19es J.E.P. - Bruxelles, 19 au 22 mai 1992.

EVALUATION PERCEPTIVE D'UN CORPUS DE VOYELLES FRANCAISES EMISES ISOLEMENT PAR PLUSIEURS LOCUTEURS SELON DIVERSES FORCES DE VOIX

Jean-Sylvain LIENARD (*) - Maria-Gabriella DI BENEDETTO (**)

(*) LIMSI-CNRS, Orsay, France
(**) Dipart. INFOCOM, Università La Sapienza, Roma, Italia

Résumé

Le but de cette étude est double. En premier lieu il s'agit de construire une base de données des voyelles du français, validées perceptivement, associant à chaque segment de signal une description linguistique et diagnostique aussi complète que possible. En second lieu il s'agit d'étudier plus précisément, à partir des résultats de l'évaluation perceptive, un descripteur linguistique (l'identité de la voyelle) et deux descripteurs diagnostiques (le genre du locuteur et la force de voix), en eux-mêmes et dans leur relations mutuelles.

I - INTRODUCTION

La présente étude se rapporte à l'élaboration d'une petite base de données sur les voyelles du français prononcées isolément par plusieurs locuteurs avec diverses forces de voix, et à son évaluation perceptive par un groupe d'auditeurs, selon divers critères. Elle est présentée en détail dans [Liénard et Di Benedetto 1992].

L'objectif à long terme est d'aborder les problèmes généraux du traitement automatique de la parole (reconnaissance, synthèse, transmission) selon une perspective qui prenne en compte les aspects non-linguistiques (ou "diagnostiques") du signal [Liénard 1990]. Malheureusement les bases de données existantes se limitent à une description linguistique normative du signal, et sont rarement validées par des tests perceptifs. Pour cette raison nous avons résolu d'enregistrer de nouvelles données et de leur associer une description symbolique validée perceptivement. Dans un premier temps nous avons choisi d'étudier les voyelles, qui sont le support de la voix et de la parole, et dont la reconnaissance est un problème encore non résolu, si l'on considère la variabilité associée à divers locuteurs et à diverses conditions d'élocution.

Cet article décrit d'abord le contenu de la base de données, les conditions d'enregistrement et le protocole d'évaluation. A partir du seul fichier d'évaluation on tente ensuite de définir plus précisément la nature des divers descripteurs mis en jeu, concernant essentiellement l'identité de la voyelle, le genre du locuteur, et la force de voix. On en conclut que les interactions entre ces descripteurs sont nombreuses et importantes, et qu'elles ne peuvent être négligées en traitement automatique de la parole.

II - LE CORPUS ET SON ENREGISTREMENT

La base de données CORENC comporte une série de 12 voyelles françaises, prononcées par 13 locuteurs d'un même groupe familial (6 hommes et 7 femmes, d'âge compris entre 19 et 88 ans) selon trois "forces de voix" (ou styles) différentes, en deux sessions distantes de six mois. L'utilisation de voyelles isolées se justifie par le fait que la plupart des voyelles correspondent en français à un mot lexical (voir par exemple "ah", "a", "euh", "eux", "hi", "y", "oh", "haut", "où", "houx", "hue", "et", "hé", "est", "haie", "en", "an", "on", "un", "hein", etc). Les seules voyelles qui n'apparaissent pas dans cette liste (mais qui apparaîtraient dans des variantes régionales) sont /ɔ/ et /oe/. Dans la préparation de la base de données numérique on a laissé de côté /õe/, /ɔ/ et les deux variantes du /A/.

Le locuteur était assis en un endroit bien défini (dans une pièce d'habitation), la bouche à 30 cm d'un microphone omnidirectionnel. Pour faire varier le style de voix du locuteur on lui a demandé de répéter les voyelles prononcées par l'opérateur, à un niveau sonore adapté à la distance à laquelle se trouvait celui-ci. Ainsi chaque ensemble de trois séries se compose d'une série "N" (voix Normale, interlocuteurs à environ 1,50 m l'un de l'autre), d'une série "P" (situation "Proche", voix plus faible, interlocuteurs à environ 40 cm), d'une série "L" (situation "Lointaine", voix plus forte, interlocuteurs à environ 6 mètres). Les variations de style se traduisent essentiellement par des variations de "force de voix" -- et nous utiliserons indifféremment les deux termes dans cette étude -- mais il ne faudrait pas en déduire

qu'elles se réduisent à de simples variations de niveau sonore. Les nombreux paramètres physiques mis en jeu dans la transmission de cette information (amplitude, mais aussi pente spectrale, position des formants, Fo, texture sonore) feront l'objet d'une étude ultérieure.

Le signal a été numérisé à 10 kHz, le niveau d'entrée sur la carte de numérisation étant maintenu constant (sauf pour les segments "S", correspondant à certains segments de la série "L", voix forte, pour lesquels on a pratiqué une atténuation de 6 dB). Chaque segment a été écouté, visualisé sous forme de spectrogramme, délimité (de -50 ms approximativement avant l'apparition du voisement, à +50 ms après son extinction), et rangé dans un fichier de signal.

III - EVALUATION PERCEPTIVE

A chaque segment de signal du corpus seront associés divers descripteurs symboliques:

- l'identité de la voyelle que l'on a demandé au locuteur de prononcer ("voyelle requise"),
- l'identité de la voyelle perçue majoritairement par un groupe d'auditeurs ("voyelle perçue"),
- l'identité du locuteur qui a effectivement produit le segment considéré ("locuteur réel", ou "locuteur"),
- le genre du locuteur effectif ("genre réel", ou "genre locuteur"),
- le caractère masculin ou féminin de la voix, tel qu'il est perçu ("genre perçu"),
- la "force de voix" qui a été suggérée au locuteur ("style requis"),
- la "force de voix", telle qu'elle a été perçue par le groupe d'auditeurs ("style perçu").

Les segments sont présentés en ordre aléatoire et écoutés par les auditeurs au moyen d'un casque professionnel. Les 40 premiers segments sont répétés à la fin du test de façon à ménager une période d'habituation. L'auditeur écoute chaque segment une seule fois et reporte ses évaluations sur un formulaire, les différentes cases étant notées par deux lettres (p.ex. "aa", "éé", "in", "ee", "eu" etc pour les voyelles, "hh" ou "ff" pour le genre, "pp", "mf" ou "ff" pour le style). Il peut cocher une case spéciale ("??") en cas d'indécision. Il peut régler à sa convenance le niveau d'écoute au début du test. Le test est relativement long (environ 3 heures) et se fait en plusieurs séances. En moyenne le temps séparant l'écoute de deux segments successifs est de l'ordre de 15 secondes. L'évaluation a été faite par 7 sujets. On a ainsi obtenu des taux bruts portant sur 5824 évaluations élémentaires pour chaque descripteur.

En suivant l'évolution de ces taux bruts par tranches de 26 segments du début à la fin du test on n'a mis en évidence qu'un très faible effet d'accoutumance, variable selon les auditeurs, concernant plus les descripteurs diagnostiques que le descripteur phonétique.

On n'a pas inclus dans les résultats définitifs les évaluations de deux auditeurs, l'un parce qu'il était aussi locuteur et l'autre parce que ses résultats se situaient nettement en dehors de la moyenne.

En ce qui concerne le style perçu chaque auditeur n'avait que trois réponses possibles (voix faible, moyenne ou forte). Les réponses des cinq auditeurs sélectionnés ne coincident pas toujours. Pour classer chaque segment dans une des trois catégories une valeur numérique est associée à chaque réponse (soit 0.0 pour une réponse "voix faible", 1.0 pour une réponse "voix moyenne" et 2.0 pour une réponse "voix forte"). A chaque segment est associée la note moyenne comprise entre 0 et 2 et variant par pas de 0.13. La limite entre les catégories (perçues) P et N est définie comme le milieu de l'intervalle séparant les notes moyennes des catégories (requises) P et N. Le même processus est utilisé pour délimiter les catégories (perçues) N et L.

Les segments S ont été utilisés pour vérifier la cohérence des évaluations de style en ce qui concerne une simple différence de niveau sonore. On a constaté que le niveau sonore jouait un rôle dans l'évaluation du style, mais que ce rôle pouvait être considéré comme secondaire pour notre propos: une voix faible, amplifiée électroniquement, ne donne pas une voix forte.

Le résultat de l'évaluation est matérialisé par le fichier descripteur des segments. Il comprend le nom de chaque segment, les trois descripteurs "voyelle requise", "genre locuteur", "style requis", les trois descripteurs perceptifs correspondants, et les notes (3, 4 ou 5) qui indiquent avec quelle majorité ont été évaluées les valeurs des descripteurs "voyelle perçue" et "genre perçu". Cette note n'est pas explicitement donnée pour le descripteur "style perçu", puisqu'elle a servi à déterminer les limites des catégories de style perçu. On a également porté en fin de ligne une évaluation résumée, sous forme de "+" et de "-", qui indiquent pour chaque descripteur si la valeur du descripteur perçu correspond ou non à la valeur du descripteur requis.

Sur l'ensemble des 792 segments de la base le nombre de segments évalués comme corrects sous tous les aspects simultanément (segments "+++") est de 491. Il y a donc 301 segments comportant au moins une erreur, soit 38.0%. Un seul segment a été évalué comme erroné sous tous les aspects ("---"), soit 0.1%. Dans tout ce qui suivra les 72 segments S seront exclus des statistiques.

Le tableau 1 montre que les erreurs phonétiques se répartissent à peu près régulièrement selon le locuteur entre 0 et 30%, ainsi que les erreurs sur le style (entre 17 et 55%) ; mais les erreurs sur le genre du locuteur sont presque uniquement relatives à un seul locuteur (JB). Le locuteur produisant le maximum d'erreurs de style est une personne âgée.

loc	gen	n	%voy	%gen	%sty	pmf
AB	F	36	11.1	0.	38.9	+ +
AM	F	72	4.1	1.4	31.9	14
CB	F	72	16.7	0.	30.6	-4
JB	F	108	4.6	25.0	16.7	0
MF	F	36	16.7	2.8	55.6	10
SA	F	36	5.6	0.	38.9	-3
SB	F	36	13.9	0.	27.8	-6
BB	H	36	8.3	2.8	22.2	-5
DB	H	36	2.8	0.	25.0	--
JP	H	72	12.5	0.	27.8	0
MB	H	72	6.9	0.	19.4	11
ML	H	36	30.6	0.	27.8	-10
OB	H	72	0.	0.	33.3	-2
ens		720	9.2	4.0	28.6	1

Tableau 1 : erreurs sur l'ensemble du corpus, segments S exclus, par locuteur; n est le nombre de segments fournis par chaque locuteur, pmf est le rapport (plus fort/moins fort), voir partie VI

Le décompte des erreurs phonétiques faites par les groupes de locuteurs masculins et féminins ne fait pas apparaître de différence significative. La comparaison des résultats d'évaluation obtenus dans les mêmes conditions par les 6 locuteurs qui ont pris part aux deux sessions montre une différence légère mais non significative.

IV - DESCRIPTEUR "IDENTITE DE LA VOYELLE"

On trouve dans la littérature scientifique de nombreuses études sur la perception des voyelles. Parmi les principales lignes de force du thème se trouvent l'influence du contexte phonétique, l'influence de la tâche assignée aux auditeurs, et l'influence de l'ordre de présentation des signaux successifs.

Lorsque l'étude est faite sur des voyelles dites isolées, il s'agit très souvent, en fait, de syllabes CVC. Dans une étude classique des voyelles américaines [Peterson and Barney 1952] celles-ci sont prononcées dans un environnement /h-d/. Pour l'analyse acoustique on ne s'intéresse qu'à la partie stable de la voyelle, mais la validation perceptive de ces voyelles est faite par écoute de l'ensemble du segment CVC. Certains auteurs ont obtenu des résultats d'identification nettement meilleurs avec un contexte CVC plutôt que hors contexte [Strange et coll. 1976]. Ces résultats ont été infirmés par la suite [Kahn 1978, Macchi 1980], avec une meilleure sélection des locuteurs et des auditeurs (pour s'assurer qu'ils ont la même origine linguistique) et un contrôle plus strict des conditions d'élocution.

L'importance de la tâche assignée aux auditeurs a été mise en évidence par divers auteurs [Assman et coll. 1982] : le comportement des auditeurs diffère selon qu'ils doivent choisir une catégorie phonétique, choisir un mot comprenant la même voyelle, répéter le son qu'ils ont perçu.

	/A/	/i/	/u/	/e/	/ɛ/	/y/	/ø/	/oe/	/o/	/ɛ̃/	/ɑ̃/	/ɔ̃/	??	tot	err	%
/A/	58 .									1	1			60	2	3
/i/		54		2		3							1	60	6	10
/u/			58						1				1	60	2	3
/e/				58									2	60	2	3
/ɛ/				1	55			1					3	60	5	8
/y/				1		59								60	1	2
/ø/							58	2						60	2	3
/oe/							13	47						60	13	22
/o/			4						55				1	60	5	8
/ɛ̃/								2		54	1		3	60	6	10
/ɑ̃/	1									1	52	2	4	60	8	13
/ɔ̃/			2				1		1		3	46	7	60	14	23
tot	59	54	64	62	55	62	72	52	57	56	57	48	22	720	66	
dif	-1	-6	4	2	-5	2	12	-8	-3	-4	-3	-12				

Tableau 2 : confusions phonétiques faites majoritairement par les auditeurs, tous locuteurs et tous styles confondus. La ligne "tot" représente le total de chaque colonne. La ligne "dif" représente le pouvoir attracteur, c'est-à-dire la différence, pour chaque voyelle, entre le total en colonne ("voyelle perçue") et le total en ligne ("voyelle requise").

L'ordre de présentation des stimuli peut être "bloqué" ("blocked"), tous les signaux d'une même série provenant d'un même locuteur, ou "mixte" ("mixed") : dans ce cas les signaux provenant de divers locuteurs sont présentés aléatoirement ([Strange et coll. 1976], [Verbrugge et al. 1976]), si bien que les auditeurs n'ont pas la possibilité d'utiliser des informations d'adaptation au nouveau locuteur. Le problème avait été déjà évoqué par Ainsworth [Ainsworth 1975] qui a introduit les termes de "perception intrinsèque" et "perception extrinsèque".

Le tableau 2 résume les confusions phonétiques pour l'ensemble des locuteurs, tous styles confondus. Le nombre total de confusions et de non-décisions est de 66 (soit 9.2%). Les voyelles les plus mal identifiées sont /$\tilde{\text{ɔ}}$/ et /oe/ , qui donnent lieu respectivement à 23% et 22% d'erreur. Les voyelles les mieux identifiées sont /A/, /u/, /e/, /y/, avec moins de 3% d'erreur.

On notera que /oe/ est souvent identifiée comme /Ø/, beaucoup plus que /Ø/ comme /oe/. Il s'agit d'une particularité régionale de nos locuteurs, dont plusieurs ont pour origine la région Sud-Est de la France. Sur les 13 erreurs /oe/ -> /Ø/, 10 proviennent de locuteurs d'origine grenobloise, et 6 parmi ces 10 sont imputables à un même locuteur. Par contre les erreurs observées sur /$\tilde{\text{ɔ}}$/ ne bénéficient pas systématiquement à une autre voyelle, et l'on observe alors un maximum de non-reconnaissances.

On a reporté au bas du tableau 2 deux lignes représentant respectivement le total observé pour chaque colonne, c'est-à-dire le score de chaque voyelle en tant que "voyelle perçue", et la différence par rapport au total en ligne (qui représente le score en tant que "voyelle requise"). Cette différence, ou "pouvoir attracteur", montre que certaines voyelles (que nous qualifierons de "fortes") attirent les suffrages qui devraient aller à d'autres, et que, réciproquement, certaines (que nous qualifierons de "faibles") sont plus fragiles. Le glissement /oe/->/Ø/ est ainsi bien mis en évidence : pour les locuteurs de notre corpus il est manifeste que /Ø/ est une voyelle forte, et /oe/ une voyelle faible. On voit aussi que /u/ est plutôt une voyelle forte, que /$\tilde{\text{ɔ}}$/ est très faible, que /i/, /ɛ/ et /$\tilde{\text{ɛ}}$/ sont plutôt faibles.

Nos résultats montrent une grande différence entre les voyelles orales et les voyelles nasales. En effet, nous avons pour les 9 voyelles orales un total de 38 erreurs, soit en moyenne 7% par voyelle, alors que les 3 voyelles nasales produisent en tout 28 erreurs, soit en moyenne 16% par voyelle. Si l'on considère le pouvoir attracteur, la différence entre orales et nasales apparaît encore plus nettement : en moyenne celui-ci est quasi-nul pour les orales (-0.1), alors qu'il est notablement négatif (-6.3) pour les nasales. Cela signifie que les nasales ont tendance à être prises pour des orales, mais pas l'inverse.

La répartition du nombre d'erreurs selon le locuteur montre que le taux d'erreur varie entre 0 et 31% (tableau 1). Certaines erreurs paraissent liées à un parler régional et sont plutôt de nature linguistique (cas de la confusion /oe/->/Ø/ chez CB, par exemple), d'autres semblent caractériser individuellement le locuteur et sont plutôt de nature diagnostique (cas de la mauvaise prononciation de /$\tilde{\text{ɔ}}$/ chez ML, de /$\tilde{\text{a}}$/ chez SB et de la confusion /i/->/y/ chez MF).

Nos résultats pour l'identification phonétique peuvent être comparés à ceux de Assman et coll. [Assman et coll. 1982], avec toutes les réserves tenant à la différence de langue et aux différences de protocole expérimental. Pour 10 voyelles de l'anglais canadien d'Edmonton, provenant de 10 locuteurs (5 F et 5 H), prononcées isolément, hors contexte consonantique, présentées en ordre aléatoire, ces auteurs ont défini des taux d'erreur de 11% (lorsque les auditeurs répondent au moyen d'un mot-clé /hVd/), et 9% (lorsqu'ils répondent avec un mot-clé /pVp/). Avec 9.2% nos résultats sont très voisins, mais cette coïncidence ne doit pas masquer les différences évoquées ci-dessus. Dans les deux cas on a constaté de forts écarts d'une voyelle à l'autre (de 1 à 43% chez Assman et coll. 1982); nous avons, en ce qui nous concerne, minimisé cet effet en excluant les deux variantes du /A/, ainsi que /ɔ/ et /$\tilde{\text{oe}}$/.

V - DESCRIPTEUR "GENRE DU LOCUTEUR"

La perception de ce descripteur diagnostique est rarement étudiée en tant que telle, mais le problème peut apparaître dans les conditions usuelles de la communication, par exemple au téléphone.

Notre but ici n'est pas de déterminer quels paramètres physiques sont à l'origine de la perception d'une voix masculine ou féminine, mais plutôt de mettre en évidence d'éventuelles relations entre ce descripteur et les descripteurs "locuteur" et "identité de la voyelle".

Le tableau 1 montre que, à deux exceptions près, seul le locuteur JB (féminin) est concerné. Sur les 108 segments produits par ce locuteur, 27 ont été perceptivement attribués à une voix masculine ou déclarés ambigus. Si l'on examine les erreurs en fonction de la voyelle requise, tous styles confondus, il apparaît que les erreurs se produisent surtout pour /y/, /u/, /Ø/ et /$\tilde{\text{ɔ}}$/, c'est-à dire lorsque la voyelle est arrondie.

Dans notre corpus nous n'avons observé qu'un petit nombre d'erreurs sur le genre du locuteur. Il ne faudrait pas en déduire que le problème de percevoir

ce descripteur est secondaire. Il est probable que si nous avions utilisé comme locuteurs des adolescents ou de jeunes enfants le problème se serait posé fréquemment.

VI - DESCRIPTEUR "STYLE DE VOIX"

La relation entre "style requis" et "style perçu" fait intervenir des considérations complexes de la part du locuteur (compréhension de la consigne, ajustement de la force de voix au niveau requis pour la voyelle requise, contrôle proprioceptif ou auditif), et de la part de l'auditeur: ce dernier doit évaluer l'intention du locuteur de parler plus ou moins fort, mais son jugement est influencé par le niveau sonore, dans l'absolu, du son qu'il écoute.

On ne trouve que peu de travaux concernant l'effort de parole ou le style de voix (Schulman 1985, Traunmüller 1985, Granström and Nord 1991). Ces travaux sont souvent en rapport avec les mécanismes de production de la parole. Du point de vue perceptif, on sait que le seuil de reconnaissance varie selon les voyelles [Wajskop 1971] : à intensité physique égale, les voyelles compactes, dont l'énergie est concentrée dans le centre du spectre (autour de 1000 Hz), sont perçues avec une plus grande intensité subjective que les voyelles diffuses, dont l'énergie est plutôt située vers le grave ou vers l'aigu. Les études menées sur ce sujet sont de nature psycho-acoustique, avec des stimuli calibrés qui n'ont qu'un rapport lointain avec les matériaux utilisés dans la présente étude.

La gamme de variation des styles de parole dans laquelle nous nous sommes situés est celle de la vie courante. Plus précisément nous avons voulu explorer l'intervalle de niveau sonore dans lequel on adapte l'effort vocal à la situation de communication de manière pratiquement inconsciente. Les premières mesures acoustiques montrent que d'un style à l'autre la variation moyenne est de 6 à 10 dB, soit moins de 20 dB entre extrêmes, ce qui est très peu par rapport à la dynamique possible de la voix.

Chaque ligne du tableau 3 représente la ventilation des 240 segments émis selon un même "style requis", dans les trois catégories de "style perçu" que nous avons déterminées plus haut.

	P	N	L	tot
P	162	70	8	240
N	66	138	36	240
L	0	27	213	240

Tableau 3 : style perçu, en fonction du style requis, tous locuteurs confondus

La somme des valeurs de la partie haute (au-dessus de la diagonale principale) représente le nombre de cas où, selon les auditeurs, la voix a été produite plus fortement que ne l'indiquait la consigne. Inversement la partie basse (au dessous de la diagonale principale) représente les cas de production moins forte que la consigne. Nous appellerons pmf ("plus fort/moins fort") le rapport des deux parties, et nous l'exprimerons sous forme logarithmique (10*log(pmf)) par symétrie. Ici ce rapport est de 1.22, soit +0.9 sous forme logarithmique. Cette valeur faible indique que les erreurs de style sont équilibrées dans l'ensemble.

On a reporté dans la figure 1 les deux indices caractérisant les erreurs faites sur le style, en fonction de l'identité de la voyelle requise. Il apparaît un quadrilatère /u/ /ɑ̃/ /ɛ̃/ /Ø/, à l'intérieur duquel les autres voyelles se regroupent de manière cohérente. Selon les ordonnées (rapport pmf) on voit que les voyelles /y/, /u/ et /Ø/ (arrondies, produites avec une protrusion des lèvres) tendent à induire une voix moins forte que ce qui est requis, alors que les nasales /ɑ̃/, /ɔ̃/ et /ɛ̃/ tendent à induire une voix plus forte. Selon les abscisses (taux d'erreur) on voit que /y/, /A/ et /ɑ̃/ donnent lieu à peu d'erreurs de style, alors que /ɛ̃/, /e/, /i/ et /Ø/ produisent plus d'erreurs, mais celles-ci sont peu caractéristiques.

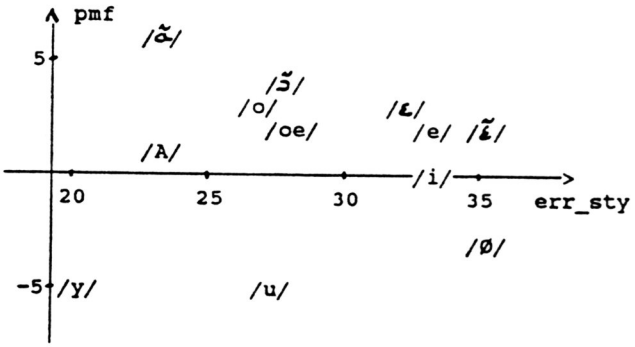

Figure 1 - Nature des erreurs de style (rapport pmf) en fonction du taux de ces erreurs, pour les diverses voyelles

Le taux d'erreur sur le style et le rapport pmf sont indiqués dans le tableau 1 pour chaque locuteur. On constate de grandes différences individuelles. Le rapport pmf indique dans quelle mesure, lorsqu'il a fait des erreurs de style, le locuteur a parlé moins fort ou plus fort que ce qui était requis.

Le locuteur JB (féminin) est le seul pour lequel on ait obtenu des évaluations de "genre perçu" masculin. Dans le tableau 4 est reporté le nombre de réponses obtenu pour chaque combinaison des descripteurs "genre perçu" et "style perçu".

	P	N	L
H	10	5	2
F	26	19	36
?	3	4	3

Tableau 4 : Répartition des 108 évaluations relatives au locuteur JB, sur les descripteurs "style perçu" et "genre perçu".

Il est clair que la confusion avec une voix masculine se produit lorsque la locutrice parle à voix moyenne ou faible. La répartition des erreurs selon la voyelle requise montre que les erreurs sur le genre du locuteur se produisent essentiellement sur les voyelles arrondies, dont on a vu plus haut qu'elles étaient moins sonores que les autres.

On peut remarquer que la locutrice JB, pratiquement seule à produire des erreurs de genre, produit très peu d'erreurs phonétiques, et très peu d'erreurs de style, et que ces erreurs sont d'autant plus importantes que la force de voix est moyenne ou faible. On peut penser que c'est justement parce qu'elle marque bien les divers styles qu'elle est amenée à exagérer la consigne, et à adopter une voix très faible quand une voix faible suffirait. Elle est amenée ainsi à baisser exagérément son fondamental, ce qui contribue à donner un caractère masculin à sa voix.

Ainsi se dessine pour ce locuteur une certaine stratégie d'élocution. Cet exemple montre bien les relations entre les trois descripteurs. Il suggère aussi que la connaissance ou la reconnaissance des descripteurs diagnostiques pourrait être mise à profit dans un système automatique pour mieux reconnaître la voyelle prononcée, ou du moins pour prédire la nature des erreurs possibles dans des circonstances données: ici la connaissance de la valeur "voix faible" du descripteur "style" permet d'augurer des erreurs de type F->H sur le genre du locuteur et des confusions sur les voyelles arrondies.

VII - CONCLUSION

Cette étude a un résultat tangible, sous la forme de deux fichiers, l'un comportant les signaux vocaliques, l'autre la description symbolique de ces éléments. Cette description, qui se veut aussi complète que possible, porte sur l'identité phonétique de la voyelle, sur le genre du locuteur, et sur le style de voix; elle reflète d'une part la consigne donnée au locuteur, d'autre part l'évaluation qui en a été faite par un groupe d'auditeurs. Ainsi ces données peuvent être utilisées en connaissance de cause dans des études de traitement automatique de la parole.

A partir des descriptions symboliques on a pu trouver des résultats perceptifs voisins de ceux obtenus dans une autre langue en ce qui concerne l'identité phonétique. On a pu aussi mettre en évidence la capacité du canal vocal à transmettre, avec de simples voyelles isolées, des informations non-linguistiques comme la force de voix et le genre du locuteur. De plus on a montré que les erreurs phonétiques, les erreurs sur le genre du locuteur et les erreurs sur la force de voix sont liées entre elles, et varient d'un locuteur à l'autre. On a ainsi justifié l'intérêt d'une approche qui prend en compte, ensemble, tous les aspects perceptifs de la parole et de la voix, au lieu de chercher dans le signal de stricts invariants linguistiques.

VII - REFERENCES

Ainsworth, W. (1975): "Intrinsic and extrinsic factors in vowel judgments", in Auditory analysis and perception of speech, ed. by G.Fant and M.Tatham, Academic, London, 103-113.

Assman, P.F., Nearey, T.M. and Hogan, J.T. (1982): "Vowel identification: orthographic, perceptual and acoustic aspects", J.Acoust.Soc.Am 71 (4), 975-989.

Granström, B. and Nord, L. (1991): "Neglected dimensions in speech synthesis", ESCA workshop on The phonetics and phonology of speaking styles, Barcelona.

Kahn, D. (1978): "On the identifiability of isolated vowels", UCLA Working Pap. Phon. 41, 26-31.

Liénard, J.S. (1990): "Perception, data variability and inductive inference", Cognitiva, AFCET, Madrid.

Liénard, J.S. et Di Benedetto, M.G. (1992): "CORENC : un corpus de voyelles françaises isolées et son évaluation perceptive selon divers descripteurs", Rapport Interne LIMSI.

Macchi, M.J. (1980): "Identification of vowels spoken in isolation vs vowels spoken in consonantal context", J.Acoust.Soc.Am. 68 (6), 1636-1642.

Peterson, G.E. and Barney, H.L. (1952): "Control methods used in a study of the vowels", J.Acoust.Soc.Am. 24 (2), 175-184.

Schulman, R. (1985): "Articulatory targeting and perceptual constancy of loud speech", Séminaire franco-suédois, ICP, Grenoble.

Traunmüller,H. (1985): "The role of the fundamental and the higher formants in the perception of speaker size, vocal effort, and vowel openness", Séminaire franco-suédois, ICP, Grenoble.

Verbrugge, R.R., Strange W., Shankweiler, D.P. and Edman T.R. (1976): "What information enables a listener to map a talker's vowel space ?", J.Acoust.Soc.Am. 60 (1), 198-212.

Wajskop, M. (1971): "Seuils de reconnaissance de voyelles isolées", Revue d'Acoustique 13, 20-22.