

EXTRACTION DE LA FONCTION D'AIRE DU CONDUIT VOCAL PAR EXCITATION EXTERNE

Hocine Teffahi *, Bernard Guerin and Amar Djeradi***

* Université des sciences et de la technologie Houari Boumédiène (USTHB),
Laboratoire de communication parlée et de traitement du signal, Faculté d'Electronique et d'Informatique, BP 32, El alia,
Alger, ALGERIE - E-mail : hteffahi@yahoo.fr

** Institut de communication parlée, INPGrenoble, 46 Avenue Félix Viallet, 38000, Grenoble, France

ABSTRACT

Knowledge of vocal tract area function is important for the understanding of phenomena occurring during speech production. We present here a new measurement method based on external excitation of the vocal tract with a known pseudo-random sequence, where the area function is obtained by a linear prediction analysis applied at the cross-correlation between the sequence and the signal measured at the lips. The advantages of this method over methods based on sweep-tones or white noise excitation are (1) a much shorter measurement time (about 100 ms), and (2) the possibility of speech sound production during the measurement. This method has been checked against classical methods through systematic comparisons on a small corpus of vowels. Moreover, it has been verified that simultaneous speech sound production does not perturb significantly the measurements. This method should thus be a very helpful tool for the investigation of the acoustic properties of the vocal-tract in various cases such as vowels.

SOMMAIRE

La connaissance de la fonction d'aire du conduit vocal est importante pour la compréhension des phénomènes qui se produisent lors d'une élocution. Nous présentons, ici, une nouvelle méthode de mesure fondée sur un mode d'excitation externe du conduit vocal par une séquence pseudo-aléatoire. La fonction d'aire est obtenue à l'aide d'une analyse par prédiction linéaire appliquée à l'intercorrélation entre le signal issu des lèvres et de la séquence pseudo-aléatoire. Les avantages de cette méthode par rapport aux méthodes à balayage de fréquence ou à excitation par bruit blanc sont (1) un temps de mesure très court (environ 100 ms) et (2) possibilité de phonation pendant la mesure. Cette méthode a été testée sur un petit corpus de voyelles. Par ailleurs, nous avons vérifié que la condition de phonation ne perturbe pas, d'une manière significative, les résultats de mesures. Enfin, cette méthode peut constituer un très bon outil pour la compréhension des propriétés acoustiques du conduit vocal lors de la production des voyelles.

1. INTRODUCTION

L'étude de la forme du conduit vocal et de son évolution au cours d'une élocution est importante pour une meilleure compréhension du phénomène phonatoire et pour ses applications dans les domaines de la reconnaissance et de la synthèse de la parole.

Plusieurs études ont été déjà faites sur l'extraction des caractéristiques géométriques du conduit vocal à partir des caractéristiques acoustiques : Schroeder (Schroeder, 1967) a décrit analytiquement la relation entre les pôles et les zéros de l'admittance du conduit vocal mesurée aux lèvres et le logarithme de la fonction d'aire des sections du conduit vocal (représenté par le développement en série de fourrier en cosinus). L'analyse a été effectuée pour des variations dans les limites de l'applicabilité de la théorie de la perturbation d'ordre 1. Pour des grandes variations, Mermelstein (Mermelstein, 1967) a développé une

procédure numérique pour estimer la fonction d'aire (paramétrer par les 6 premiers coefficients de la série de fourrier en cosinus) des singularités de l'admittance. Il a été démontré que les fréquences des formants, qui correspondent aux pôles de l'admittance, sont insuffisantes pour déterminer, uniquement, le logarithme de la fonction d'aire. Les informations nécessaires restantes peuvent être obtenues à partir des zéros de l'admittance, lesquelles, malheureusement ne peuvent être estimés à partir du signal de parole. Schroeder (1967) a alors développé un dispositif expérimental pour mesurer l'admittance du conduit vocal aux lèvres, et en utilisant une approche fréquentielle, il a été capable de déterminer de bonnes approximations de la fonction d'aire. Cependant, le problème d'estimation de la fonction d'aire à partir du signal de parole reste sans solutions.

Avec l'apparition de la prédiction linéaire (LPC) appliquée au signal de parole (Atal et al., 1971) Wakita

(Wakita, 1973, 1979) a développé une technique de filtrage inverse pour estimer la fonction d'aire à partir du signal vocal. Cependant, cette technique fait usage d'informations sur la source voisée, distribution des pertes, longueur du conduit vocal et rayonnement aux lèvres qui ne peuvent être supposés connus, à priori, avec précisions. En fait, Sondhi (Sondhi, 1979) a montré que le signal de parole, seul, ne contient pas assez d'informations pour déterminer une fonction d'aire unique, confirmant les conclusions de Mermelstein (1967) et Schroeder (1967). Yehia (Yehia et al., 1996) propose une solution se basant sur une combinaison explicite des contraintes morphologiques et acoustiques du conduit vocal.

Nous proposons dans ce travail une nouvelle méthode de mesure basée sur un mode d'excitation externe du conduit vocal par une séquence pseudo-aléatoire (Djeradi et al., 1991). La fonction d'aire du conduit vocal est obtenue à l'aide d'une analyse par prédiction linéaire de l'intercorrélation entre la séquence pseudo-aléatoire et le signal mesuré aux lèvres.

Les avantages de cette méthode sont : (1) un temps de mesure très court (100ms), (2) possibilité de phonation pendant la mesure.

2. FORMALISME THEORIQUE

Le conduit vocal peut être considéré comme un filtre acoustique linéaire de réponse impulsionnelle notée $h(t)$. Les échantillons de cette réponse sont notés $h(n)$, n étant le numéro de l'échantillon, et la fonction de transfert du filtre est H . La sortie, $y(n)$, du filtre peut s'écrire alors :

$$y(n) = [b(n) + x(n)] * h(n) \quad (1)$$

Où $b(n)$ est un bruit, indésirable, superposé à l'excitation $x(n)$. La corrélation, R_{xy} , entre $x(n)$ et le signal de sortie $y(n)$, s'exprime par:

$$R_{xy} = R_1(n) + R_2(n) \quad (2)$$

Où

$$R_1(n) = \sum_{k=-\infty}^{k=\infty} h(k) \left[\sum_{m=-\infty}^{m=+\infty} x(m) \cdot x(m+n-k) \right] \quad (3)$$

Et

$$R_2(n) = \sum_{k=-\infty}^{k=\infty} h(k) \left[\sum_{m=-\infty}^{m=+\infty} x(m) \cdot b(m+n-k) \right] \quad (4)$$

Par ailleurs, soit $\varphi_{xx}(k)$ l'autocorrélation du signal $x(n)$ et $\varphi_{xb}(n)$ la corrélation entre $x(n)$ et $b(n)$. L'équation (2) s'écrit alors :

$$R_{xy} = [h * \varphi_{xx}](n) + [h * \varphi_{xb}](n) \quad (5)$$

Sachant que $x(n)$ et $b(n)$ sont décorrelés, $\varphi_{xb}(n) = 0$
Alors

$$R_{xy}(n) = [h * \varphi_{xx}](n) \quad (6)$$

Il est évident que si $\varphi_{xx}(n)$ est une impulsion de Dirac, la réponse impulsionnelle $h(n)$ serait une image exacte de $R_{xy}(n)$. Une bonne approximation des propriétés statistiques d'un bruit blanc est une séquence pseudo-aléatoire (Julien et al., 1972), dont l'autocorrélation est un train d'impulsions e_N de période N . De ce fait :

$$R_{xy}(n) = [h * e_N](n) \quad (7)$$

Dans le cas où la longueur de la réponse impulsionnelle serait plus petite que N , la séquence $R_{xy}(n)$ correspond exactement à $h(n)$ pour n allant de 0 à $N-1$.

En utilisant le modèle de prédiction linéaire, il a été démontré que le conduit vocal est modélisé par un filtre tout pôles $H(Z)$ donné par (Atal et al., 1971) :

$$H(z) = 1 / (1 + \sum_{i=1}^p a_i \cdot z^{-i}) \quad (8)$$

Où : p : est l'ordre de prédiction et a_i : coefficients de prédiction (paramètres du filtre).

Le problème consiste à trouver un ensemble de coefficients a_i ($1 \leq i \leq p$) tel que l'erreur e_n , entre le signal original et le signal prédit, soit minimale. En traitement de la parole, le critère usuel utilisé est la minimisation de l'énergie de l'erreur (ou critère des moindres carrés), car il conduit, souvent, à une solution mathématique intéressante. Il suffit donc de minimiser l'erreur quadratique totale ou énergie de l'erreur, que nous désignons par E , qui s'écrit:

$$E = \sum_n e_n^2 = \sum_n (h_n + \sum_{i=1}^p a_i h_{n-i})^2 \quad (9)$$

On obtient le système d'équations suivantes:

$$\sum_{i=1}^p a_i \cdot R_{|k-i|} = -R_k; 1 \leq k \leq p \quad (10)$$

R_k : est l'autocorrélation de la réponse impulsionnelle du filtre, donnée par:

$$R_k = \sum_{n=0}^{N-1-|k|} h_n \cdot h_{n+|k|} \quad (11)$$

Ce système d'équations peut être résolu par l'algorithme récursif de Levinson (Levinson, 1947). L'énergie de l'erreur minimale est donnée par:

$$E_i = (1 - k_i^2) \cdot E_{i-1} \quad (12)$$

k_i : sont les coefficients de corrélation partiels ou coefficients de réflexions, et sont calculés par la méthode de Leroux (Leroux et al., 1977).

L'énergie de l'erreur normalisée, notée V_p , est définie par:

$$V_p = E_p / R_0 = \prod_{i=1}^p (1 - k_i)^2 \quad (13)$$

Avec E_p : énergie de l'erreur d'ordre p et R_0 : l'autocorrélation d'ordre zéro.

Wakita (1979) a montré que le jeu de coefficients k_i , représentant les coefficients de réflexion aux jonctions entre les sections cylindriques d'égales longueurs constituant un tube acoustique, peuvent sous certaines conditions représentés une approximation de la fonction d'aire du conduit vocal. En effet, si A_i représente l'aire de la section d'indice i , prise à partir des lèvres, alors :

$$k_i = (A_i - A_{i+1}) / (A_i + A_{i+1}) \quad (14)$$

Les aires des p sections sont calculées en posant $A_{p+1} = 1$, les aires suivantes sont, alors, estimées dans une échelle relative, pour obtenir un profil de fonction d'aire significative.

3. SIMULATION DE LA METHODE DE MESURE DE LA FONCTION D'AIRES

3.1.1. Principe

Le schéma de principe est donné à la figure 1. Pour réaliser cette simulation (Djeradi et al., 1991), il faut disposer du signal pseudo-aléatoire. Ce signal $x(n)$ est construit à partir d'une suite de nombres binaires, constituant un 'champ de nombres finis' nommé communément 'champ de Galois' (Schroeder, 1979). Dans le cadre de ce travail, qui se limite à une étude de quelques voyelles, la largeur spectrale intéressante est égale à 5Khz, nous prendrons ainsi une fréquence d'échantillonnage de 10 KHz et la durée de la séquence pseudo-aléatoire sera de 1023 échantillons. Le conduit vocal est simulé ensuite par le modèle S.I.M.O.N.D (Castelli, 1989). Ce modèle utilise les coefficients de réflexion telle que proposé par Kelly (Kelly et al., 1962). Dans ce modèle à paramètres localisés, les variations de géométrie du conduit vocal sont représentées par une succession de coefficients de réflexions. Par ailleurs, ce modèle inclut toutes les pertes : à savoir, pertes par viscosité- chaleur, par vibrations des parois et par rayonnement aux lèvres. Pour une fréquence d'échantillonnage de 10 KHz, la longueur de chaque section élémentaire est de $L = 1.765 \text{ cm}$ ($Fe = 2C/L$), où C est la vitesse du son.

Dans un premier temps, afin de valider la méthode, on utilisera, en simulation, un modèle sans pertes. Pour cela, on excite le modèle SIMOND (sans pertes) par une entrée pseudo-aléatoire pour une configuration du conduit vocal bien déterminée. La figure (2.a) donne le résultat du calcul de la fonction d'aire obtenu à l'aide d'une analyse par prédiction linéaire appliquée sur l'intercorrélacion entre le signal issu du conduit vocal, sans pertes, et de l'excitation pseudo-aléatoire. Il apparaît que l'on retrouve exactement la fonction d'aire de configuration. On suppose connu le

nombre de sections, c'est à dire la longueur du conduit vocal.

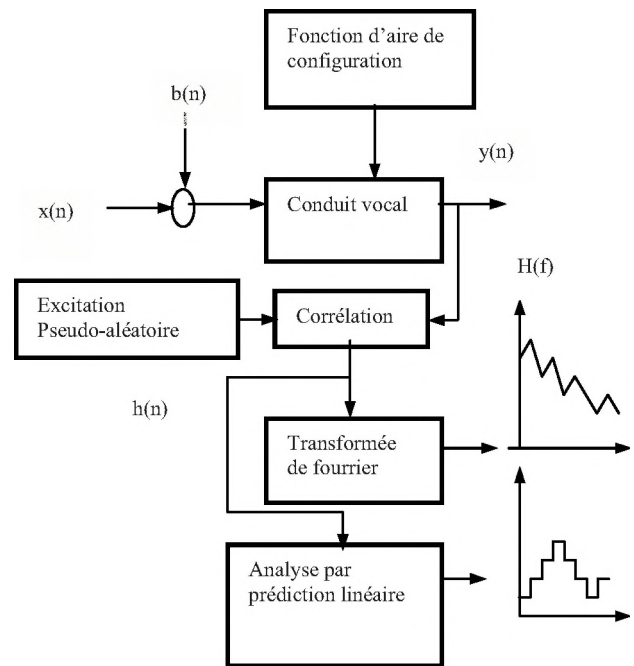


Figure 1. Diagramme du principe de mesure de la fonction d'aire et de la fonction de transfert du conduit vocal.

Pour approcher le cas réel, nous avons étudié le comportement du modèle d'inversion dans le cas d'un conduit vocal avec pertes. Les résultats obtenus ont montré que la configuration calculée ne correspondait pas à la fonction d'aire initiale. La figure (2.b) donne un exemple de résultat.

En ne prenant en compte, successivement, que l'une des sources de pertes, nos études ont montré que les pertes par viscosité chaleur et vibrations des parois n'ont pas beaucoup d'influences sur la fonction d'aire calculée, mais que les pertes par rayonnement aux lèvres jouent un rôle important lors de l'inversion pour retrouver la fonction d'aire (Teffahi, 2000).

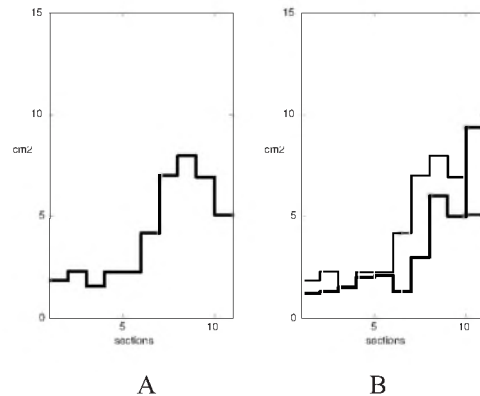


Figure 2. Résultats de simulations. (En trait fin: sans pertes; en trait gras: avec pertes. A- cas sans pertes ; B- cas avec pertes.

4. COMPENSATION DES DIFFERENTES PERTES

Pour prendre en compte les différentes pertes dans le modèle d'inversion, nous avons étudié différentes stratégies d'égalisation. Trois mécanismes de compensation vont être utilisés pour se ramener dans les conditions qui simulent la réponse d'un modèle sans pertes.

4.1 Blanchiment du spectre

La première compensation consiste à réduire la pente du spectre de la fonction de transfert du conduit vocal par application d'une préaccentuation à sa réponse impulsionnelle, représentée par h . En effet, cet affaiblissement global est dû au filtrage de la séquence pseudo-aléatoire par les tissus du cou (Pham et al., 1994). Cette fonction est de la forme:

$$Q(z) = 1 - b \cdot z^{-1} \text{ avec } |b| < 1 \quad (15)$$

Par conséquent, si l'on souhaite une préaccentuation adaptée au signal, il convient d'estimer la valeur de b . On peut remarquer que $Q(z)$ est un prédicteur d'ordre 1. Dans ce cas simple, nous avons d'après le modèle de prédiction linéaire:

$$b = R_1 / R_0 \text{ avec } R_i = \sum_n h_n \cdot h_{n-i} \quad 1 \leq i \leq p$$

4.2. Amortissement de la réponse impulsionnelle

Pour éviter des variations trop importantes d'une section à la suivante (non réalistes) et assurer une meilleure stabilité du filtre modèle, on élargit les bandes passantes des formants (Elmallawany, 1975). L'approche consiste à optimiser par prédiction linéaire le filtre modèle sur une séquence de signal, soit calculer les coefficients a_k modèle, dont la réponse impulsionnelle est de la forme :

$$D(z) = 1 - \sum_{k=1}^p a_k \cdot z^{-k} \quad (16)$$

Ensuite, en multipliant la réponse impulsionnelle $D(z)$, qui est directement fonction des a_k , par une exponentielle décroissante, on réalise un élargissement équivalent des bandes passantes. On obtient ainsi de nouveaux coefficients

$$a_k^i = a_k \cdot \exp(-c \cdot k) \quad (17)$$

La valeur de c est liée à l'élargissement B des bandes passantes des formants par la relation:

$$c = \pi \cdot B \cdot T_e \text{ (} T_e \text{ est la période d'échantillonnage).} \quad (18)$$

La meilleure adéquation est obtenue en prenant $B = 50$ Hz.

4.3 Compensation du rayonnement aux lèvres

Dans le modèle du conduit vocal à réflexion, le rayonnement aux lèvres est approché comme une connexion en parallèle sur le dernier tube du conduit vocal, d'une inductance L et d'une résistance R (Degryse, 1981). Cette impédance est équivalente, alors, à une section supplémentaire qui vient s'ajouter à la dernière section du conduit vocal.

Pour cela, pour un conduit ayant N sections, on calcule à l'aide du modèle d'inversion ($N+1$) sections et on considère que la section d'ordre ($N+1$) est équivalente à l'effet du rayonnement aux lèvres. Il suffit alors de ne considérer que les N premières sections pour avoir la forme du conduit vocal.

5. MESURE DE LA LONGUEUR DU CONDUIT VOCAL

Nous proposons un critère simple pour l'estimation du nombre N de sections du conduit vocal. Cette approche fait appel à la définition d'un intervalle de valeurs possibles, avec les bornes suivantes :

$$N_1 = 2 \cdot 12 / 35300 \cdot T_e \text{ (} l=12 \text{ cm, valeur minimale)}$$

$$N_2 = 2 \cdot 20 / 35300 \cdot T_e \text{ (} l=20 \text{ cm, valeur maximale).}$$

Pour $F_e = 10$ KHz, on a $N_1 = 7$ et $N_2 = 12$. La méthode consiste à calculer pour toutes les valeurs de l'intervalle, le critère suivant:

$$J(p) = (V_{p-1} - V_p) / V_{p-1} \quad (19)$$

Où V_p est l'énergie de l'erreur normalisée (équation 13). Le nombre de sections, N , sera égal à la valeur de p qui correspond au maximum de la fonction $J(p)$. En effet augmenter N au-delà de la valeur recherchée revient à modéliser la source d'excitation sous forme d'une source de débit, ce qui ne fait plus partie de la configuration du conduit vocal. L'erreur normalisée a alors atteint son minimum, on est donc dans le cas optimal pour la modélisation par analyse par prédiction linéaire. On rappelle que le calcul de V_p porte sur le signal issu du conduit vocal après application d'une excitation pseudo-aléatoire.

La figure (3), représente la courbe d'évolution de l'erreur normalisée V_p en fonction de p pour deux voyelles (/a/ et /u/).

Cet exemple illustre les différences qui peuvent exister entre les courbes de V_p pour différents sons. Néanmoins, la courbe part toujours de l'unité pour $p=0$ et décroît en permanence jusqu'à sa valeur minimale quand p tend vers l'infini. Chaque courbe présente plusieurs segments de transitions rapides qui signifient que l'incrément de la valeur de p d'une unité améliore sensiblement l'approximation du spectre. A titre d'exemple, le "a" connaît des transitions relatives importantes pour $p < 3$ et puis de $p=8$ à $p=10$. Pour $p > 11$, V_p est quasi-constante, ce qui tend à

signifier que $p=10$ est la valeur pour laquelle l'approximation de l'enveloppe du spectre est optimale. Ainsi, pour des valeurs de p , plus faibles, l'approximation serait plus grossière.

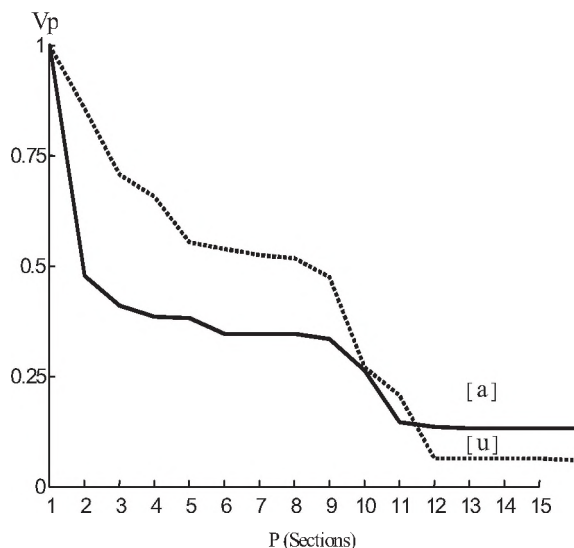


Figure 3: Courbes de l'erreur normalisée pour deux voyelles /a/ et /u/.

Alors, pour $p > 10$, le conduit vocal sera complété par un tube uniforme de grande longueur et de très faible section, ce qui ne modifie pas la fonction de transfert.

Nous donnons sur la figure (4), quelques exemples de fonctions $J(p)$ obtenues après analyse, par simulation, de quelques voyelles.

Nous remarquons que le maximum de la fonction $J(p)$ dans l'intervalle considéré permet de retrouver, exactement, le nombre de sections de la voyelle analysée.

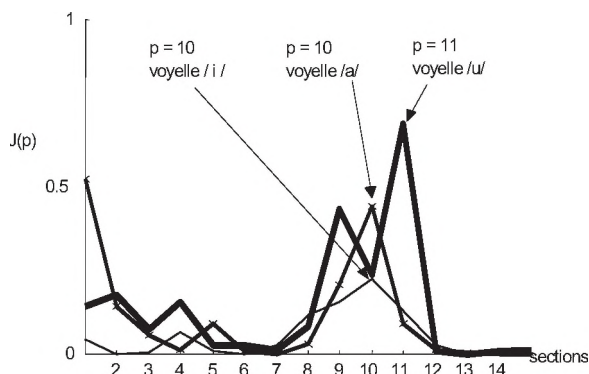


Figure 4: Courbes $J(p)$ pour les voyelles /a, i et u/.

6. APPLICATION A LA MESURE DE LA FONCTION D'AIRE DE QUELQUES VOYELLES DU FRANÇAIS

Nous présentons sur la figure 5, un exemple de résultats obtenus, par simulation, pour trois configurations du conduit vocal [a, i, u]. Nous remarquons que la méthode, après application du triple mécanisme de compensation, donne des fonctions d'aires calculées (colonne B) qui se rapprochent beaucoup des fonctions d'aire initiales. D'une part la forme générale est bien conservée, d'autre part, le calcul de la fonction de transfert à partir de la fonction estimée se superpose également assez bien avec celle de la fonction d'aire de départ.

7. EXPERIMENTATION

Le dispositif expérimental est présenté à la figure 6. La chaîne de mesure se compose de : (1) une carte de traitement de signal connectée à un micro-ordinateur, possédant des cartes de convertisseurs A/D et D/A, et qui génère un signal d'excitation pseudo-aléatoire numérique, (2) un amplificateur, (3) un exciteur, (4) un microphone avec son préamplificateur connecté à la voie A/D de la carte de traitement de signal, ainsi qu'au casque audio porté par le sujet. Une plaque en fibre de verre est utilisée afin de minimiser le rayonnement acoustique de l'excitateur vers le microphone.

Le conduit vocal est excité de manière externe au niveau du larynx ; le microphone placé à environ 2 cm des lèvres capte le signal modulé par les cavités supra-glottiques à la sortie du conduit vocal.

Pour une articulation donnée, l'opération se déroule en quatre phases :

- la première est une phase d'excitation par bruit blanc, qui aide le sujet à positionner correctement ses articulateurs en écoutant le retour ;
- le sujet maintient son articulation pendant une phase de silence ;
- excitation par le signal pseudo-aléatoire et enregistrement du signal issu des lèvres ;
- calcul de l'intercorrélation et de la fonction d'aire.

Dans le cas de mesure de la fonction d'aire en condition de phonation, la production du son commence juste avant l'application de l'excitation pseudo-aléatoire.

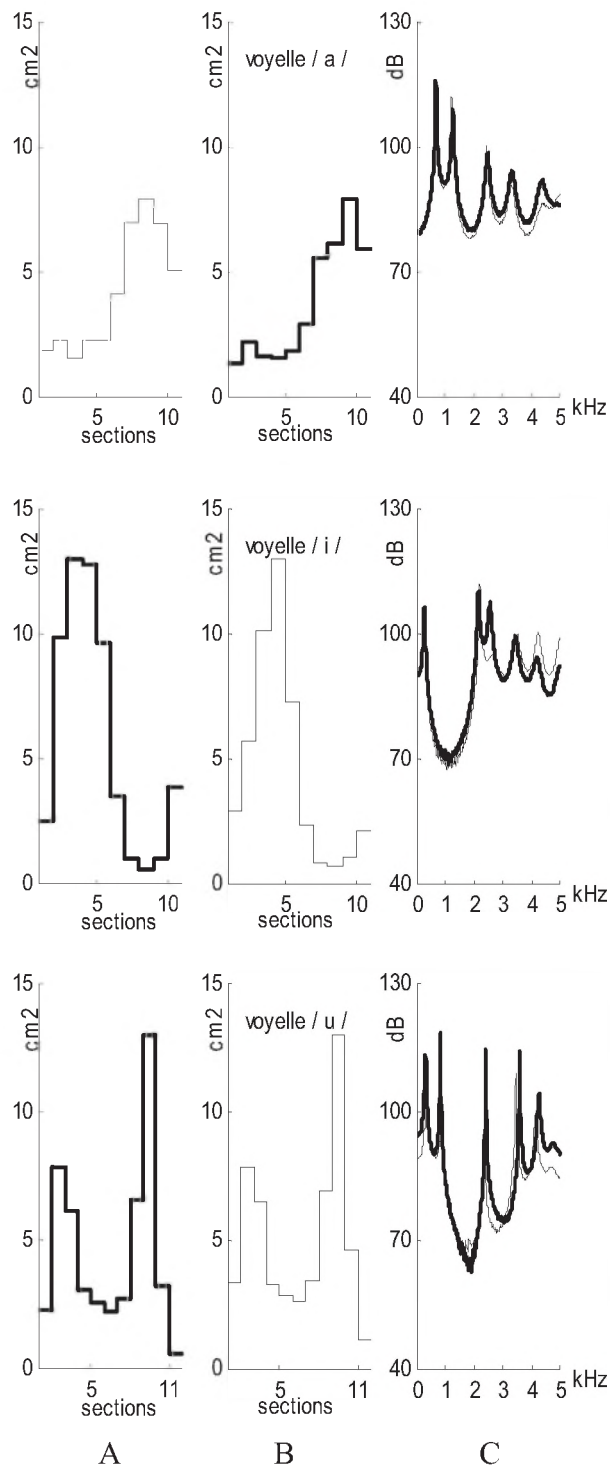


Figure 5. Résultats de simulations après compensations

A: Fonction d'aire de configuration,

B: Fonction d'aire estimée

C: Fonction de transfert de configuration en gras, calculée en fin

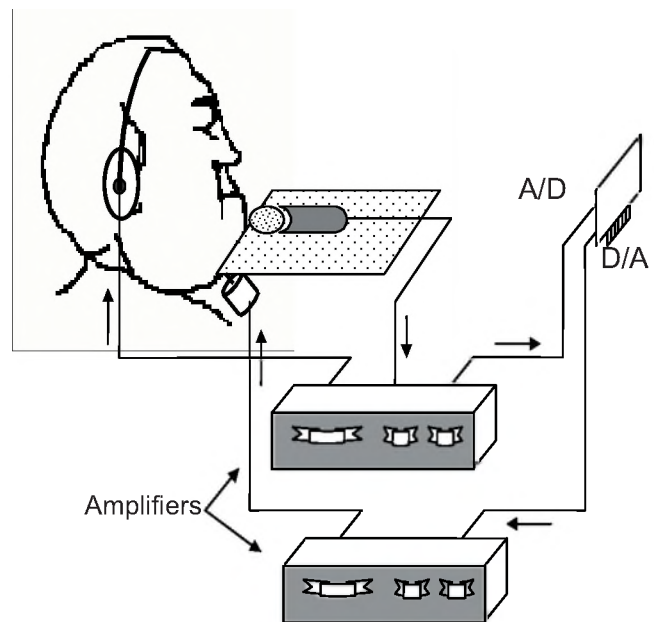


Figure 6: Dispositif expérimental (Djeradi et al, 1991).

Les fonctions de transfert du conduit vocal des onze voyelles orales du français [i, e, E, a, A, O, o, u, y, eu, oe] ont été enregistrées systématiquement par deux sujets masculins de langue maternelle française. Les enregistrements ont été effectués dans un studio isolé phonétiquement. Le signal est directement numérisé à la fréquence d'échantillonnage de 10 KHz, ce qui est suffisant pour les voyelles dans le domaine de validité du modèle de production. La durée de la mesure est d'environ 100 ms, correspondant à une séquence pseudo-aléatoire de 1023 échantillons. Chaque voyelle est enregistrée douze fois afin de vérifier la stabilité, et de différencier les pics ayant une réalité physique des pics liés à des artefacts de mesure. Deux fonctions de transfert sont mesurées successivement pour la même articulation dans deux conditions différentes : en phonation (production d'une voyelle voisée), et à glotte fermée. Pour cela, le sujet soutient l'articulation en phonation, pour la première mesure, puis ferme sa glotte sans bouger les autres articulateurs, pour permettre la deuxième mesure (Pham et al., 1994). Les fonctions d'aires sont mesurées, ensuite, par la méthode de prédiction linéaire.

Notons qu'il est extrêmement difficile de réaliser une source de débit acoustique ayant une impédance interne assez grande pour assurer que son signal de sortie serait indépendant de l'impédance acoustique le chargeant. Il est encore plus difficile d'envisager d'insérer une telle source dans le conduit vocal du sujet près de la glotte. C'est pourquoi nous sommes obligés d'exciter le conduit vocal de manière externe. Il en résulte que la fonction de transfert inconnue de la peau, des cartilages et du cou est ajoutée à la fonction de transfert du conduit vocal mesurée.

Il a été vérifié de manière expérimentale que la peau et les cartilages du sujet se comportent comme un filtre passe-bas à bande passante large, ce qui assure qu'aucun pôle ou zéro supplémentaire à bande passante étroite dans la

fonction de transfert mesurée ne peut provenir de la fonction de transfert du cou (Pham et al., 1994). Le blanchiment du spectre décrit au paragraphe 4.1 permet de réduire la pente spectrale, parasite, introduite par le filtrage de l'excitation pseudo-aléatoire par les tissus du cou.

7. RESULTATS

Pour apprécier les fonctions d'aire obtenues, nous avons utilisé les données publiées par Majid (Majid, 1986). Nous avons tout particulièrement porté notre attention sur le lieu d'articulation, l'aire correspondante et l'aperture aux lèvres. Notons que les fonctions d'aire obtenues ont été normalisées, en volume, par rapport à celles publiées par Majid, pour les voyelles du français. Nous présentons ici les résultats correspondant aux trois voyelles cardinales (/i/, /a/ et /u/), pour douze mesures successives. Ces fonctions d'aire nous ont servi pour sélectionner, pour chaque voyelle, une configuration "type".

A première vue, les fonctions d'aire que nous avons obtenus présentent des variations, mais une lecture articulatoire permet de remarquer que (figure 7):

- Pour la voyelle /i/, le lieu d'articulation (la position de la constriction) est bien situé à l'avant du conduit vocal (à 13 cm de la glotte), et l'aire aux lèvres varie dans une gamme limitée (2 à 4.5 cm²).
- Pour la voyelle /a/, le lieu d'articulation se trouve bien dans la zone pharyngale (8 cm de la glotte) et l'ouverture des lèvres est plus grande et relativement précise (4 à 5 cm²).
- Nous avons bien trouvé, pour la voyelle arrière /u/, deux constriction : l'une à l'intérieur du conduit vocal bien localisée (7 cm de la glotte) et l'autre aux niveaux des lèvres (0.5 à 1 cm²).

Ces résultats peuvent se résumer comme suit:

- Les voyelles cardinales /i/, /a/ et /u/ produites par le modèle présentent toutes les caractéristiques bien connues du lieu d'articulation et de l'aperture aux lèvres, c'est une première validation du modèle. Il y a, parfois, des différences entre les fonctions d'aire, mais pas au point de présenter des configurations ayant des lieux d'articulations différents, et des aires aux lèvres très variables.
- Dans cette génération extensive fournie par le modèle, la description vocalique classique en terme de lieu et d'aperture aux lèvres n'est donc pas prise en défaut.

Sur la colonne B de la figure 7 nous présentons les résultats obtenus en phonation. Notons que l'existence d'un signal parasite lié à une intercorrélation non rigoureusement nulle entre le signal pseudo-aléatoire et les signaux extérieurs (signal glottique, bruit ambiant...) induit des perturbations sur la fonction de transfert mesurée, par conséquent sur la fonction d'aire. Nous avons vu que l'intercorrélation entre la sortie $y(n)$ et le signal pseudo-aléatoire $x(n)$ est la somme de deux composants $R1(n)$ et

$R2(n)$: $R1$ correspond à la réponse impulsionnelle du conduit vocal, s'atténue au bout d'une certaine durée D ; $R2$ provenant d'une intercorrélation non rigoureusement nulle en pratique entre $x(n)$ et le signal parasite, $b(n)$, dont l'amplitude maximale est à peu près constante mais faible relativement au maximum de l'amplitude de la réponse impulsionnelle. Cette dernière composante n'apparaît vraiment que dans la partie où la réponse impulsionnelle est suffisamment amortie, et elle se présente sous la forme de l'apparition d'un pic de résonance à la fréquence fondamentale du signal perturbateur (Djeradi et al., 1991). Pour atténuer l'amplitude de ce pic, nous pondérons par une fenêtre de Hanning (Bellanger, 1984) l'intercorrélation entre $x(n)$ et $y(n)$. Ce traitement présente l'inconvénient de modifier aussi les valeurs des bandes passantes des résonances. Des essais ont montré que le meilleur compromis entre l'erreur introduite dans la mesure des bandes passantes et la diminution du perturbateur, sachant que la résolution spectrale est de l'ordre d'une dizaine de Hz, correspond à une fenêtre de Hanning de largeur 30 ms.

La comparaison des fonctions d'aire mesurées à glotte fermée (colonne A) avec celles mesurées en phonation (colonne B), pour les trois cas, montre que les profils sont bien retrouvés et que les constatations articulatoires restent valables.

Sur la figure 8 nous avons superposé les fonctions de transferts obtenus, colonne A, et les fonctions d'aire, colonne B, pour les deux conditions de mesure (en phonation et à glotte fermée). Ces fonctions d'aire se rapportent aux trois voyelles cardinales du français, prononcées par un locuteur masculin. Sur la colonne C, nous présentons les résultats publiés par Majid (1986), pour les mêmes voyelles. On notera l'assez bonne cohérence des lieux d'articulation, de leur degré d'aperture, de l'aire aux lèvres et de la longueur du conduit vocal.

Le dispositif mis en place, nous a permis d'extraire les fonctions d'aire et de transfert des onze voyelles du français, pour deux locuteurs masculins, dans les deux conditions d'enregistrements. Les résultats obtenus ont été jugés réalistes.

8. FORMANTS

A l'aide d'un simulateur analogue du conduit vocal et pour valider nos fonctions d'aire, nous avons calculé les valeurs des trois premiers formants des onze voyelles correspondants aux configurations obtenues, pour deux locuteurs. Sur les figures 9.a et 9.b, nous avons superposé les valeurs calculées à partir des fonctions d'aire extraites avec celles mesurées sur les fonctions de transfert. Nous remarquons que les valeurs sont assez proches, avec un léger déplacement du troisième formant pour certaines voyelles.

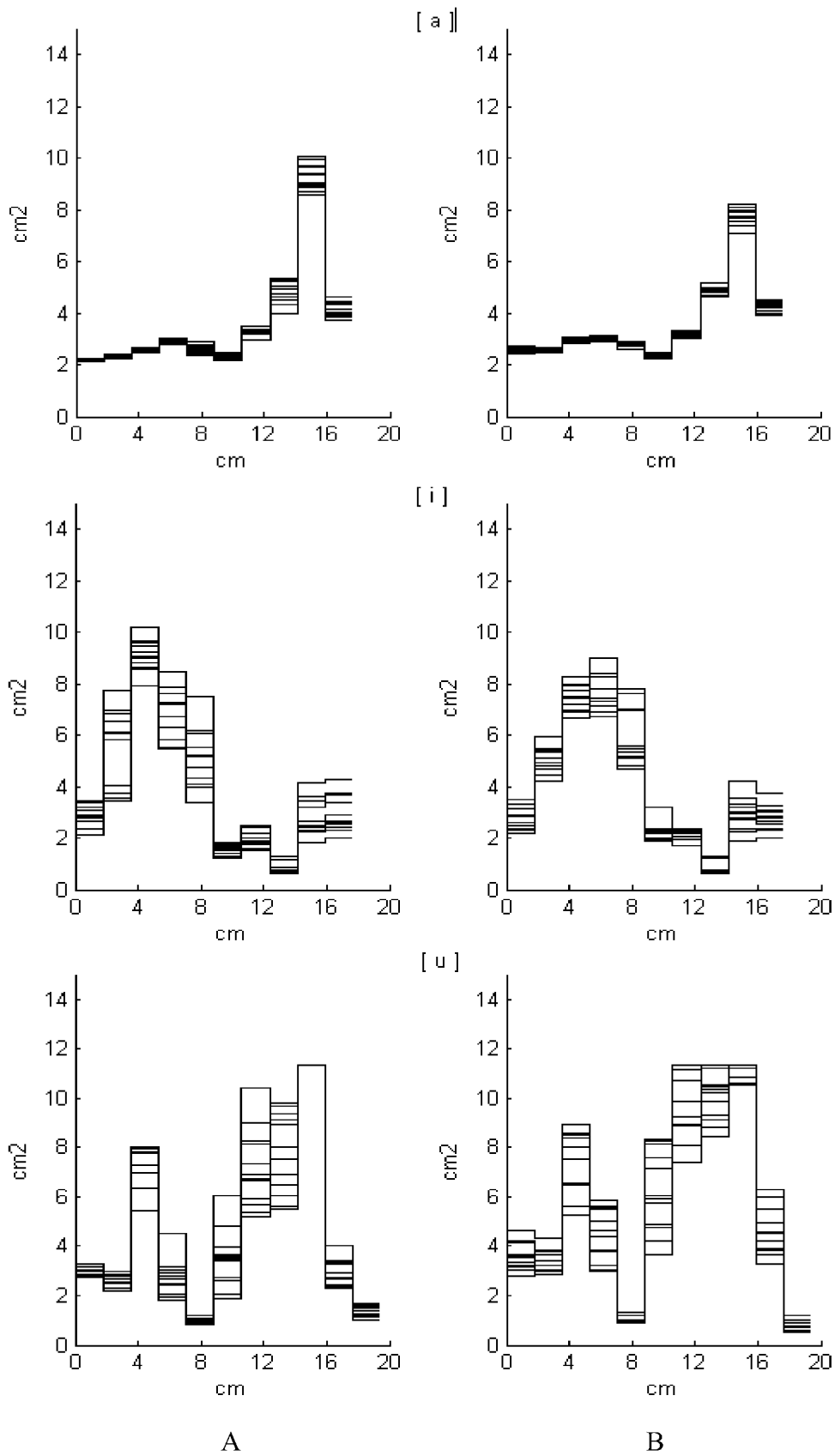


Figure 7: Fonctions d'aire des voyelles / a /, / i / et / u / A: glotte fermée; B: En phonation.

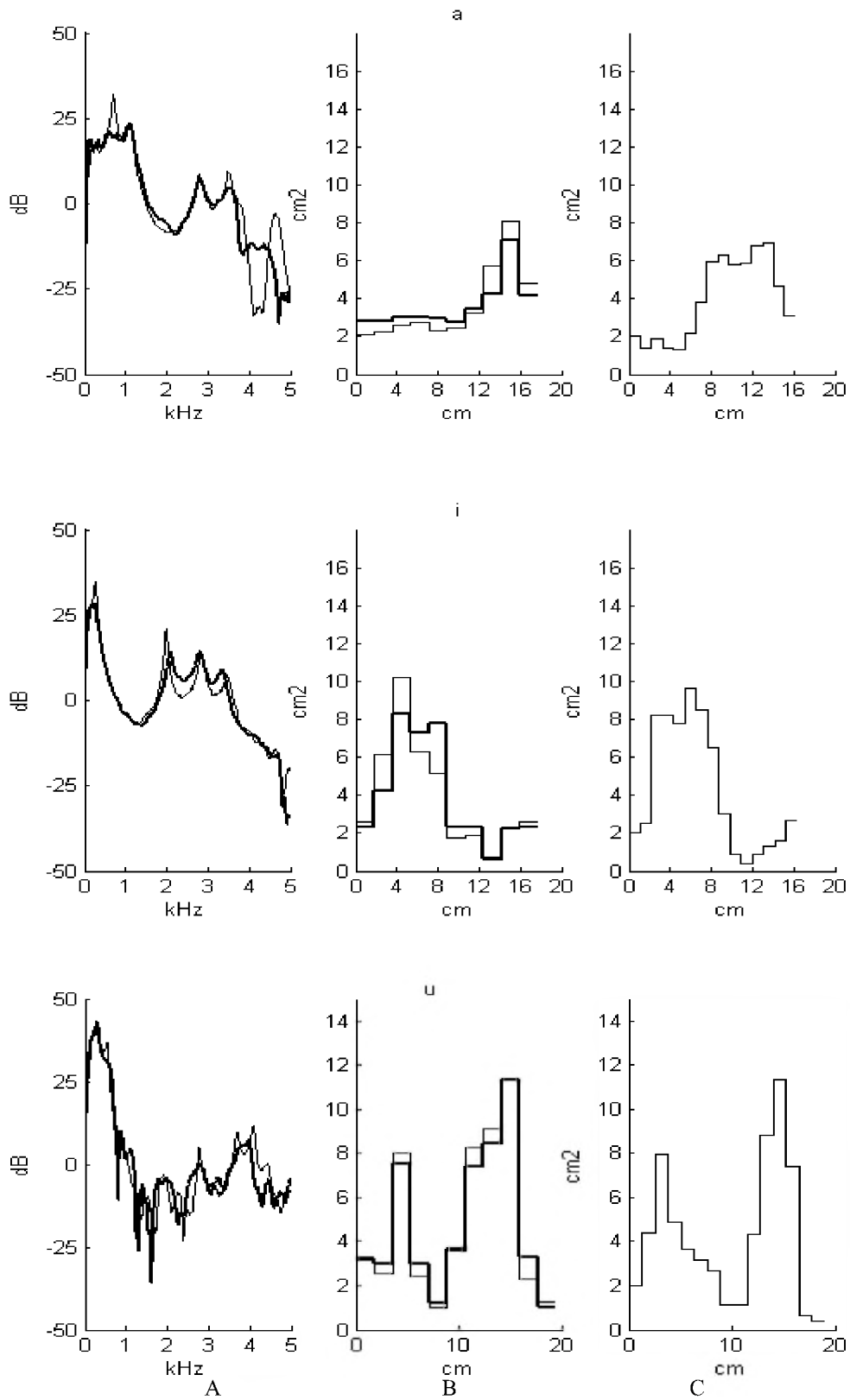
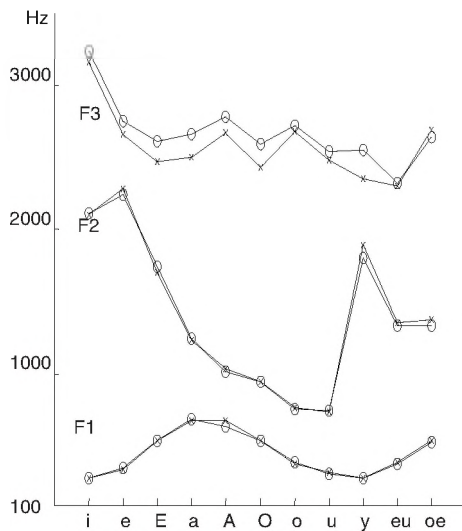
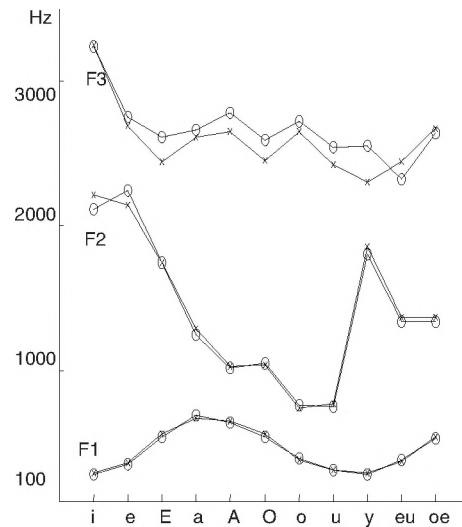


Figure 8: Fonctions de transfert et d'aire du conduit vocal pour les voyelles [a, i, u], (sujet masculin). (En trait gras : phonation, en trait fin : glotte fermée) - A: Fonctions de transfert; B : Fonctions d'aires; C: Fonctions d'aires d'après Majid (1986).



Locuteur 1



Locuteur 2

Figure 9a : Présentation des voyelles orales du français selon les valeurs des trois premiers formants. (En haut: glotte fermée ; en bas : en phonation)
 x: valeurs calculées à partir des fonctions d'aire o: valeurs mesurées sur les fonctions de transfert.

Figure 9b: Présentation des voyelles orales du français selon les valeurs des trois premiers formants. (En haut: glotte fermée; en bas: en phonation)
 x: valeurs calculées à partir des fonctions d'aire o: valeurs mesurées sur les fonctions de transfert.

9. CONCLUSIONS

Les résultats de simulation, puis de mesure sur quelques voyelles, montrent que la méthode de mesure de la fonction d'aire du conduit vocal par excitation pseudo-aléatoire est fiable.

Ainsi, nous avons montré à travers ce travail que l'extraction d'une fonction d'aire unique directement de la réponse impulsionnelle du conduit vocal était possible à l'aide d'une méthode de prédiction linéaire en simulation dans le cas sans pertes.

Dans le cas avec pertes, plusieurs traitements complémentaires sont nécessaires. Nous avons notamment étudié les différentes stratégies d'égalisation en vue de la prise en compte des influences des différentes pertes, en particulier celles par rayonnement aux lèvres.

Enfin nous avons proposé et testé un critère pour l'estimation du nombre de sections du conduit vocal. Les longueurs obtenues pour les onze voyelles du français sont valables.

Les résultats de mesure sur des cas réels montrent que les profils de fonctions d'aire obtenus pour les onze voyelles sont très satisfaisants. Les lieux de constriction et l'aperture

aux lèvres sont très proches de ceux fournis en littérature (Majid, 1986), (Yehia et al., 1996).

Par ailleurs, nous avons vérifié qu'en condition de phonation, la perturbation apportée par une intercorrélation non nulle pouvait être atténuée et que les résultats finaux restaient réalistes et exploitables.

Il est clair que le système que nous venons de développer présente des limites. Celles-ci résident dans la difficulté à positionner l'excitateur, exactement au niveau du larynx et de s'assurer que le conduit vocal est stable pendant la mesure. Par ailleurs, la technique peut être améliorée en comparant les fonctions d'aire du sujet à ceux que nous mesurerons directement par la technique d'imagerie (IRM). Cette option est difficile à réaliser mais pas impossible.

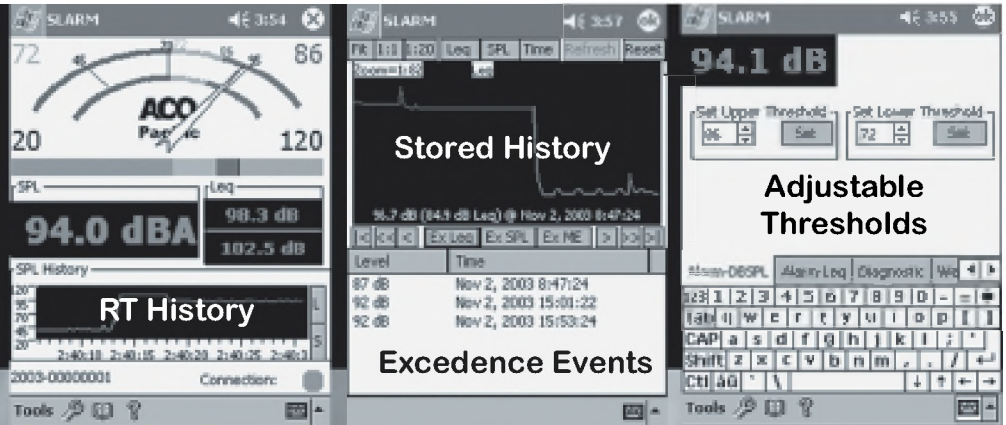
REFERENCES

- Atal, B. and Hanaeur, S. (1971). "Speech analysis and synthesis by linear prediction of speech wave," *J. Acoust. Soc. Amer.*, 50, 637-655.
- Bellanger, M. (1984). "Traitement numérique du signal," Ed. Masson, Paris.
- Castelli, E. (1989) "Acoustic considerations upon the low nasal formant based on nasopharyngeal tract transfer function measurement," in proceeding of the European conference on speech communication and technology, J.P. Tubach & J.J. Meriani editors, 2, 412-415.
- Degryse, D. (1981) "Temporal simulations of wave propagation in the lossy vocal tract", 4th FASE symposium, 193-196.
- Djeradi, A., Guerin, B. and Badin, P. (1991) "Measurement of the acoustic transfer function of the vocal tract: a fast and accurate method," *Journal of phonetics*, 19, 387-395.
- Elmallawany, I. (1975) "Etude des vocodeurs à prédiction linéaire, détermination de l'intervalle de fermeture de la glotte", docteur ingénieur thesis, INPGrenoble, France.
- Leroux, J. and Gueguen, G. (1977) "A fixed point computation of partial correlation," *IEEE Tans ASSP*, 25, 257-259.
- Levinson, N. (1947) "the wiener root mean square error criterion in filter design and prediction," *Journal of Math. Phys.*, 25, 21-278.
- Julien, J., Gilloire, A. and Saliou, A. (1972) "Mesure des réponses impulsionnelles en acoustique," *NT/LAA/TSS/181*.
- Kelly, J. and Lockbaum, C. (1962) "Speech synthesis," *Proc. Stockholm- speech communication seminar*, 193-196.
- Majid, R. (1986) "Modélisation articulaire du conduit vocal, Exploration et exploitation des fonctions de macro sensibilités paramétriques et voyelles du français," doctorat thesis, INPGrenoble, France.
- Mermelstein, P. (1967) "Determination of vocal-tract shape from measured formant frequencies," *J. Acoust. Soc. Amer.*, 41(5), 1283-1294.
- Pham thi ngoc, Y. and Badin, P. (1994) "Vocal-tract acoustic transfer function measurements: further developments and applications". *Journal de physique IV, Colloque C5, supplément au journal de physique III*, 4, 549-552.
- Schroeder, M. (1967) "Determination of the geometry of the human vocal-tract by acoustical measurements," *J. Acoust. Soc. Amer.*, 41(4), 1002-1010.
- Schroeder, M. (1979) "Integrated-impulse method measuring sound decay without using impulses," *J. Acoust. Soc. Amer.*, 66, 497-500.
- Sondhi, M. (1979) "Estimation of vocal-tract areas: the need for acoustical measurements," *IEEE Trans. Acous. Speech. Signal. Process.*, 27(3), 268-273.
- Sondhi, M. (1974) "Model for wave propagation in a lossy vocal-tract," *J. Acoust. Soc. Amer.*, 55(5), 1070-1075.
- Teffahi, H. (2000) "Vers une mesure directe de la fonction d'aire du conduit vocal," *CIMASI 2000, Casablanca, Maroc*.
- Wakita, H. (1973) "Direct estimation of the vocal tract shape by inverse filtering of the acoustic speech waveforms," *IEEE Trans Audio and Electroacoustics*, 21(5), 417-427.
- Wakita, H. (1979) "Estimation of vocal-tract shapes from acoustical analysis of the speech wave: the state of art," *IEEE Trans. Acoust. Speech. Signal. Process.*, 27(3), 281-285.
- YEHIA, H. and ITAKURA, F. (1996) "A method to combine acoustic and morphological constraints in the speech production inverse problem," *Speech communication*, 19, 151-174.

Noise Pollution

The SLARM™ Solution

PDA & Laptop
Displays
Wired
Wireless



The SLARM™ developed in response to increased emphasis on hearing conservation and comfort in the community and workplace incorporates ACOustAlert™ and ACOustAlarm™ technology. Making the SLARM™ a powerful and versatile sound monitoring/alarm system.

Typical Applications Include:

Community

- ◆ Amphitheatres
- ◆ Outdoor Events
- ◆ Nightclubs/Discos
- ◆ Churches
- ◆ Classrooms

Industrial

- ◆ Machine/Plant Noise
- ◆ Fault Detection
- ◆ Marshalling Yards
- ◆ Construction Sites
- ◆ Product Testing

FEATURES

- ✓ Wired and Wireless (opt)
- ✓ USB, Serial, and LAN(opt) Connectivity
- ✓ Remote Displays and Programming
- ✓ SPL, Leq, Thresholds, Alert and Alarm
- ✓ Filters (A,C,Z), Thresholds, Calibration
- ✓ Multiple Profiles (opt)
- ✓ 100 dB Display Range:
- ✓ 20-120 dBSPL and 40-140 dBSPL
- ✓ Real-time Clock/Calendar
- ✓ Internal Storage: 10+days @1/sec
- ✓ Remote Storage of 1/8 second events
- ✓ 7052S Type 1.5™ Titanium Measurement Mic



2604 Read Ave., Belmont, CA 94002 Tel: 650-595-8588 FAX: 650-591-2891
www.acopacific.com acopac@acopacific.com

ACOustics Begins With ACO™