c'est le fait que, à l'exception des vélares devant voyelles arrondies, leurs lieux d'articulation peuvent se reconnaître, à la synthèse, sans aucun appui du bruit d'explosion. La vitesse des transitions, surtout dans le premier formant, suffit à donner l'impression d'occlusion.

Dans le cas des liquides et des semi-voyelles, la difficulté était encore moindre puisque aucun bruit distinctif n'entre en jeu dans leur production.

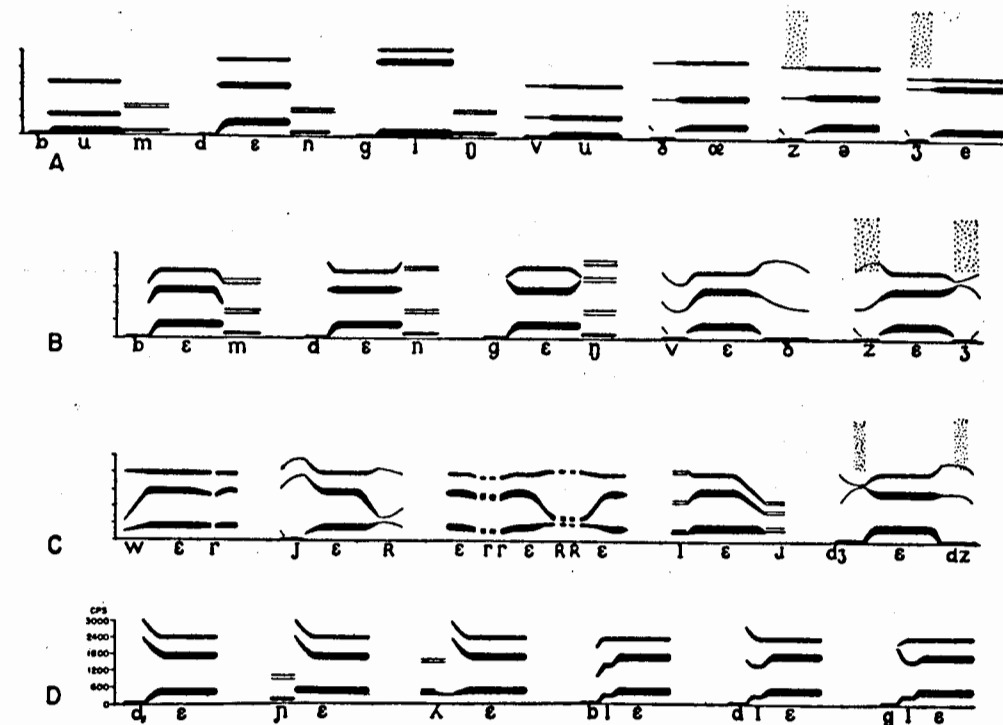


Fig. 4. A. Tableau des Locus de deuxième et troisième formants, pour les principales occlusives et fricatives anglaises, sous la forme de voyelles dont les transitions zéro visent leurs Locus respectifs B, C, D. Schémas typiques pour la synthèse d'occlusives, de fricatives, de semi-voyelles, de liquides, d'affriquées, de palatalisées et de complexes, composés par extension du concept du Locus.

Mais pour les fricatives, une grosse difficulté se présentait du fait que les bandes de friction, leur fréquence, leur largeur, leur intensité, interviennent nettement dans la perception du lieu d'articulation, comme l'a fort bien montré l'étude de Catherine Harris. Cette difficulté a pu être surmontée, tout dernièrement, grâce à l'observation qui suit. En étudiant les facteurs du voisement dans les fricatives, on a discerné l'importance insoupçonnée des formants vocaliques qui traversent la tenue, c'est-à-dire la friction — formants à peine visibles sur les spectrogrammes. On peut les observer sur la figure 3 dans les schémas de synthèse des quatre fricatives sonores anglaises [v], [ð], [z], [ʒ]. Grâce au léger renforcement du lieu d'articulation que permettent ces formants vocaliques de voisement, il a été possible de "neutraliser" le facteur friction de ces quatre fricatives et de faire varier seules les transitions des 2e et 3e formants. Dans les tests, on a pu opposer /v/ à /ð/ sans aucune friction, et /z/ à /ʒ/ avec une friction neutre intermédiaire entre celles des deux fricatives. Dans cette recherche, on a combiné la technique de variations de formants droits (transition zéro) à celle de variations de degrés de transition pour une voyelle fixe. Les résultats ainsi obtenus pour les 2e et 3e transitions de fricatives sonores semblent

bien s'appliquer aussi aux mêmes transitions des fricatives sourdes correspondantes [f], [θ], [s], [ʃ].

Les Locus des 2e et 3e formants pour les 6 occlusives [b d g m n ŋ] et pour les 4 fricatives [v ð z ʒ] sont présentés dans la figure 4-A sous la forme de voyelles dont les formants les viseraient. Les premiers formants sont choisis arbitrairement pour s'accorder vocaliquement avec les 2e. Ainsi les voyelles des syllabes [bum ðen gin vu ðæ zə ʒe] peuvent servir d'aide-mémoire approximatif aux Locus des 2e formants de ces 10 consonnes — 20 consonnes si nous comptons les sourdes et aspirées correspondantes. En écoutant le son que produisent ces schémas en passant sous les faisceaux de lumière de la machine à synthèse, rappelons-nous qu'ils ne peuvent que vaguement suggérer le son des consonnes correspondantes, car dans la parole humaine quand l'un des formants est droit, les autres sont normalement courbés.

(son)

Si nous appliquons à la voyelle [e] le principe des Locus de 4-A, nous obtenons les transitions de 4-B qui produisent des syllabes plus intelligibles que les précédentes: [bem ðen geŋ veð zεʒ].

(son)

Enfin si nous appliquons des principes semblables aux semi-voyelles, aux liquides, aux affriquées, aux palatalisées et aux complexes consonantiques, nous obtenons les formes schématiques 5-C et 5-D et les sons qui correspondent [wer jeR lεɹ dʒedz d,e ne λe ble dle gle].

(son)

Les semi-voyelles ont naturellement leurs Locus près des formants des voyelles fermées qui leur correspondent.

Les liquides ont des Locus bas pour le 2e formant. Les plus bas sont ceux de r pharyngal et l sombre, les moins bas ceux de r apico-alvéolaire et l clair. La transition du 3e formant n'est caractéristique que dans l'r palatal, rétroflexe ou dorsal, pour lequel elle descend très bas.

Il va sans dire que les transitions du [x] et du [ç] sont les mêmes que celles du [R] et du [j].

Les affriquées ont les transitions du second élément, l'élément fricatif.

Les palatalisées ont les 2e et 3e transitions du [j] mais se distinguent entre elles par le mode d'articulation qui est dans le premier formant. Cela confirme la définition classique qui veut que les palatalisées gardent leur mode d'articulation original et convertissent leur lieu d'articulation en celui du *jod*.

Quant aux complexes *muta plus liquida* on peut les synthétiser en traitant la tenue liquide comme une voyelle et en appliquant à cette voyelle le concept du Locus.

Tout ce qui précède se rapporte à la perception des lieux d'articulation. Les 2e et 3e formants ont en effet pour fonction essentielle de distinguer le lieu d'articulation au moyen de la direction des transitions. Ces formants jouent pourtant un certain rôle dans la perception des *modes* d'articulation, et cela par le régime des transitions,

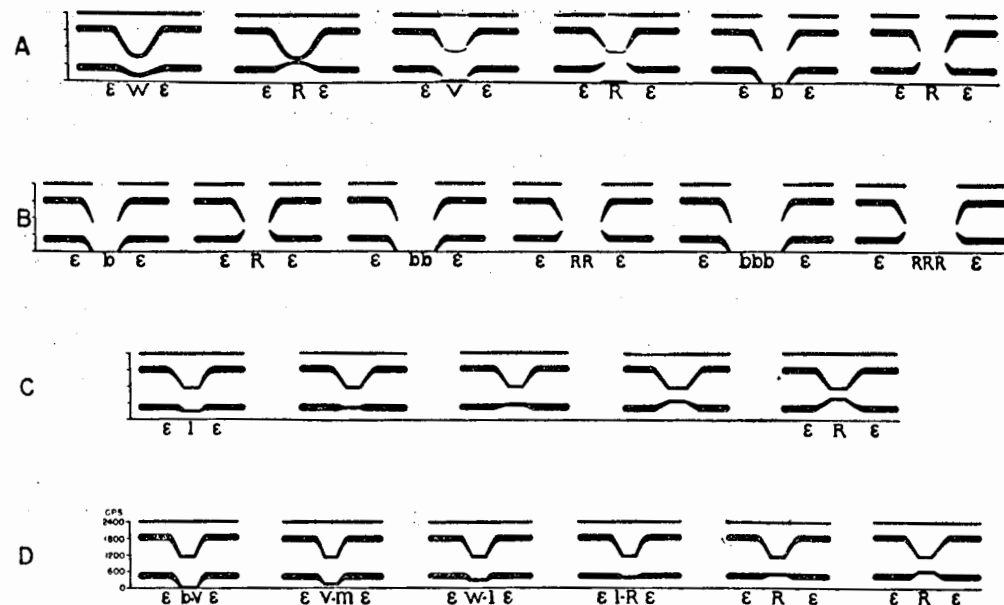


Fig. 5. A, B, C. Effet des transitions de premier formant pour la perception du *lieu* d'articulation. Dans chacune des trois séries respectives A, B, C, les deuxième et troisième formants sont fixes; seules les transitions des premiers formants varient. Devant [e], les transitions de premier formant qui sont négatives font entendre une consonne buccale, celles qui sont positives une consonne pharyngale. D. Effet des transitions de premier formant pour la perception du *mode* d'articulation. Plus le Locus du premier formant est haut, plus la consonne est vocalique.

c'est-à-dire le degré de vitesse avec lequel elles changent de fréquence. (Ce facteur joue d'ailleurs pour toutes les transitions, la première aussi bien que les 2e et 3e.) D'une manière générale, le régime est relié à la vocalisation de la consonne. Ainsi les consonnes sonores ont en moyenne des transitions plus lentes que les sourdes; les fricatives ont en moyenne des transitions plus lentes que les occlusives; et les semi-voyelles et les liquides ont en moyenne des transitions plus lentes que les fricatives. Dans la série vocalisante: [pe be ve we ue], par exemple, les transitions sont de moins en moins rapides.

D'ailleurs le régime des transitions a aussi quelque rapport avec le lieu d'articulation. Ainsi parmi les occlusives, les labiales ont des transitions plus rapides que les dentales devant voyelles antérieures, ces dernières que les vélaires, et ces dernières que les dentales devant voyelles postérieures.

LES TRANSITIONS DU PREMIER FORMANT

La fonction des transitions du premier formant est partagée: le premier formant permet de distinguer le lieu aussi bien que le mode d'articulation. La distinction de lieu est la plus claire mais la distinction de mode est la plus importante du point de vue statistique.

Examinons d'abord la relation du premier formant avec le lieu d'articulation.

Lorsque la constriction consonantique est dans la moitié antérieure du pavillon (des lèvres au palais mou) la transition du premier formant est négative par rapport à la voyelle neutre (autrement dit elle commence initialement plus bas que 500 cycles environ), et lorsque la constriction consonantique est dans la moitié postérieure du pavillon (dans le pharynx) la transition du premier formant est positive par rapport à la voyelle neutre (autrement dit elle commence plus haut que 500 cycles environ). La figure 5-A, B, C, démontre cela d'une manière dramatique grâce à la coïncidence qui veut que les labiales, ou *l* sombre, aient à peu près le même Locus de 2e formant (Locus très bas) que les pharyngales. Pour ces trois classes de consonnes en effet la cavité buccale est grande au départ. La distinction buccale/pharyngale est donc entièrement produite ici par la transition du premier formant.

(son)

Pour comprendre ce que nous venons d'observer, rappelons que la voyelle neutre est celle pour laquelle le pavillon n'a pas de constriction bien marquée: le pavillon prend alors théoriquement l'aspect d'un tube, uniforme de diamètre, fermé à un bout (la glotte) et ouvert à l'autre (les lèvres), et qui résonne au quart d'onde, aux trois-quarts d'onde, aux cinq-quarts d'onde, etc., indépendamment de son diamètre. Quand la longueur du tractus vocal en tube uniforme est de quelque 17.5 cm, ses trois premiers modes de résonance, c'est-à-dire ses trois premiers formants sont à environ 500 cycles, 1500 cycles et 2500 cycles, dans un rapport proche de 1-3-5. Le timbre de la voyelle produite par ce tube uniforme est assimilable à un [ə]; elle est donc doublement qualifiée pour le titre de "voyelle neutre."

Chiba et Kajiyama ont démontré théoriquement qu'on pouvait considérer toutes les autres voyelles comme des modifications de cette voyelle neutre par application du concept du lieu d'articulation qui veut que si la constriction est près d'un ventre la fréquence du formant baisse par rapport à la voyelle neutre, et inversement si la constriction est près d'un noeud la fréquence du formant monte par rapport à la voyelle neutre. Cette hypothèse des Japonais a d'ailleurs été clairement confirmée par Gunnar Fant et Kenneth Stevens pour les voyelles. Ce que nous venons d'exposer pour l'R pharyngal indiquerait que le concept des Japonais peut s'appliquer non seulement aux voyelles mais aussi aux consonnes. En effet, si le Locus du premier formant de R pharyngal est plus haut que le premier formant de la voyelle neutre, c'est que la constriction du R pharyngal est plus près d'un noeud que d'un ventre, le ventre du quart d'onde étant aux lèvres et le noeud à la glotte. Si par contre la plupart des consonnes ont un Locus de premier formant plus bas que 500 cycles, c'est que la plupart des consonnes ont leur constriction dans la portion antérieure du pavillon, c'est-à-dire plus près d'un ventre que d'un noeud.

Ainsi en ce qui concerne la perception du lieu d'articulation, les transitions du premier formant servent à distinguer les consonnes pharyngales des consonnes buccales.

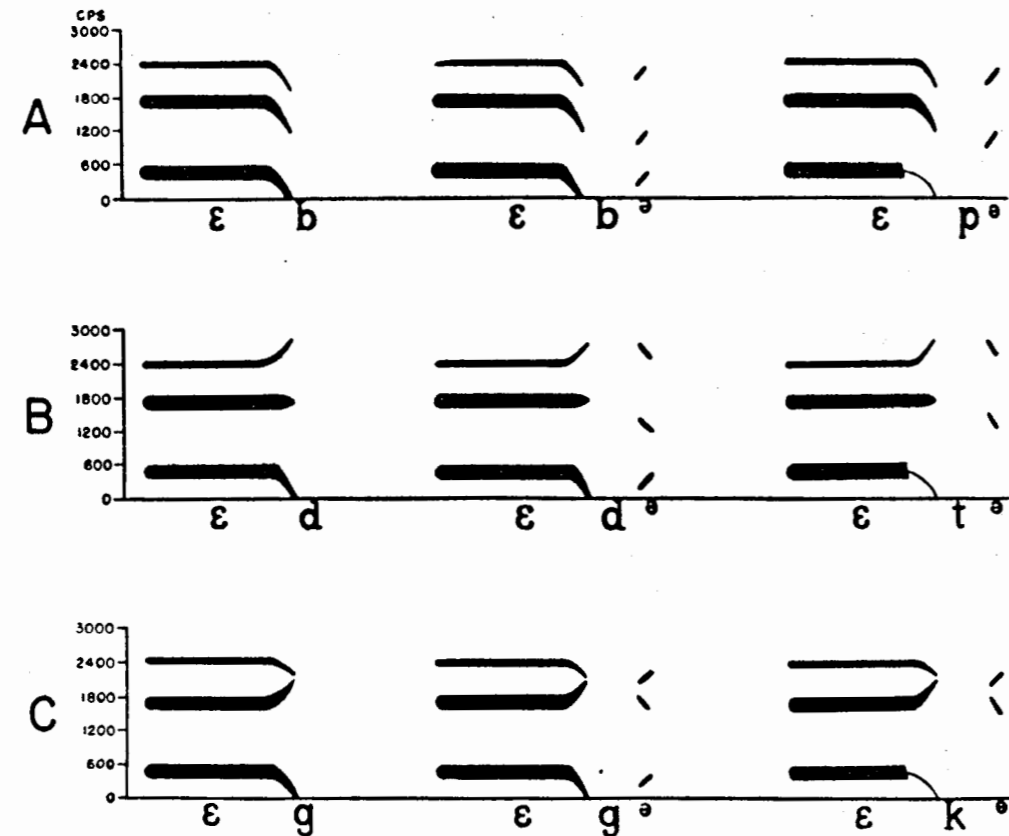


Fig. 6. A, B, C. Effet de l'intensité dans la transition du premier formant sur la perception du voisement. Dans chaque série, le troisième schéma doit son dévoisement à la faible intensité de la transition du premier formant, tant dans l'implosion que dans la détente.

En ce qui concerne la perception du mode d'articulation des consonnes buccales, la vocalisation consonantique est reliée à deux indices acoustiques. L'un, le régime de transition, fonctionne, comme nous l'avons déjà dit, pour les trois formants: plus les transitions sont lentes, plus la consonne est vocalisée. L'autre indice ne fonctionne que pour le premier formant: en général, et jusqu'à une limite d'environ 500 cycles, plus le Locus du premier formant est élevé, plus la consonne est vocalisée. Les schémas de la Figure 5-D démontrent à la fois la fonction du premier formant comme indice de mode et comme indice de lieu. Le schéma de base combine des indices d'occlusive et de fricative. Ainsi le plus bas des Locus de premier formant produit un son intermédiaire entre [b] et [v], le 2e, à 240 cycles, un son intermédiaire entre [v] et [m], le 3e, à 360 cycles, un son intermédiaire entre [w] et [l], le 4e, à 480 cycles, un son intermédiaire entre [l] et [R]. A partir d'ici on est au-dessus de 500 cycles, le lieu d'articulation change donc de buccal à pharyngal, et les sons qui suivent sont de plus en plus pharyngaux à mesure que le Locus du premier formant s'élève.

(son)

Tout ce qui a été dit jusqu'ici pourrait faire croire que les indices des transitions ne sont fonction que de la fréquence et de la durée. Il existe au moins un cas où la troisième dimension, l'intensité, semble jouer un rôle dans les transitions. C'est dans le dévoisement: plus la transition du premier formant est faible, plus la consonne est sourde. La Figure 6 en donne un bon exemple. La presque absence de première transition, tant dans les implosions que dans les détentes, se traduit par un dévoisement bien marqué.

(son)

Cette Figure montre aussi l'importance du facteur *durée*: pour que les transitions des détentes fonctionnent comme telles, il leur faut une durée d'au moins 3 cs. En dessous de cela, elles fonctionnent non comme des transitions mais comme des bruits d'explosion, ce qui indique que la direction des transitions ne joue plus dans la perception.

Pour terminer nous allons entendre un poème de Guillaume Apollinaire intitulé *L'anémone et l'ancolie*, synthétisé d'après les notions qui précèdent.

L'anémone et l'ancolie
 Ont poussé dans le jardin
 Où dort la mélancolie
 Entre l'amour et le dédain
 (son)

Laboratoire Haskins
 Université du Colorado