

UN NOUVEL ALGORITHME A CONVERGENCE DE POLE OFFRANT
UNE MEILLEURE IMMUNITÉ AU BRUIT DES SPECTRES LPC

G. Duncan et M.A. Jack

Centre for Speech Technology Research, University of Edinburgh,
80 South Bridge, Edimbourg EH1 1HN, Ecosse

RESUME

L'analyse spectrale par prédiction linéaire, bien que communément employée et fournissant un excellent modèle du conduit vocal, est traditionnellement peu performante et dans la séparation de résonances proches en fréquence et dans les milieux bruités.

On présente une nouvelle méthode LPC d'analyse qui est beaucoup moins affectée par le bruit et offre une meilleure résolution d'attributs spectraux du signal. De plus, la technique à convergence de pôle demande des paramètres d'analyse qui pourraient dégrader l'analyse LPC classique, et ne nécessite aucune préaccentuation du signal vocal.

SUMMARY

Spectral analysis of speech by linear prediction, although in widespread use and furnishing an excellent model of the human vocal tract, nonetheless performs poorly both in its ability to adequately separate spectral resonances which are close in frequency, and in the analysis of noisy signals.

A novel method of LPC-based spectral analysis is presented here which offers a substantial improvement in signal-to-noise performance and also provides enhanced resolution in the extraction of spectral features from the signal. In contrast to accepted standards, the pole focussing technique requires analysis parameters which would otherwise be considered to degrade the performance of LPC spectral estimation. Additionally, the technique does not require any prior preemphasis of the speech signal.

1 - INTRODUCTION

L'extraction et la mesure de détails spectraux par la modélisation autorégressive ont des applications particulièrement importantes dans les domaines du traitement de la parole ainsi que de l'analyse sismique. Par exemple, dans la reconnaissance de la parole par l'analyse d'attributs phonétiques, la mesure des résonances (formants) dans le spectre d'amplitude à court temps de la parole, ainsi que la capacité de suivre la trajectoire évolutive de ces attributs, sont de grande importance pour la reconnaissance des sons voisins. Le calcul des trajectoires des formants fournit également des renseignements acoustiques secondaires en ce qui concerne l'environnement segmentaire, et donc aide à identifier des sons consonantiques et avant et après la vocalisation.

La nouvelle technique exposée dans cet article est une méthode de mesure des formants de spectres obtenus par l'analyse autorégressive. En se référant à l'analyse par prédiction linéaire (LPC) classique, on trouve que la technique améliore la détection et la séparation de formants qui sinon se confondraient dans le plan temps-fréquence, et offre une meilleure performance dans les milieux bruités.

La technique comprend le calcul de plusieurs spectres à rayons réduits lié à un accroissement progressif de l'ordre du modèle autorégressif, améliorant ainsi l'estimation de la fonction transferte du système. Contrairement à l'analyse LPC classique, on trouve que la préaccentuation empire la performance de la technique à convergence de pôle. L'étape finale de l'algorithme consiste d'un filtre à pondération d'attributs, qui est appliqué aux spectres accumulés dans chaque tranche d'analyse, et qui aide directement à fournir une forte immunité au bruit au processus d'estimation des fréquences formantiques.

On trouve que la technique à convergence de pôle, comparée à l'estimation par LPC classique, offre une augmentation de plus de 10dB dans l'immunité au bruit dans son application à l'estimation des fréquences des

formants. Suite à cette propriété, on trouve que la technique améliore également la performance des méthodes d'estimation de trajectoires évolutives des formants, de telle sorte qu'on peut donc obtenir une amélioration supérieure à 15dB dans la cohérence de ces trajectoires. Finalement, la technique à convergence de pôle est relativement insensible au choix de l'ordre du modèle LPC, ce qui fondamentalement mène à un algorithme d'analyse qui est indépendant du locuteur. On démontre les propriétés mentionnées ci-dessus de la nouvelle technique en l'appliquant à une voyelle extraite de la parole continue, en comparaison avec une autre méthode LPC avancée d'analyse des formants.

2 - LA TECHNIQUE A CONVERGENCE DE POLE

Dans l'analyse de la parole par la méthode LPC, on représente le conduit vocal par une chaîne de tubes acoustiques de la glotte aux lèvres, le nombre de tubes étant lié directement au choix de l'ordre du modèle LPC. On assume que l'onde acoustique qui excite ces tubes est soit due à la vibration périodique des cordes vocales (train d'impulsions) soit à une constriction quelquepart dans le conduit vocal créant un courant d'air turbulent (bruit blanc). En général, le choix de l'ordre du modèle a un effet critique sur la performance de l'estimation du spectre, et donc d'habitude on choisit un ordre entre $p=12$ et $p=16$ inclusif: avec moins de coefficients, l'analyse LPC ne réussit pas à séparer les formants qui se rapprochent dans le plan temps-fréquence; avec plus de coefficients, la performance de l'estimation du spectre est dégradée dans les milieux bruités à cause des pics aléatoires qui, dûs au bruit, apparaissent dans le spectre LPC.

La technique à convergence de pôle évite la confusion et la perte de formants de faible intensité en utilisant le calcul de plusieurs spectres à rayons réduits. C'est-à-dire que, pour une seule tranche d'analyse, plusieurs spectres sont calculés dans le plan complexe des z en suivant des chemins tel que $|z| \leq 1$. Des candidats pour les formants sont extraits de chacun de ces spectres à rayons réduits en utilis-



nt des méthodes bien connues de l'interpolation parabolique. Cependant, les améliorations offertes par la technique sont fournies par l'accroissement progressif de l'ordre du modèle autorégressif avec chaque réduction du rayon du spectre, de telle sorte qu'il est de plus-en-plus probable que le modèle du conduit vocal contiendra des termes représentant des formants faibles. La technique réussit donc bien à séparer des formants qui sinon se confondraient, et d'améliorer la détection de formants qui sont traditionnellement extrêmement durs à obtenir, tel que les formants nasaux. On peut considérer également que l'accroissement progressif de l'ordre du modèle aide directement à fournir une forte immunité au bruit au processus pourvu qu'on applique en même temps un simple filtre à pondération d'attributs. L'objectif de cette étape finale est de faire ressortir les attributs spectraux qui sont restés consistants avec chaque changement d'ordre, et d'atténuer les candidats qui ont eu un comportement non stationnaire avec chaque changement de rayon et d'ordre du modèle.

On peut expliquer le comportement du processus de convergence de pôle en examinant le résultat bien connu dans l'étude des systèmes résonants à deux pôles qui indique que le moins le facteur d'amortissement, γ , le moins est la largeur de bande du pic qui caractérise la réponse fréquentielle du système. De plus, la fréquence centrale de la résonance, f_m , se déplace en fréquence vers la fréquence naturelle du système, f_n , au fur et à mesure que le facteur d'amortissement décroît:

$$f_m = f_n \sqrt{1 - 2\gamma^2} \quad (1)$$

Par analogie, on peut dire que si le chemin, à travers lequel on calcul la réponse fréquentielle du système, dévie de son chemin habituel à un rayon dans le plan complexe des z de $|z|=1$ de sorte qu'il s'approche aux points singuliers qui caractérise le système, les pics associés à ces pôles dans le spectre à rayon réduit sembleront appartenir à un système qui possède un facteur d'amortissement qui est plus faible. C'est-à-dire que leur largeur de bande diminuera et au même temps leur fréquence centrale accroîtra. Cependant, dans les systèmes complexes caractérisés par plusieurs paires de pôles, il est peu probable que la variation de la fréquence centrale d'un pic quelconque soit monotone au fur et à mesure que le chemin de calcul du spectre s'approche du point singulier associé au pic. En général, on peut s'attendre à ce que la fréquence centrale d'un pic obéisse à la formule d'équation (1) qu'à des largeurs de bandes très petites, quand la paire de pôles associés au pic domineront effectivement la réponse fréquentielle et donc peut être considéré découplé du reste du système.

3 - RESULTATS ET PERFORMANCE COMPARATIVE

Afin d'évaluer la performance de la technique à convergence de pôle, on l'applique à un segment de la parole continue - la voyelle /i/ du mot anglais "deed" - énoncé par un interlocuteur masculin. Un spectrogramme à bande large du segment est indiqué en figure 1(a). Pour obtenir des résultats comparatifs, on emploie une technique qui est également basée sur l'analyse LPC et qui comprend le calcul de plusieurs spectres à rayons réduits [1], mais qui ne cherche ni à augmenter l'ordre du modèle autorégressif ni à appliquer un filtre à pondération aux spectres. Dans cette dernière technique, un modèle fixe est choisi (ici on utilise $p=16$) et employé pour n'importe quel valeur du rayon dans le plan complexe des z . Cependant, contrairement à la technique à convergence de pôle, le calcul de spectres à rayons réduits ne se déroule pas automatiquement dans chaque tranche d'analyse, n'étant invoqué que s'il manque au moins un formant dans le spectre calculé à travers le chemin $|z|=1$ d'une tranche d'analyse à une autre. Dans ce cas, la technique tente de retrouver le formant manquant en supposant que celui-ci soit profondément enfoncé à l'intérieur du plan complexe des z . L'algorithme peut ensuite calculer jusqu'à 25 spectres

à des rayons à partir de $r = |z| = 0.98$ jusqu'à $r = 0.88$, décroissant à chaque étape de $\delta r = -0.004$. Les valeurs extraites pour les formants présents à $r = 1$ ne sont pas affectés par ces opérations. L'algorithme demande que le signal subisse une préaccentuation avant d'effectuer l'analyse.

Dans la technique à convergence de pôle, le calcul de spectres à rayons réduits se déroule automatiquement dans chaque tranche d'analyse de $0.9 \leq r \leq 1$, décroissant à chaque étape de $\delta r = -0.01$, c'est-à-dire qu'il y aura 11 spectres pour chaque tranche d'analyse de 25.6ms. L'algorithme ne demande pas que le signal soit préaccentué (et en effet un tel prétraitement empire sa performance), mais par contre chaque réduction du rayon de calcul de la transformation des z doit être accompagnée d'une augmentation de l'ordre du modèle LPC. Cependant, l'algorithme paraît peu sensible au choix de l'ordre initial sur le cercle unité, et des valeurs entre $p=8$ et $p=12$ produiront des résultats très proches à ceux présentés ici. L'accroissement de l'ordre du modèle, δp , avec chaque réduction de rayon doit être $\delta p = +2$ afin d'introduire la possibilité que la technique réussisse à modéliser encore un formant dans chaque spectre. Tous les pics sont extraits de chaque spectre, et ensuite on applique une pondération à la fréquence centrale de chaque pic selon sa largeur de bande. La sommation de ces valeurs par un filtre balayant à bande étroite réduit le taux de pics à quelques candidats qui pourraient être considérés comme les formants.

Une comparaison de la performance de ces deux algorithmes est présentée dans les figures 1(b) et (c). Les deux techniques partagent exactement la même méthode de calcul des trajectoires formantiques selon celle décrite dans la référence [1]. Donc les différences de performance présentées sont dues uniquement aux algorithmes d'analyse spectrale employés dans chaque technique. Le rapport signal-à-bruit (S/B) correspond à celui du processus de quantification, et a été estimé à $S/B = 60\text{dB}$. On constate bien que la technique à convergence de pôle a bien réussi à séparer les formants F_2 et F_3 qui se rapprochent vers la fin de la voyelle, et a aussi réussi à fournir des trajectoires continues pour F_4 et F_5 . Vu que les techniques utilisent un régime similaire de recalcul de spectres à rayons réduits, on peut donc conclure que la meilleure performance de la technique à convergence de pôle est due directement à l'accroissement de l'ordre du modèle LPC employé dans le calcul de chaque spectre. Etant donné que la classification détaillée de sons voisés nécessite au moins les trajectoires continues pour F_1 , F_2 et F_3 , ce résultat est donc d'importance pour des systèmes de reconnaissance de la parole par l'analyse de traits phonétiques.

Afin d'étudier la performance comparative en ce qui concerne le rapport signal-à-bruit, le signal était contaminé par du bruit Gaussien de différentes intensités. Les figures 1(d) et (e) démontrent les résultats comparatifs obtenus par les deux techniques pour un $S/B = 15\text{dB}$. Bien que les estimations de formants pour chaque technique aient été affectées par l'intensité du bruit, le caractère général des trajectoires a été mieux conservé par la technique à convergence de pôle, surtout dans le cas de F_1 et F_2 . L'intégrité de la trajectoire pour F_2 a été totalement conservée, ce qui est particulièrement intéressant car les transitions de F_2 dans des environnements consonantiques peuvent aider à identifier la consonne elle-même. On a trouvé que l'intégrité de F_2 pour la technique des figures 1(c) et (e) est interrompue à un $S/B = 29\text{dB}$, tandis que celle associée avec la technique à convergence de pôle subit un S/B de 13dB , offrant donc une augmentation de l'immunité au bruit de 17dB . On peut considérer que cette amélioration de performance dans les milieux bruités est due à l'accroissement de l'ordre du modèle et surtout à l'application du filtre à pondération d'attributs. Il est connu que l'analyse LPC modèle d'abord les attributs les plus puissants du spectre. Plus l'ordre du modèle



est grand, le plus d'attributs faibles peuvent être caractérisés par le spectre, mais en utilisant un seul spectre il est difficile de distinguer les attributs spectraux dus au signal de ceux dus au bruit. En augmentant l'ordre du modèle par la récursion de Levinson-Durban, il est probable que les pics dus au bruit Gaussien occuperont des positions différentes et aléatoires dans la réponse fréquentielle de chaque modèle, tandis que les attributs du signal auront une tendance à rester dans une localité restreinte suivant des chemins définis par l'équation (1). Ceci entraîne l'assomption que la fonction transferte associée au signal peut être considérée stationnaire pendant la durée de la fenêtre d'analyse. L'application d'un filtre à pondération moyenne aux attributs renforcera donc ces attributs qui sont relativement fixes de spectre-en-spectre, et atténuera ceux qui démontrent un comportement non stationnaire, une propriété qui aide à restituer même les attributs faibles en admettant qu'ils fassent partie de la structure du signal. En effet, bien qu'on n'en présente pas les résultats ici, la technique à convergence de pôle a fait ses preuves dans l'extraction de formants nasaux [2], un aspect de performance qui est traditionnellement difficile pour l'analyse de la parole par les méthodes LPC.

4 - CONCLUSIONS

On a présenté dans cet article une nouvelle méthode d'estimation de formants du signal vocal qui est basée sur l'analyse spectrale par prédiction linéaire. La détection de formants est améliorée par l'augmentation de l'effet des pôles de la fonction transferte du conduit vocal sur le spectre. L'utilisation d'un ordre croissant du modèle LPC au fur et au mesure que le rayon de calcul du spectre décroît dans le plan complexe des z aide à séparer des formants proches en fréquence. Le filtrage des résultats par un simple filtre à pondération moyenne fait ressortir ces attributs du signal à court temps qui sont stationnaires et atténue les attributs non stationnaires du spectre dus au bruit à bande large.

On trouve que la technique à convergence de pôle offre une hausse de l'immunité au bruit de l'ordre de 10 à 15dB dans la performance de l'analyse spectrale, ce qui mène à une augmentation égale ou supérieure dans la performance de la méthode de calcul des trajectoires des formants.

Enfin, la structure de l'algorithme à convergence de pôle fournit une insensibilité inhérente en ce qui concerne le choix de l'ordre du modèle LPC, par rapport à l'analyse LPC classique. On peut considérer que suite à cette propriété, la performance de la technique sera moins affectée par les variations des caractéristiques acoustiques de différents interlocuteurs.

REMERCIEMENTS

La présente étude a été réalisée sous contrat de la Science and Engineering Research Council du Royaume Uni.

BIBLIOGRAPHIE

- 1 S.S. McCandless, "An algorithm for formant extraction using linear prediction spectra", *I.E.E. Transactions*, v. ASSP-22, pp.135 - 141 (Avril 1974)
- 2 G. Duncan and M.A. Jack, "An improved algorithm based on pole enhancement for estimation of the vocal tract frequency response", *Electronics Letters*, v.22, n.23, pp.1213 - 1214 (Nov. 1986)

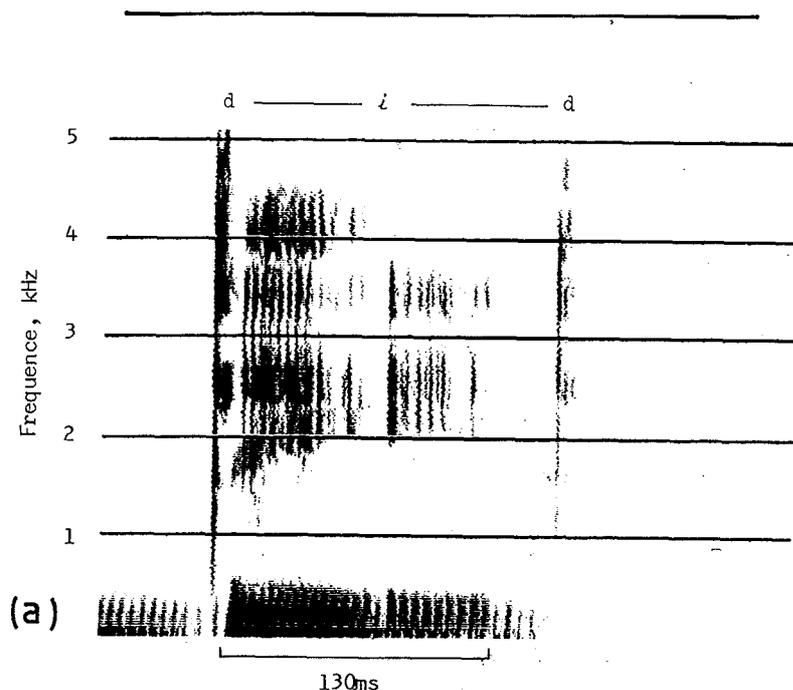
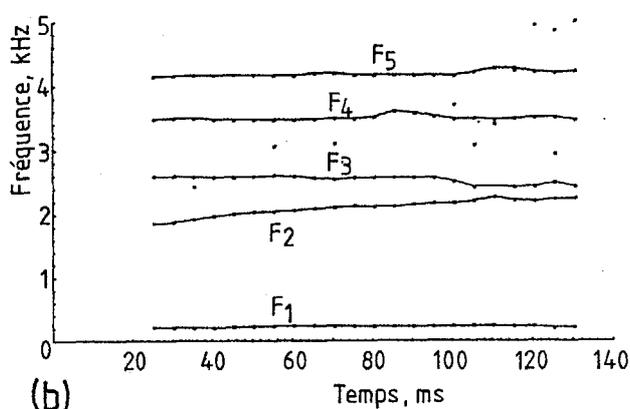
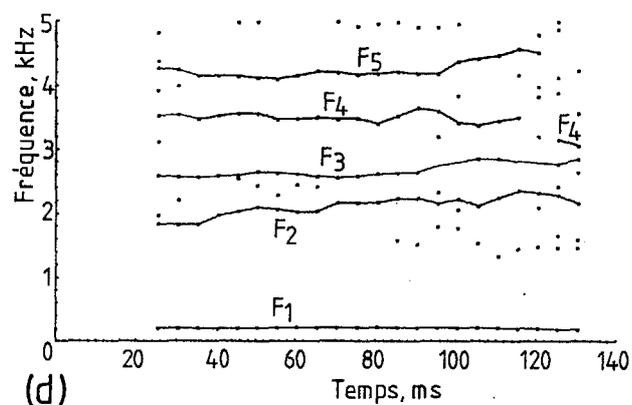


Fig.1 (a) Sonagramme du mot anglais "deed" énoncé par un interlocuteur masculin. Ceci peut être considéré comme référence de laquelle on peut juger la performance des deux techniques d'estimation des formants employées dans les figures 1(b)-(e).



S/B = 60dB



S/B = 15dB

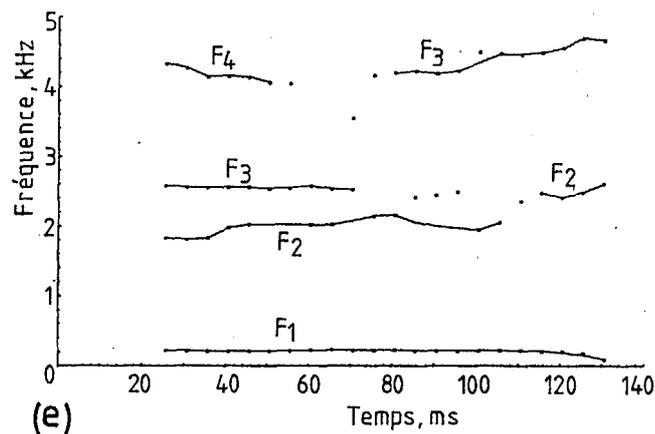
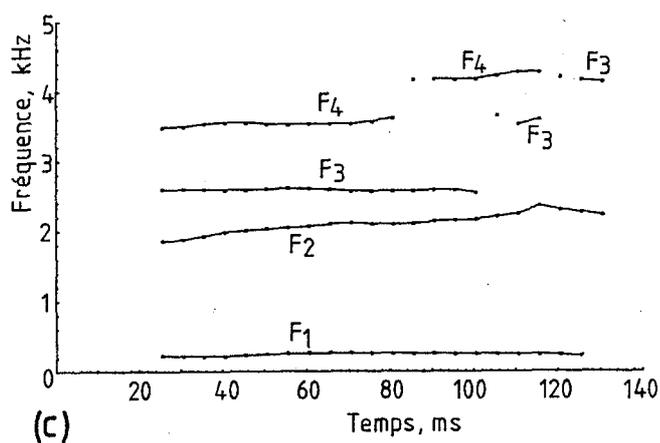


Fig.1 (b) Technique à convergence de pôle appliquée à l'analyse de la voyelle /i/ de la figure 1(a). $\delta r = -0.01$; $\delta p = +2$, $1 \leq r \leq 0.9$; $p(r=1) = 12$.

(c) Technique de la référence [1] appliquée au même segment voisé. $\delta r = -0.004$, $0.98 \leq r \leq 0.88$; $\delta p = 0$, $p = 16$ pour chaque spectre à rayon réduit.

Les deux techniques d'analyse spectrale partagent la même méthode de calcul des trajectoires évolutives des formants.

(d) - (e) Ces figures correspondent aux mêmes techniques employées dans les figures (b) - (c), mais avec un rapport signal-à-bruit du segment de 15dB.