

# Phonétique acoustique

C. Meunier - Chargée de Recherche

Laboratoire Parole et Langage, CNRS UMR 6057  
Université de Provence  
29 avenue Robert Schuman  
13621 Aix-en-Provence Cedex 1, France.

Courriel : christine.meunier@pl.univ-aix.fr

## I. Les sons du langage

L'ensemble des configurations articulatoires possibles pour la production du langage parlé n'est pas extensible. Elles sont dépendantes du matériel anatomique dont dispose l'être humain. Et comme tous les êtres humains disposent du même appareil anatomique, les sons utilisés dans chacune des langues du monde relèvent de configurations articulatoires communes. Il est donc possible de procéder à des regroupements de classes de sons dans lesquels chaque langue puise son propre inventaire. Ces classes de sons se font selon des critères articulatoires ou acoustiques.

On observe sur la figure 1 les critères articulatoires retenus selon l'Alphabet Phonétique International (International Phonetic Alphabet, IPA) pour le classement des consonnes. Ainsi, dans la classe des

fricatives, le français utilise seulement une partie des possibilités offertes par l'appareil anatomique. On remarquera sur cette figure 1 que le français utilise peu la partie très postérieure du tractus vocal, la consonne la plus arrière étant le /R/ uvulaire.

De même, dans l'inventaire des voyelles françaises il n'y a aucune voyelle centrale, excepté le /ə/ (Fig. 2). En fait, la structure de l'inventaire des sons dans une langue n'est pas aléatoire mais suit une logique universelle relative aux contraintes articulatoires et à la distinctivité nécessaire au sein de chaque système (Schwartz *et al.*, 1997). De même, la description acoustique des sons du langage tente de présenter un panorama de valeurs permettant de distinguer l'ensemble des sons. Nous tenterons, dans ce chapitre, de dresser un inventaire des descriptions acoustiques des sons du français en les reliant avec une explication articulatoire.

ALPHABET PHONÉTIQUE INTERNATIONAL - CONSONNES

	bilabiales	labiodentales	dentales	alvéolaires	post-alvéolaires	rétroflexes	palatales	vélaires	uvulaires	pharyngales	glottales
Occlusives	<b>p</b> <b>b</b>			<b>t</b> <b>d</b>		<b>t̠</b> <b>ɖ</b>	<b>c</b> <b>ɟ</b>	<b>k</b> <b>g</b>	<b>q</b> <b>ɢ</b>		<b>ʔ</b>
Nasales	<b>m</b>	<b>ɱ</b>		<b>n</b>		<b>ɳ</b>	<b>ɲ</b>	<b>ŋ</b>	<b>ɴ</b>		
Vibrantes	<b>ʙ</b>			<b>r</b>					<b>ʀ</b>		
Battues		<b>ɸ</b>		<b>ɾ</b>		<b>ɽ</b>					
Fricatives	<b>ɸ</b> <b>β</b>	<b>f</b> <b>v</b>	<b>θ</b> <b>ð</b>	<b>s</b> <b>z</b>	<b>ʃ</b> <b>ʒ</b>	<b>ʂ</b> <b>ʐ</b>	<b>ç</b> <b>ʝ</b>	<b>x</b> <b>χ</b>	<b>ɣ</b> <b>ʁ</b>	<b>ħ</b> <b>ʕ</b>	<b>h</b> <b>ɦ</b>
Fricatives latérales				<b>ɬ</b> <b>ɮ</b>							
Approximantes		<b>ʋ</b>		<b>ɹ</b>		<b>ɻ</b>	<b>ɻ</b>	<b>ɰ</b>			
Approximantes latérales				<b>ɭ</b>		<b>ɭ</b>	<b>ʎ</b>	<b>ʎ</b>			

Fig. 1. Liste des principales consonnes des langues du monde dans l'Alphabet Phonétique International classées selon leur mode d'articulation (en colonne) et leur lieu d'articulation (en ligne). Lorsqu'il y a deux consonnes par case, celle de gauche est la consonne sourde, celle de droite est son équivalente sonore. D'après International Phonetic Alphabet (International Phonetic Association, 1999). Les consonnes du français sont entourées en gris.

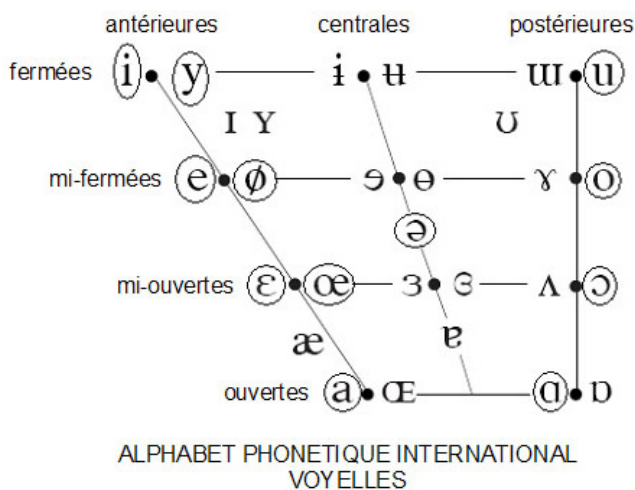


Fig. 2. Trapèze vocalique représentant un classement des voyelles orales des langues du monde selon des critères articulatoires. Lorsque deux voyelles apparaissent de part et d'autre d'un point, la première est une réalisation étirée, la deuxième une réalisation arrondie (voir plus loin dans ce chapitre II.1 Les voyelles). D'après International Phonetic Alphabet (International Phonetic Association, 1999). Les voyelles du français sont entourées en gris.

## II. L'acoustique des sons

Il est classique de distinguer les voyelles des consonnes selon des critères articulatoires : la production des voyelles suppose une libre circulation de l'air dans le tractus vocal, tandis que lors de la production des consonnes, un rétrécissement du tractus en un point entraîne une interruption ou une perturbation dans la circulation de l'air. Cette distinction de bas niveau reflète mal ce qui différencie fondamentalement voyelles et consonnes. En effet, en s'appuyant uniquement sur des critères articulatoires ou acoustiques, la différence entre une occlusive et une approximante est sans doute plus importante que la différence entre cette même approximante et une voyelle. Il apparaît en fait que la distinction la plus probante entre les voyelles et les consonnes est de nature linguistique : les voyelles et les consonnes se distinguent selon leur place dans la syllabe, les voyelles peuvent être le noyau (centre de la syllabe et seul élément indispensable pour une syllabe), ce que les consonnes, à de rares exceptions près, ne peuvent pas (cf. chapitre 14).

Si les voyelles représentent un ensemble de productions plutôt homogène, il n'en va pas de même pour la catégorie des consonnes constituées de classes de sons assez différents aussi bien d'un point de vue articulatoire qu'acoustique. On parle alors souvent de macroclasses de sons (Rossi, 1990) pour essayer

de rendre compte de groupes de sons homogènes du point de vue acoustique. Si la classe des voyelles est homogène, il est classique de distinguer les voyelles fermées des voyelles ouvertes et les voyelles antérieures des postérieures. Au sein des consonnes, deux macroclasses sont elles-mêmes assez homogènes, il s'agit des occlusives et des fricatives. Reste ensuite un ensemble de sons, ayant peu de points communs, faisant office d'intermédiaire entre consonnes et voyelles que l'on peut regrouper dans la macro classe des consonnes vocaliques. Dans cette macro classe, on trouve les approximantes, les glissantes, les vibrantes, etc.

Nous allons voir ci-dessous plus en détails les caractéristiques de l'ensemble de ces sons. Pour une description générale de l'acoustique des sons, cf. chapitres 6 et 8.

### 1. Les voyelles

Toutes les voyelles sont voisées, c'est-à-dire qu'il s'agit de sons périodiques. Cette périodicité est issue de la vibration des cordes vocales. Les voyelles sont caractérisées par la présence de zones d'harmoniques renforcées appelées « formants ». La configuration des cavités bucco-pharyngales étant différente pour chaque voyelle, chacune d'elle va être caractérisée par des valeurs de formants différentes. En d'autres termes, la modulation de la forme du tractus et donc de la taille des cavités, va occasionner des valeurs spécifiques de formants correspondant à la forme que prend le tractus vocal pour chaque voyelle. Il existe un lien entre la taille d'un résonateur (cavités du tractus dans la parole) et sa fréquence : plus la taille du résonateur est importante, plus sa fréquence est basse. Les valeurs de formants vont donc, en partie, dépendre de la taille des cavités de résonance durant la production des voyelles.

Pour les voyelles, trois cavités sont pertinentes : les cavités pharyngale, buccale et nasale. Quatre dimensions permettent de modifier la forme ou l'accès à ces cavités : 1/ le degré d'ouverture de la mandibule ; 2/ la position de la langue ; 3/ la position des lèvres ; 4/ la position du velum (autorisant ou non le passage de l'air dans les fosses nasales). Les variations au sein de ces quatre dimensions entraînent des variations de timbre (valeurs de formants). Dans d'autres langues, la durée ou encore la valeur tonale (variations de F0) vont constituer des dimensions supplémentaires permettant la distinctivité au sein du système.

Le premier formant (F1) est déterminé par le degré d'ouverture de la mandibule (et de la hauteur de la langue). Le deuxième formant (F2) varie en fonction

de la position avant ou arrière de la langue, mais aussi en fonction de la position étirée ou arrondie des lèvres. Enfin, les valeurs du troisième formant (F3) sont déterminées par la position étirée ou arrondie des lèvres. Notons que la voyelle /ə/, présente en français mais avec un statut phonologique très particulier, présente des caractéristiques articulatoires totalement neutres et ne sera donc pas mentionnée dans nos descriptions.

Tableau 1. Valeurs formantiques moyennes des voyelles orales du français (d'après Tubach, 1989).

		F1	F2	F3
voy. fermées	i	308	2064	2976
	y	300	1750	2120
	u	315	764	2027
voy. mi-fermées	e	365	1961	2644
	ø	381	1417	2235
	o	383	793	2283
voy. mi-ouvertes	ɛ̃	530	1718	2558
	œ	517	1391	2379
	ɔ	531	998	2399
voy. ouverte	a	684	1256	2503

### a. L'aperture

L'aperture de la mandibule ne fonctionne pas de façon isolée. Elle entraîne, de fait, un abaissement de la langue ainsi qu'un rétrécissement de la taille du pharynx. On oppose ainsi les voyelles ouvertes (ou basses), caractérisées par une aperture maximale de la mandibule et un abaissement de la langue (entraînant une très petite cavité pharyngale), aux voyelles fermées (ou hautes), caractérisées par une aperture très réduite de la mandibule et la langue en position haute (laissant place à une grande cavité pharyngale). Ainsi, les voyelles ouvertes sont caractérisées par un F1 élevé (/a/ = 700 Hz), tandis que les voyelles fermées ont un F1 bas (/i/ = 300 Hz) (Tableau 1 et Fig. 3). Les voyelles ouvertes sont /a/ et /ɑ/, /ã/. Les voyelles mi-ouvertes sont /ɛ/, /œ/, /ɔ/, /ɛ̃/, /œ̃/, /ɔ̃/. Les voyelles mi-fermées sont /e/, /ø/, /o/. Les voyelles fermées sont /i/, /y/, /u/.

### b. La position de la langue

Le mouvement de la langue en position avant ou arrière entraîne une modification globale du tractus vocal et plus précisément de la taille de la cavité buccale. Ce mouvement a une incidence sur la valeur du F2 (Fig. 3). Les voyelles antérieures sont caractérisées par un F2 élevé (/i/ = 2000 Hz), tandis que les voyelles postérieures laissent place à une cavité buccale plus ample et ont un F2 bas (/u/ = 750 Hz). Les voyelles d'avant sont /i/, /y/, /e/, /ø/, /ɛ/, /œ/, /a/, /ɛ̃/, /œ̃/. Les voyelles d'arrière sont /u/, /o/, /ɔ/, /ɑ/, /ɔ̃/, /ã/.

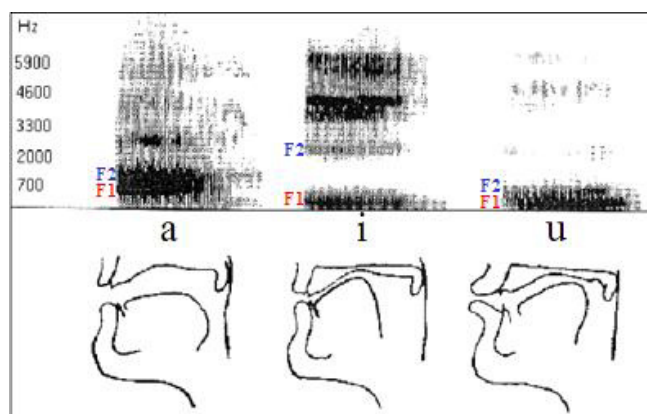


Fig. 3. Valeurs formantiques des voyelles a, i, u (en haut) et coupe sagittale du tractus vocal (en bas) avec les différentes positions de la langue pour la production de ces trois voyelles, les plus distinctes du point de vue articulatoire.

### c. La position des lèvres

Certaines voyelles sont réalisées avec une projection des lèvres en avant. On parle alors « d'arrondissement » des lèvres, de « projection des lèvres » ou encore de « protrusion ». Ce geste articulatoire a pour conséquence l'allongement de la cavité buccale par la création d'une petite cavité, la cavité labiale. Cet allongement occasionnel du conduit vocal entraîne un abaissement des F2 et F3. Ainsi, la seule différence articulatoire entre /i/ et /y/, qui sont toutes les deux des voyelles fermées antérieures, est l'arrondissement des lèvres pour la production de /y/. Le F3 de /y/ (2100 Hz) est alors plus bas que celui de /i/ (3000 Hz). Il en va de même pour le F2 (/y/ = 1750 Hz ; /i/ = 2000 Hz). Les voyelles arrondies sont /y/, /u/, /ø/, /o/, /œ/, /ɔ/, /ɔ̃/, /œ̃/. Les voyelles non-arrondies (étirées) sont /i/, /e/, /ɛ/, /a/, /ɑ/, /ã/, /ɛ̃/.

### d. La nasalité

Les voyelles nasales présentent des caractéristiques plus complexes concernant leur description acoustique. Lors de la production des voyelles nasales, le voile du palais s'abaisse laissant le flux d'air accéder aux fosses nasales. Les fosses nasales entrent donc en communication avec le conduit oral. Le couplage acoustique de ces cavités donne lieu à un formant supplémentaire, nommé formant nasal (vers 500 Hz – 700 Hz) ainsi qu'à des « anti-formants », autrement dit des zones où l'on observe une forte chute de l'intensité des harmoniques. Par ailleurs une modification des valeurs formantiques des voyelles orales correspondantes est observée. Les voyelles nasales sont /ã/, /ɔ̃/, /ɛ̃/, /œ̃/. Les voyelles nonarrondies (étirées) sont /i/, /y/, /u/, /e/, /ø/, /o/, /ɛ/, /œ/, /ɔ/, /a/, /ɑ/.

## 2. Les consonnes

Les consonnes peuvent être périodiques ou apériodiques. Par rapport aux voyelles, elles sont produites avec un resserrement du conduit vocal. Les consonnes constituent un ensemble de sons très hétérogène aussi bien du point de vue articuloire qu'acoustique. Le type de resserrement réalisé va permettre de distinguer les macroclasses de consonnes. D'une certaine façon les deux macroclasses occlusives et fricatives sont des sous-ensembles homogènes et organisés. Nous regroupons les autres types de consonnes dans la catégorie des consonnes vocaliques en raison de la présence d'une structure de formant dans leur réalisation acoustique. Toutefois, cette catégorie est constituée de sons très différents d'un point de vue articuloire.

En acoustique, trois grandes classes sont donc distinguées : 1/ pour les occlusives (ou plosives), le conduit phonatoire est fermée durant une partie de l'émission de la consonnes ; 2/ pour les fricatives (ou constrictives), le resserrement est important et donne lieu à un bruit turbulent ; 3/ les consonnes vocaliques présentent un rétrécissement ponctuel et/ou peu important du conduit vocal. Ce faible rétrécissement permet la présence d'une structure formantique visible pour ces consonnes.

La périodicité ou l'apériodicité des sons du langage fait référence, respectivement, à la vibration ou l'absence de vibration des cordes vocales. On trouve également les termes *voisé/non voisé* ou *sonore/sourd*. On lit parfois dans des ouvrages que les consonnes sont des sons apériodiques. Toutefois, à partir du moment où une source sonore est présente, on observe, sur le signal de parole, une onde périodique à laquelle peut être ajoutée une source bruitée. Nous considérons que ces signaux sont périodiques.

### a. Les occlusives

On distingue généralement les occlusives orales des occlusives nasales. Effectivement, d'un point de vue articuloire les occlusives se caractérisent par la fermeture complète du conduit oral. Toutefois pour les nasales le voile du palais est abaissé. L'occlusion devient alors relative puisque l'air passe par les cavités nasales qui entrent en résonance donnant lieu à une structure de formant. Sur le simple critère de l'analyse acoustique, les occlusives nasales devraient être classées parmi les consonnes vocaliques (Meunier, 1994). Pour cette raison, nous décrivons les nasales plus bas (3. Les consonnes vocaliques).

Les occlusives orales sont constituées de deux événements temporels. D'abord une tenue, durée pendant laquelle l'occlusion est maintenue, puis un relâchement de l'occlusion. Ces deux événements donnent lieu à une trace acoustique différente selon que l'occlusive est périodique ou apériodique. Pour ces dernières, le conduit vocal étant totalement fermé, ces consonnes sont caractérisées par un silence durant toute la durée de l'occlusion. A la fin de l'occlusion, le contact de l'occlusion (lèvres, palais ou autre) est brutalement relâché donnant lieu à un bruit d'explosion dont la fréquence est caractéristique de la consonne. Il en va de même pour les occlusives périodiques à la différence près que la vibration des cordes vocales donne lieu à une onde sonore (remplaçant le silence des occlusives apériodiques et visible sur un spectrogramme sous la forme d'une « barre de voisement ») dont la fréquence est celle de la fréquence fondamentale (F0) (Fig. 4). Les caractéristiques de la barre de voisement ont une grande importance d'un point de vue perceptif. Le Voice Onset Time (V.O.T.) est le temps d'établissement du voisement, mesuré à partir de la barre d'explosion (Serniclaes, 1987). En français, les occlusives sourdes ont un V.O.T. positif (le voisement s'installe après l'explosion), alors que les sonores ont un V.O.T. négatif (l'établissement du voisement précède l'explosion). Ce n'est pas le cas dans toutes les langues (Lisker et Abramson, 1964). Les occlusives périodiques

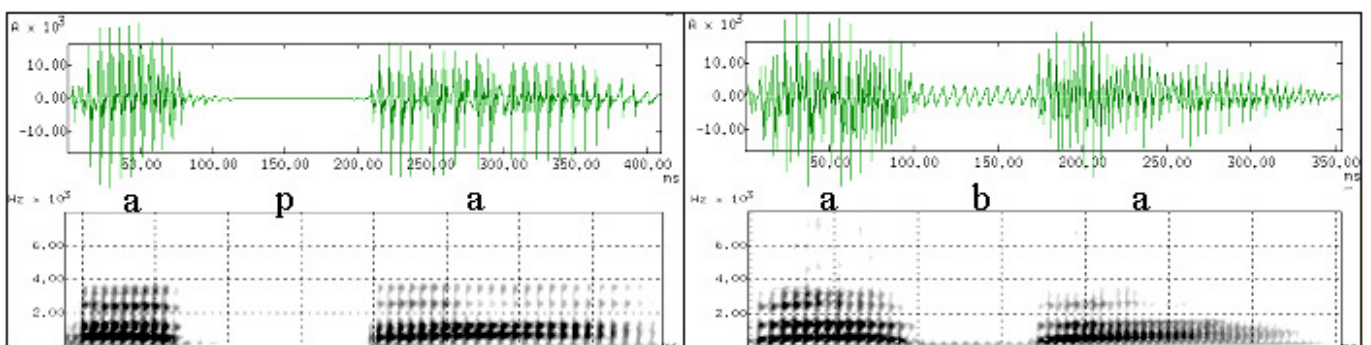


Fig. 4. Signaux temporel (en haut) et spectrographique (en bas) des séquences /apa/ (à gauche) et /aba/ (à droite). On observe la barre de voisement permettant de distinguer le /b/ du /p/.

sont généralement plus courtes que les occlusives apériodiques. Leur bruit d'explosion est également de moindre intensité que celui de leur équivalente sourde. Les occlusives orales apériodiques sont /p/, /t/, /k/. Les occlusives orales périodiques sont /b/, /d/, /g/.

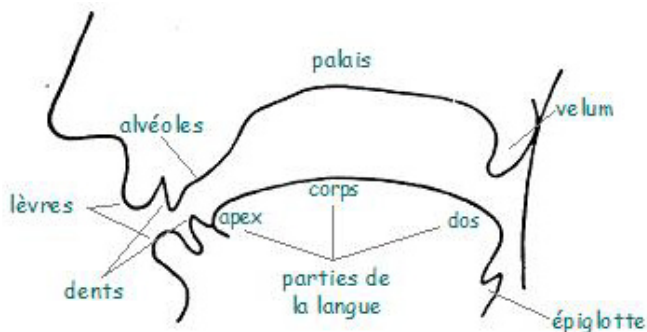


Fig. 5. Coupe sagittale du tractus vocal et localisation des lieux d'articulation et articulateurs utilisés pour la description des sons des langues du monde.

Les occlusives se distinguent également en fonction de leur lieu d'articulation, autrement dit du lieu où est effectué la fermeture du conduit vocal (Fig. 5). Il existe trois lieux d'articulation pour les occlusives en français : les bilabiales (les deux lèvres viennent en contact), les alvéolaires (l'apex de la langue vient obturer le conduit oral en se plaçant sur les alvéoles, c'est-à-dire, juste derrière les incisives centrales supérieures), les vélares (le dos de la langue vient en contact du velum, c'est-à-dire la partie molle et postérieure du palais). Les caractéristiques acoustiques qui permettent de différencier ces trois types d'occlusives portent a priori sur une durée très brève (5 à 30 ms), le bruit d'explosion. En effet, la tenue est constituée soit d'un silence, soit d'une onde non caractéristique du lieu d'articulation de la consonne produite. La fréquence, la durée et l'énergie du bruit d'explosion sont typiques de la consonne produite. Les bilabiales ont un bruit d'explosion de très faible énergie (parfois même non identifiable), court et de basse fréquence. Les alvéolaires produisent

un bruit intense, court et plutôt de haute fréquence. Enfin, les vélares ont un bruit intense, long et dont la fréquence dépend étroitement de la voyelle suivante. En effet, la fréquence du bruit dépend de la taille de la cavité antérieure au lieu d'occlusion et pour les vélares, le lieu d'occlusion va légèrement se déplacer selon que la voyelle suivante est une voyelle antérieure (/i/) ou postérieure (/u/), ceci en raison de la coarticulation (cf. plus loin dans ce chapitre). Toutefois, on considère généralement que les traces acoustiques les plus saillantes pour l'identification du lieu d'articulation des occlusives ne sont pas internes au segment consonantique, mais dans les « pentes de transition » des voyelles (cf. chapitre 8, Fig. 8). Les transitions sont les variations fréquentielles (sous forme de pente) des formants de la voyelle au contact de la consonne. Le type de pente (montant, descendant ou plat) va dépendre du lieu d'articulation de l'occlusive ainsi que du type de voyelle produit. Les occlusives bilabiales sont /p/ et /b/. Les occlusives alvéolaires sont /t/ et /d/. Les occlusives vélares sont /k/ et /g/.

### b. Les fricatives

Les consonnes fricatives sont caractérisées par la présence d'un bruit turbulent durant toute leur tenue. Ce bruit est issu d'un resserrement important, mais incomplet du tractus vocal. Les fricatives peuvent être apériodiques ou périodiques. Dans ce deuxième cas, le bruit se superpose à une onde sinusoïdale caractéristique de la fréquence fondamentale (Fig. 6). Tout comme les occlusives, les fricatives sonores sont plus brèves que les fricatives sourdes. De même, on note une intensité globale du bruit plus faible pour les sonores en raison de la vibration des cordes vocales qui diminue la pression supraglottique. Les fricatives apériodiques sont /f/, /s/, /ʃ/. Les fricatives périodiques sont /v/, /z/, /ʒ/.

La fréquence et l'intensité du bruit de friction dépendent du lieu d'articulation de la consonne,

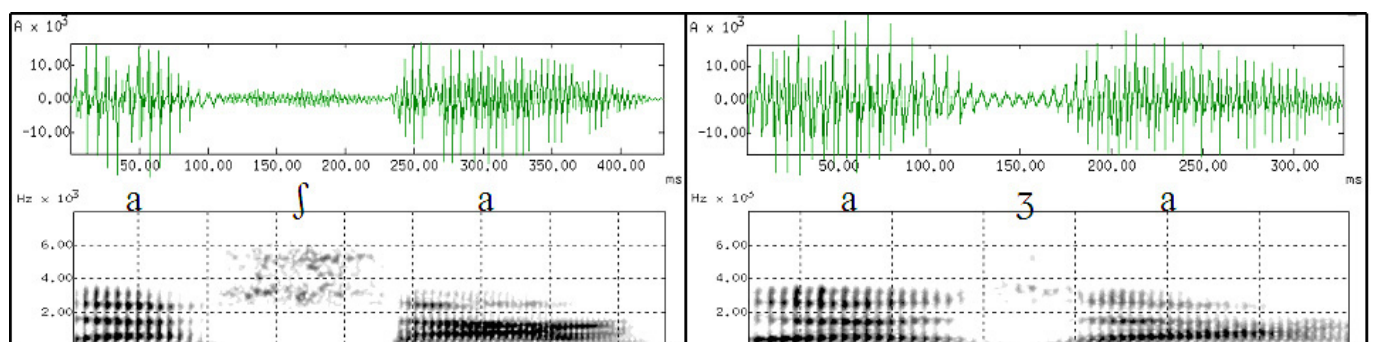


Fig. 6. Signaux temporel (en haut) et spectrographique (en bas) des séquences /aʃa/ (à gauche) et /aʒa/ (à droite). On constate le voisement bruité de la fricative /ʒ/. A noter également la durée plus courte et l'intensité plus faible de la fricative sonore.

c'est-à-dire de l'endroit du conduit où le passage de l'air est le plus étroit. Il existe trois lieux d'articulation en français pour les fricatives : l'espace entre la lèvre inférieure et les incisives supérieures (labio-dentales), l'espace entre la langue et les alvéoles (les alvéolaires) et l'espace entre la langue et le palais dur (les palatales). Notons que pour les alvéolaires et les palatales, la langue entre en contact avec le lieu d'articulation, mais uniquement sur les parties latérales, laissant passer l'air au milieu. Le lieu de resserrement donne à la cavité buccale une configuration particulière qui détermine la fréquence du bruit. Le bruit des labio-dentales est de faible intensité (c'est d'ailleurs leur principale caractéristique), il présente des pics diffus vers 3, 5 et 8 kHz. Celui des alvéolaires, plus intense, se situe entre 4 et 8 kHz avec des pics autour de 5 kHz et 8 kHz. Enfin pour les palatales, le bruit de turbulence se situe entre 2 et 7 kHz avec un pic diffus dont la moyenne se trouve autour de 4 kHz. Les fricatives labio-dentales sont /f/, /v/. Les fricatives alvéolaires sont /s/, /z/. Les fricatives palatales sont /ʃ/, /ʒ/.

### c. Les consonnes vocaliques

L'appellation « consonnes vocaliques » est peu répandue. Certains auteurs utilisent les termes « Consonnes sonantes » ou d'autres ne regroupent pas ces consonnes dans une seule classe mais détaillent chaque type. Nous avons choisi d'utiliser ces termes car ils reflètent à la fois l'aspect consonantique

(rétrécissement du tractus) et vocalique (structure de formant) de ces consonnes. D'une certaine façon, il s'agit de sons intermédiaires plutôt considérés comme des consonnes en raison de la place qu'ils occupent dans la syllabe (attaque ou coda, cf. chapitre 14).

Cette catégorie est bien plus complexe à décrire que les autres catégories de consonnes. Elle est en fait constituée d'éléments très divers. On distingue les approximantes latérales, les vibrantes, les nasales et les glissantes. Ces types de consonnes ont peu de caractéristiques communes, si ce n'est le rétrécissement partiel ou transitoire du tractus vocal qui permet l'apparition de structures de formants moins stables et moins intenses que celles des voyelles. Même si chacune de ces consonnes possède des structures formantiques spécifiques, les fréquences des formants sont relatives aux consonnes et voyelles adjacentes.

Il existe une approximante en français, la latérale /l/. Les approximantes présentent un rétrécissement du tractus vocal, mais moins important que celui des fricatives. Ce faible rétrécissement, intermédiaire entre l'articulation vocalique et consonantique, ne génère pas de bruit. Le lieu d'articulation de /l/ est situé sur l'avant du palais : l'apex de la langue va se positionner sur les alvéoles, laissant passer l'air sur l'un ou les deux côtés de la langue, d'où son nom de latérale. Le F1 est autour de 300 Hz tandis que le F2 est fonction des voyelles adjacentes (Fig. 7).

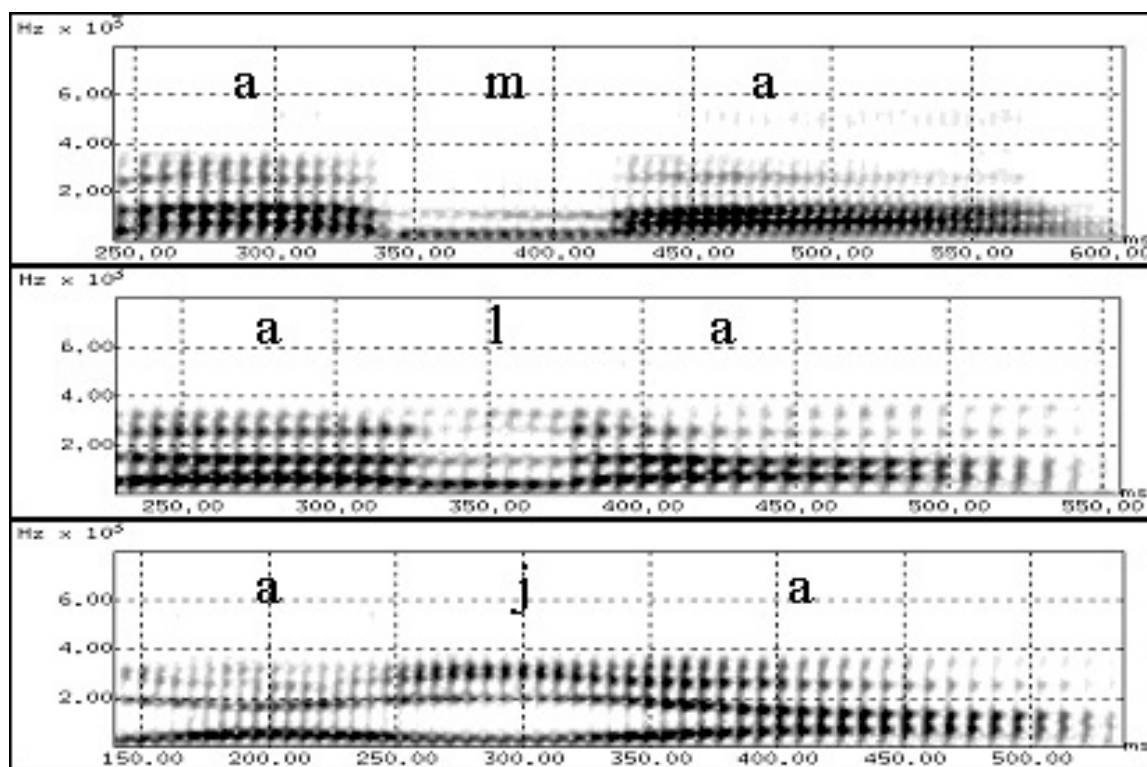


Fig. 7. Signaux spectrographiques des séquences /ama/, /ala/ et /aja/. On note la faible intensité des formants des consonnes comparés à ceux des voyelles. On observe également l'aspect totalement continu des variations formantiques de la

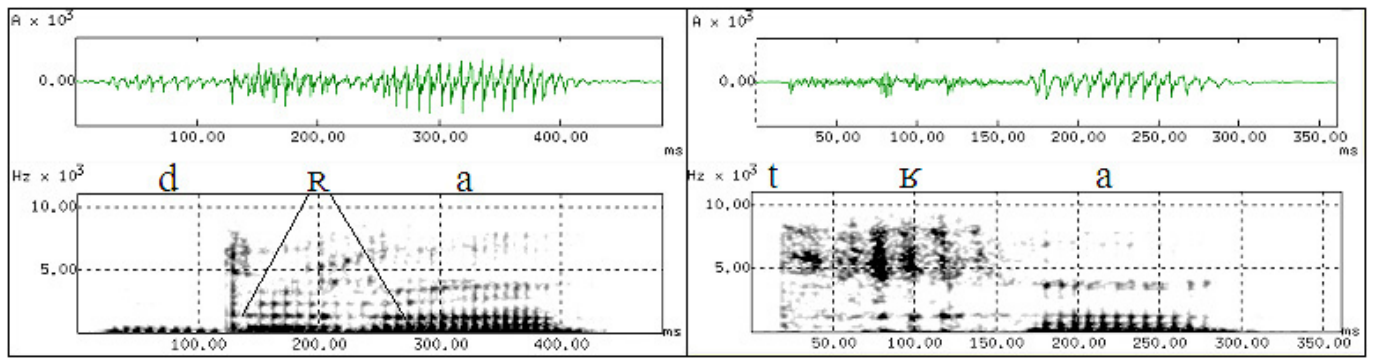


Fig. 8. Signaux temporel (en haut) et spectrographique (en bas) des séquences /dRa/ (à gauche) et /tʁa/ (à droite). On observe la partie vocalique puis battue (chute d'intensité) du /R/ dans la séquence voisée /dRa/. En revanche, dans la séquence dévoisée /tʁa/, le /ʁ/ est totalement dévoisé et fricatif.

Tout comme pour l'approximante, il n'existe qu'une vibrante (ou trill en anglais) en français, il s'agit du /r/. La consonne /r/ a un statut tout à fait particulier car c'est le son du français le plus multiforme. Il prend en effet des aspects extrêmement variés en fonction du contexte phonétique dans lequel il se trouve. Traditionnellement considéré comme une vibrante en français, il est, en fait, rarement réalisé ainsi. Une vibrante est une consonne pour laquelle la langue va entrer en battements le plus souvent multiples au niveau du lieu d'articulation (alvéolaire ou uvulaire). La vibrante alvéolaire /r/ (« r roulé ») n'est présente que dans certains accents régionaux et tend à disparaître. Le lieu d'articulation du « r » le plus fréquent en français est uvulaire (donc le son le plus postérieur du français). Il est réalisé vibrant (/R/ dans le tableau de l'IPA) uniquement en groupe consonantique lorsqu'il est précédé d'une occlusive ou d'une fricative sonore (Fig. 8). Ce /R/ est alors constitué de trois parties : une partie voyelle suivie d'une partie battue suivie elle-même d'une autre partie voyelle (pour une description complète, cf. Meunier, 1994). En revanche, on observe une réalisation fricative /ʁ/ dans les groupes consonantiques où les occlusives ou fricatives sont sourdes. Dans ce dernier cas, le /ʁ/ est totalement dévoisé (assimilation de voisement avec la consonne qui précède) et prend le statut d'une consonne fricative (Fig. 8). Enfin, en contexte intervocalique, le « r » est en général réalisé comme une approximante : il n'y a pas de pas de battement et ce n'est pas une fricative. Le lieu d'articulation reste uvulaire dans tous les cas.

Nous avons mentionné les consonnes nasales dans la partie relative aux occlusives en raison de l'obstruction complète du conduit oral. De ce fait, ces consonnes sont souvent considérées comme des occlusives. Il existe deux consonnes nasales en français : /m/ et /n/. D'un point de vue articulatoire, elles sont totalement comparables aux occlusives orales si ce n'est que l'abaissement du velum autorise

la sortie de l'air par les fosses nasales. Ainsi, /m/ est une bilabiale. Son articulation est identique à celle de /p/ et /b/ excepté l'abaissement du voile du palais. De même, /n/ est une alvéolaire comme /t/ et /d/. Les formants de /m/ sont plus graves que ceux de /n/. Comme pour les voyelles nasales, on note pour ces consonnes la présence d'anti-formants dus au couplage des cavités orale et nasale (Fig. 7).

Deux autres nasales sont souvent mentionnées comme étant des phonèmes du français. Il s'agit de /ɲ/ et de /ŋ/. Elles sont constituées chacune de deux éléments consonantiques, /n/ + /j/ (pour /ɲ/) et /n/ + /g/ (pour /ŋ/). Leur apparition dans le discours est relativement rare et fortement conditionné par le lexique. /ɲ/ est censé apparaître dans des mots comme « pagne » (/paɲ/) ou « peigne » (/peɲ/). Toutefois, aucun argument convaincant ne peut expliquer pourquoi il s'agit d'un phonème spécifique et non une réalisation allophonique de /nj/. Le phonème /ŋ/, quant à lui, n'apparaît que dans des mots empruntés à l'anglais. De plus, ces phonèmes consonantiques ont un statut syllabique peu cohérent avec les autres consonnes : toutes les autres consonnes du français peuvent se combiner pour constituer un groupe de consonnes soit en attaque, soit en coda de syllabe. Les deux consonnes /ɲ/ et de /ŋ/ ne peuvent, elles, se combiner avec d'autres consonnes. Ces arguments nous amènent à considérer qu'il s'agit, en français, de groupes de consonnes plutôt que de consonnes simples.

Les glissantes (ou glides en anglais) sont par excellence des éléments phonétiques intermédiaires entre consonnes et voyelles. Certains auteurs les font appartenir à la classe des approximantes. Elles sont, de ce fait, également appelées semi-voyelles ou semi-consonnes. Elles sont au nombre de trois en français : /j/, /ɥ/ et /w/. Leur articulation se fait en référence à l'articulation des trois voyelles fermées : /i/ pour

/j/ (palatale), /y/ pour /ɥ/ (labio-palatale) et /u/ pour /w/ (labio-vélaire). Les glissantes sont caractérisées par une structure formantique faible, en variation constante et donc sans partie stable, à la différence des voyelles. Cette instabilité s'explique par un resserrement des articulateurs plus important pour les glissantes ainsi que par une articulation en mouvement. Leur caractère transitoire et vocalique fait qu'il est impossible de distinguer avec assurance la fin d'une glissante du début de la voyelle qui suit (Fig. 7).

On remarquera parfois une nomenclature articuloire peu cohérente pour les glissantes. Elle s'explique par une appartenance conjointe de ces phonèmes à la classe des voyelles comme à celle des consonnes. Ainsi, en tant que consonnes, elles trouvent leur description articuloire dotée des termes propres aux lieux d'articulation consonantiques : palatale, labio-palatale et labio-vélaire. Ces termes, peu explicites pour des consonnes du français (le terme « labio » réfère en général, pour les consonnes, au lieu de constriction du tractus et non à l'arrondissement des lèvres), trouvent leurs équivalents dans la nomenclature propre à la description des voyelles de cette façon : palatale = antérieur ; vélaire = postérieur ; labio = arrondi. Il est nécessaire d'éclairer cette équivalence dans la mesure où le lieu d'articulation des glissantes se réfère à celui des trois voyelles fermées.

### III. Prototypes acoustiques et réalités de la parole

L'ensemble des descriptions que nous avons proposé ci-dessus représentent des prototypes des sons du français. En d'autres termes, il s'agit de descriptions obtenues à partir de productions très contrôlées (appelées « parole de laboratoire ») auprès de groupes de populations homogènes. Pourquoi est-il nécessaire de contrôler les productions de parole pour obtenir des prototypes acoustiques stables ? Dans la parole conversationnelle classique, les sources de variation de la parole sont nombreuses et leur impact sur les réalisations phonétiques peut être de nature à changer considérablement les valeurs prototypiques des sons décrits. En quelque sorte, l'objectif de la description des sons étant la distinctivité, il est nécessaire pour augmenter cette distinctivité de réduire les sources de variation. Toutefois, pour éviter de donner dans ce chapitre l'impression que tout écart par rapport à ces prototypes est déviant, ce qui serait erroné, nous nous

proposons d'évoquer rapidement quelques sources de variations normales de la parole.

#### 1. Enchaînements articuloires et coarticulation

Parmi les différents facteurs qui occasionnent la variabilité des sons de la parole, celui qui a probablement donné lieu au plus grand nombre d'études est la coarticulation. Elle peut être décrite comme le chevauchement et l'interaction des différents articulateurs au cours de la production de segments phonétiques successifs (Farnetani, 1997). La vitesse de production des sons de la parole (12 sons par seconde, donc environ 700 sons par minute) a pour conséquence que les gestes articuloires ne peuvent être produits indépendamment les uns des autres. Ils sont donc souvent soit anticipés, soit persistants. Par exemple, lors de la réalisation du mot « sous » (/su/), la configuration du conduit vocal présente, dès le début de la fricative, les caractéristiques du trait d'arrondissement de la voyelle /u/ dont l'articulation est anticipée, alors que dans la production de /si/, les lèvres sont en position étirées au début de l'articulation du /s/. Ces deux configurations donnent des résultats acoustiques très différents : l'arrondissement des lèvres durant la réalisation du /s/ dans /su/ augmente la taille de la cavité buccale ; la fréquence du bruit de /s/ dans /su/ est donc plus grave que celui de /si/. Dans ce cas, on observe donc un décalage par rapport aux valeurs prototypiques. Cet exemple n'est pas isolé. La production de la parole implique la présence systématique de ce type de configuration. Le cas de la coarticulation est inhérent à la production de la parole. C'est pourquoi la prise en compte des variations dues à la coarticulation est désormais incontournable dans les descriptions de la parole. Ces chevauchements systématiques des gestes articuloires trouvent un cadre privilégié au sein de la syllabe (cf. chapitre 14). Toutefois, des travaux récents ont pu montrer des phénomènes d'anticipation ou de persistance dans des unités dépassant largement le cadre de la syllabe.

#### 2. Caractéristiques du locuteur

L'anatomie du conduit vocal, qui change selon le sexe et l'âge, est à l'origine d'une autre forme de variation totalement inhérente à la production de la parole. La hauteur de la voix est le paramètre le plus apparemment affecté par l'âge ou le sexe des locuteurs. Si la F0 des enfants de deux ans est autour de 400 Hz, elle baisse continuellement pour se stabiliser après la puberté entre 200 Hz et 250 Hz pour les femmes tandis que celle des hommes se situe entre 100 Hz et 150 Hz. Si chaque auditeur est capable d'identifier le

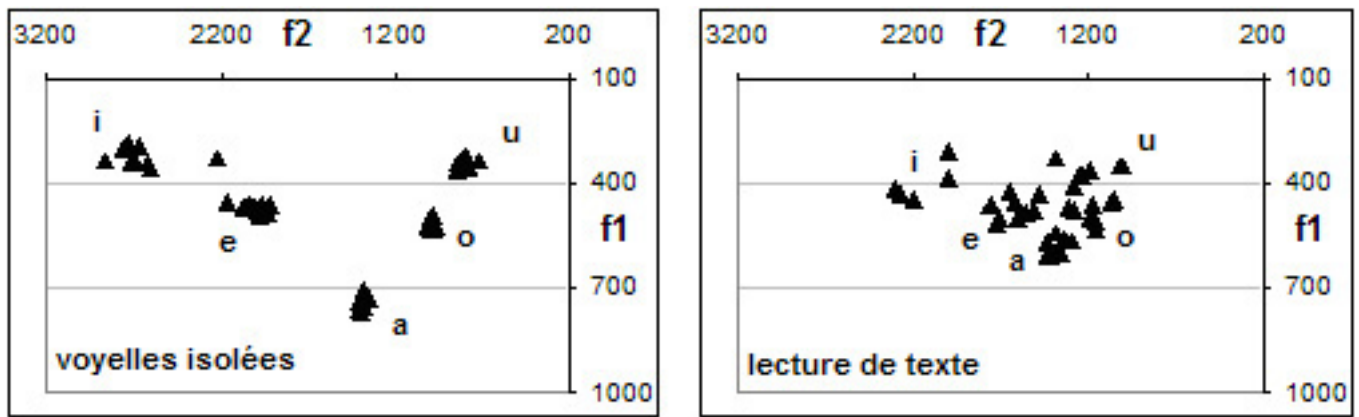


Fig. 9. Réalisations acoustiques des voyelles /i/, /e/, /a/, /o/, /u/ prononcées isolément (à gauche) et dans un texte. On observe la réduction considérable de la taille globale du système vocalique ; les valeurs moyennes de formants observées en lecture de texte sont très éloignées des valeurs prototypiques (d'après Meunier *et al.*, 2006).

sexe du locuteur d'après la hauteur de la voix, il lui sera beaucoup plus difficile de prendre conscience que les hommes et les femmes produisent des voyelles dont les valeurs acoustiques sont différentes. Pourtant ces variations, dues aux tailles des résonateurs (plus larges pour les hommes), sont bien réelles et engendrent des valeurs de formant plus élevées pour les femmes (Tubach, 1989). Ainsi, alors que les trois formants de /e/ pour un homme sont en moyenne de 365 Hz, 1961 Hz et 2644 Hz (cf. Tableau 1), ils augmentent pour une femme à 417 Hz, 2351 Hz et 3128 Hz en raison de la taille réduite des différentes cavités.

Indépendamment des caractéristiques anatomiques des locuteurs, il existe également des variations dues à la spécificité de chaque locuteur, son origine géographique, ses caractéristiques sociales, etc. Par exemple, les locuteurs du nord-est de la France produisent des /a/ (/ɑ/) plus postérieurs que ceux de la moyenne des Français.

### 3. Type de parole

On oppose depuis quelques années la parole de laboratoire à la parole « naturelle » ou « spontanée ». L'analyse de la parole produite dans des conditions réelles est relativement récente. Une des raisons majeures de cette marginalité est la très forte variabilité des sons produits en parole spontanée, non contrôlée, mais aussi l'identification délicate des sources de variations qui affectent le signal. Les réalisations des unités phonétiques sont ainsi très éloignées de leurs valeurs prototypiques. Dans une étude portant sur la réalisation et la perception des occlusives du français, Duez (1995) montre qu'en parole spontanée seulement la moitié des consonnes sont identifiées correctement. Parler est une adaptation constante aux situations de communication (Lindblom, 1990). L'articulation peut être soutenue si la situation le

demande (contexte peu informatif, milieu bruité, etc.) ou relâchée si la situation est peu ambiguë. Plusieurs travaux (Lindblom, 1963 ; Gendrot et Adda-Decker, 2006) ont ainsi pu montrer que plus le débit de parole augmente, moins les cibles articulatoires sont atteintes. De même, l'articulation des voyelles tend à se réduire en même temps que la production de la parole est contextualisée (Meunier *et al.*, 2006). Le triangle caractéristique de la représentation acoustique des voyelles montre que les voyelles sont produites en périphérie lorsqu'elles sont prononcées isolément, tandis qu'elles se centralisent lorsqu'elles sont produites dans un texte lu (Fig. 9). Cela signifie que l'articulation des voyelles en lecture de texte tend vers une voyelle centrale /ə/, pour laquelle les gestes articulatoires sont les plus neutres. En conséquence, la parole très contrôlée (celle qui permet de fournir des mesures prototypiques) est caractérisée par une hyperarticulation, alors que la parole contextualisée est sous-articulée.

## IV. Conclusion

Nous avons cherché, dans ce chapitre, à dresser un portrait des réalisations acoustiques des sons du français. Ce portrait n'est pas toujours simple étant donné la complexité de l'organisation des sons dans une langue. Si certaines classes de sons semblent former un système articulatoire et acoustique organisé en français (les voyelles, les occlusives, les fricatives), d'autres sons semblent plus isolés et n'entrent pas dans l'organisation d'un système de traits aussi régulier. D'une certaine façon, il existe des sons très ouverts (les voyelles) et d'autres très fermés (les occlusives et les fricatives). Entre les deux, les réalisations articulatoires semblent plus mouvantes et moins simples à expliciter.

Nous insistons sur le fait que les réalisations acoustiques des sons d'une langue sont des phénomènes bien plus mouvants qu'une simple description pourrait le laisser penser. La production des sons est une composante de l'organisation plus générale de la production du langage. Cette organisation suppose des interactions complexes entre les différents niveaux linguistiques ainsi que des contraintes cognitives mettant en relation un locuteur avec un (ou des) auditeurs. Aussi, la présentation de valeurs acoustiques prototypiques ne représente que la partie émergente de l'iceberg des productions sonores du langage en condition normale. Cette partie émergente est celle qui apparaît lorsque l'on isole la production des sons du contexte linguistique.

## Références

1. Duez D. *On spontaneous French speech : Aspects of the reduction and contextual assimilation of voiced stops*. *J Phonetics* 1995, 23: 407-427.
2. Farnetani E. *Coarticulation and connected speech*. In Hardcastle WJ, Laver J (eds). *Handbook of Phonetic Sciences*. Blackwell. Oxford. 1997, pp 371-404.
3. Gendrot C, Adda-Decker M. *Analyses formantiques automatiques en français: périphéralité des voyelles orales en fonction de la position prosodique*. *Actes des 26èmes Journées d'études sur la Parole, Dinard, juin 2006* : 407-410
4. *International Phonetic Association. (1999). Handbook of the International Phonetic Association. A Guide to the Use of the International Phonetic Alphabet*. Cambridge University Press. Cambridge.
5. Lindblom B. *Spectrographic study of vowel reduction*. *J Acoust Soc Am* 1963, 35: 1773-1781.
6. Lindblom B. *Explaining phonetic variation : a sketch of the hyper- and hypospeech theory*. In Hardcastle WJ, Marchal A (eds). *Speech Production and Speech Modelling*. Kluwer Academic Publishers. 1990, pp 403-439.
7. Lisker L, Abramson AS. *A cross-language study of voicing in initial stops: acoustical measurements*. *Word* 1964, 20: 384-422.
8. Meunier C. (1994). *Les groupes de consonnes : problématique de la segmentation et variabilité acoustique*. Thèse de l'Université de Provence. Aix-en-provence. France.
9. Meunier C, Espesser R, Frenck-Mestre C. *Aspects phonologique et dynamique de la distinctivité au sein des systèmes vocaliques : une étude inter-langues*. *Actes des 26èmes Journées d'études sur la Parole, Dinard, juin 2006* : 333-336.
10. Rossi M. *Segmentation automatique de la parole. Pourquoi ? Quel segments ? Traitement du signal* 1990, 7: 315-326.
11. Schwartz JL, Boë LJ, Vallée N. *Major trends in vowel system inventories*. *J Phonetics* 1997, 25: 233-253.
12. Serniclaes W. (1987). *Etude expérimentale de la perception du trait de voisement des occlusives du français*. Thèse de l'Université Libre de Bruxelles. Bruxelles. Belgique.
13. Tubach JP. (1989). *La parole et son traitement automatique*. Masson. Paris.