JUSTIFICATION PERCEPTIVE DU SPECTROGRAPHE AUDITIF

Christophe d'Alessandro & Denis Beautemps

LIMSI-CNRS BP133-91403 Orsay Cédex, France.
ICP-INPG 46, avenue Félix-Viallet 38031 Grenoble Cédex, France.

ABSTRACT

An auditory spectrograph is presented and discussed, which is quite different from the initial proposal of [4]. A set of descriptive acoustic parameters are derived from speech signals analysed according to the time and frequency resolution caracteristics of the spectrograph: sinusoids in the area of F0 and F1, dominant frequencies and envelope modulation above about 1 kHz, for voiced speech. These parameters are used in an analysis/synthesis system which delivers a synthetic signal perceptively equivalent to the original signal. This preliminary work demonstrates the possibility of using alternative auditory-based acoustic parameters for speech synthesis and analysis instead of production-based acoustic parameters.

1 Introduction

L'avènement du spectrographe a permis un développement considérable de l'étude phonético-acoustique descriptive de la parole. Deux raisons conceptuellement distinctes ont contribué à ce succès: 1. le spectrographe permet l'observation de corrélats acoustiques importants du mécanisme de production de la parole, comme les formants, la vibration des cordes vocales; 2. il existe une analogie entre analyse spectrale à court terme et analyse du signal par le système auditif périphérique: les objets visibles sur un spectrogramme correspondent à des caractéristiques acoustiques perceptivement pertinentes. Les relations entre ces deux aspects ne sont pas toujours clairement considérées, et des notions qui relèvent du modèle acoustique de production, comme les formants ou l'onde de débit glottique s'avèrent d'une importance variable, voire contestable, du point de vue perceptif [3]. Nous pensons ainsi que le spectrographe a contribué à exagérer l'importance de paramètres acoustiques qui ne sont pas perceptivement pertinents.

La modélisation de l'analyse du signal par le système auditif périphérique a conduit à proposer de nouvelles représentations assimilables à la représentation spectrographique, sous forme de "spectrogrammes auditifs" [4], ou le "cochléogrammes" [5]. Le but de cet article est de discuter d'une forme de spectrographe auditif, et de justifier par une évaluation perceptive, en utilisant la resynthèse, les paramètres acoustiques auditivement pertinents qui apparaissent sur cette représentation.

2 Spectrographe auditif

Le spectrographe utilisé est présenté dans cette section, et confronté aux formes de spectrographes auditifs proposées précédemment. Les "spectrogrammes auditifs" de [4] combinent deux types d'informations spectrales: 1. l'énergie spectrale à la sortie d'un banc de filtres passe-bandes s'appuyant sur des échelles fréquencielles (Bark) et d'intensité (phon) auditives; 2. les fréquences dominantes dans chaque canal d'analyse, issues du modèle d'analyse temporelle DOMIN. D'autres auteurs [5], proposent comme représentation spectrographique la visualisation des signaux en sortie d'un banc de filtres auditifs. C'est une variante de cette seconde solution que nous avons adopté pour les figures 1 et 3. Ce tracé, équivalent au redressement simple alternance de signaux filtrés par un banc de filtres auditifs, est obtenu en portant le produit de l'amplitude par la phase principale (entre $-\pi$ et π), pour les phases positives seulement, d'une analyse par ondelettes sur une échelle auditive [1]. 250 filtres sont régulièrement répartis en échelle Bark, de largeur de bande constante 1 bark. Une échelle logarithmique est employée pour porter les amplitudes. La différence de lisibilité visuelle avec une échelle auditive Phon apparaît tout à fait négligeable. Nous avons préféré ce type de tracé à celui proposé dans [4] pour deux raisons:

1. Une erreur d'interprétation de la dualité temps/fréquence semble à la base du procédé de calcul utilisé pour les figures de [4]. Les spectrogrammes semblent (le procédé de calcul n'est pas explicite dans l'article) calculés par transformée de Fourier à court terme et visualisés en utilisant les échelles Bark et Phon. Alors seule la résolution spectrale est celle d'une analyse en échelle Bark, et augmente en raison de la fréquence centrale d'analyse. Par contre la résolution temporelle est manifestement fixe sur les spectrogrammes publiés, égale à celle d'une analyse spectrographique classique en bande large. 2. Il n'apparaît pas nécessaire d'introduire un modèle supplémentaire, comme DOMIN, pour rendre compte qualitativement (visuellement) de l'analyse temporelle: la visualisation de la phase d'analyse dans chaque bande permet de distinguer les fréquences dominantes, temporellement dans le grave du spectre, et au delà d'environ 1-1.5 kHz grâce à l'amplitude. Il faut ajouter que la resynthèse ou la modification du signal à partir de notre représentation est directe, ce qui est montré dans [1], mais ce qui est hors du propos de cette communication.

3 Justification perceptive

La lecture des spectrogrammes auditifs suggèrent un ensemble de paramètres acoustiques descriptifs, liés à la fois à l'appareil phonatoire et à la résolution spectro-temporelle du dispositif d'analyse utilisé. Les relations entre la résolution d'analyse, c'est-à-dire la largeur de bande effective des filtres auditifs, et les grandeurs fréquencielles types produites par l'appareil phonatoire sont résumées figure 4. L'abcisse représente les fréquence centrales d'analyse (en Bark), l'ordonnée les bandes passantes des filtres auditifs (en Hz) correspondant. Les deux courbes partagent le plan en deux zones: dans la zone intérieure aux deux courbes, deux fréquences pures présentes pour une fréquence centrale donnée ne sont pas distinguées par le filtre d'analyse; dans la zone extérieure deux fréquences pures sont distinguées par le filtre d'analyse. Si l'on applique un spectre harmonique, comme pour de la parole voisée, comportant un ensemble de raies spectrales équidistantes la courbe de résolution prédit le nombre de raies séparées, en fonction de la fréquence d'analyse. Deux situation types se manifestent: dans le cas A de la figure 4 (F0=100 Hz, fréquence d'analyse 200 Hz) le filtre auditif isole une composante spectrale et des sinusoïdes redressées en simple alternance sont représentées sur le spectrogramme; dans le cas B (F0=100 Hz, fréquence d'analyse 2000 hz), plusieurs raies sont intégrées par un même filtre. Dans ce cas, des battements sont visibles sur le spectrogramme. La période de la modulation d'amplitude du signal filtré, des battements, est l'inverse de la différence de fréquence entre les composantes, soit la période fondamentale 1/F0. Les battements donnent naissance à un signal possédant une fréquence dominante, obtenue par la moyenne les fréquences des raies spectrales pondérées par leurs amplitudes. Ainsi, lorsqu'un pic spectral (ou formant) est présent dans la bande d'analyse du filtre, la fréquence dominante due aux battements est approximativement égale à la fréquence centrale de ce formant. Une première dimension fréquencielle relève de la fréquence fondamentale. Une seconde dimension est donnée par l'espacement des formants. Lorsque la résolution des filtres d'analyse est plus fine que l'espacement entre deux formants, la séparation des harmoniques ou les battements formantiques se produisent: c'est la situation rencontrée respectivement pour F1, cas C et F2, cas D de la figure 4. Lorsque cette résolution diminue, ou lorsque plusieurs formants sont proches, les battement deviennent plus complexe, et une masse spectrale apparaît sur le spectrogramme. La modulation d'amplitude est plus rapide que la fréquence fondamentale, et la fréquence dominante peu saillante. Pour illustrer ce propos, les figures 1 et 3 présentent des spectrogrammes auditifs d'une voix féminine prononçant /wiski/ et d'une voix masculine prononçant /lopotiʃa/. Si l'on considère les parties voisées, pour les deux exemples, le grave du spectre est décomposé en harmoniques. Les amplitudes et phases de ces harmoniques dépendent d'une part de la source de voisement et d'autre part de l'influence du premier formant. Dans la région du second formant, l'intégration de plusieurs harmoniques dans un même filtre auditif provoque l'apparition de battements, avec une fréquence dominante et une modulation d'amplitude à la période fondamentale. Au delà d'environ 3 kHz pour la figure 1, deux formants s'agglomèrent et la fréquence dominante comme la période de modulation devient plus difficile à définir. Il est probable que cette masse spectrale est perçue par son centre de gravité, et que sa contribution se limite à des aspects non-linguistiques du signal, comme la brillance, le degré de souffle etc. La validité de ces observations peut être testée par un procédé d'analyse/synthèse. Un signal naturel est décomposé en utilisant les paramètres acoustiques précédents:

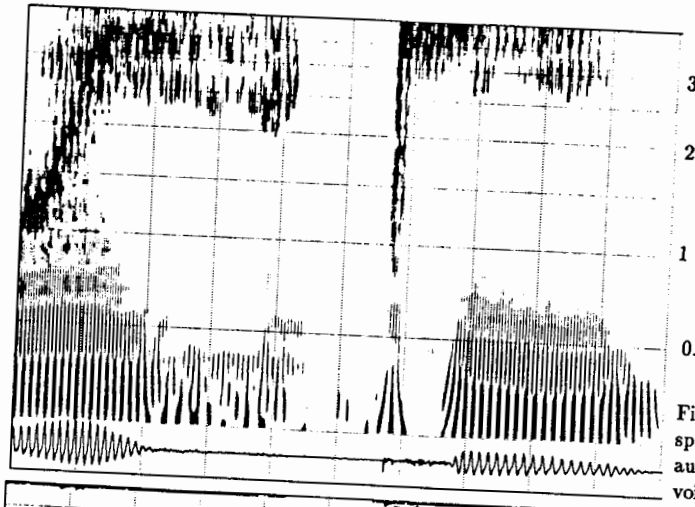

Figure 1: spectrogramme auditif /wiski/ voix féminine

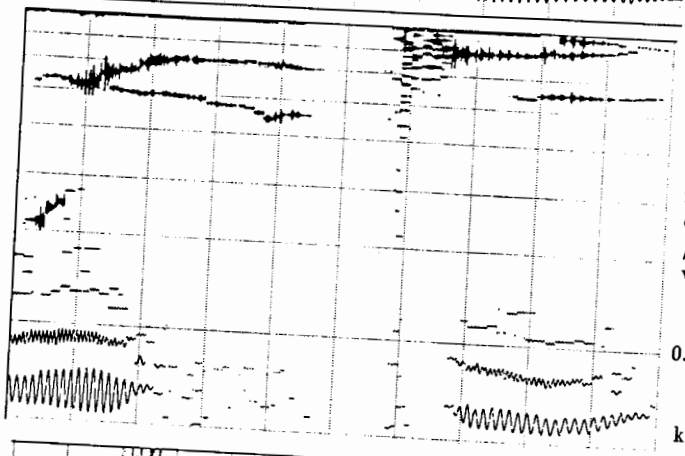

Figure 2: formes d'ondes de synthèse /wiski/ voix féminine

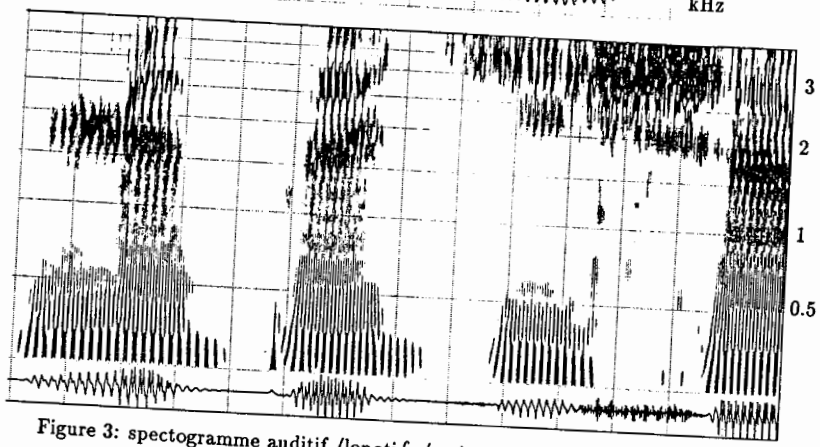

Figure 3: spectogramme auditif /lopotiʃa/ voix masculine

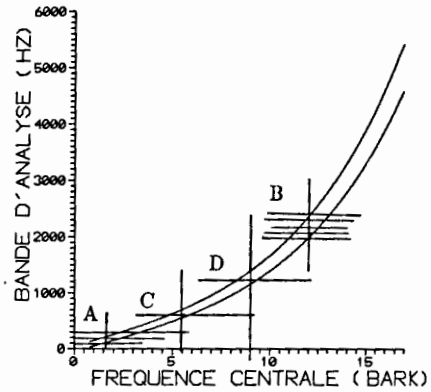

Figure 4: résolution des filtres auditifs

des sinusoïdes (amplitudes et phases) pour le grave du spectre, dans la région du premier formant et en dessous; les fréquences dominantes et la modulation d'amplitude au dessus du premier formant, par recherche des fréquences formantiques et par modélisation de la modulation d'enveloppe temporelle dans chaque filtre. Le système complet, conçu pour un but différent, est décrit dans [2]. Les signaux synthétiques obtenus sont perceptivement équivalents aux signaux originaux: seule une écoute attentive permet de les distinguer. La figure 2 montre le signal synthétique correspondant à la figures 1: les formes d'ondes utilisées pour la synthèse sont portées dans le plan temps-fréquence. Cette image montre les paramètres acoustiques sinusoïdaux et formantiques déduit de l'analyse et eessemble à un squelette de la figure 1.

4 conclusion

Ce papier présente une justification perceptive des paramètres acoustiques apparents sur une représentation spectrographique auditive. Après une discussion sur le type de spectrographe utilisé, les relations entre la résolution du spectrographe auditif et les grandeurs acoustiques du signal de parole sont examinées. Des paramètres acoustiques descriptifs sont proposés pour une représentation acoustique auditive de la parole: pour de la parole voisée, le grave du spectre (en dessous d'environ 1 kHz) peut se décomposer comme une somme d'harmoniques, qui subissent l'influence du premier formant et de la source de voisement; dans une région spectrale moyenne (entre environ 1 et 3 kHz), région des seconds et troisièmes formants les signaux filtrés apparaissent comme des signaux de fréquences dominantes approximativement égales aux fréquences formantiques, modulés en amplitudes à la fréquence du fondamental; au delà d'environ 3 kHz, les formants supérieurs perdent de leur individualité et se regroupent en masses spectrales dont les fréquences dominantes sont peu saillantes et dont la modulation d'amplitude est complexe. Un système d'analyse/synthèse par formes d'ondes élémentaires a permis de montrer l'équivalence perceptive entre le signal naturel et un signal synthétique obtenu en utilisant ces paramètres. Le spectrographe auditif propose une représentation acoustique descriptive qui se démarque de celle basée sur un modèle de production, mais qui paraît perceptivement justifiée. L'avenir permettra de juger de l'efficacité de cette représentation, grâce à un synthétiseur de parole utilisant ce type de paramètre qui est actuellement à l'étude pour la synthèse à partir du texte.

Références

[1] d'ALESSANDRO, C. et BEAUTEMPS, D. (1990), "Representation et modification du signal de parole par transformée en ondelettes utilisant des contraintes auditives", Notes et Documents LIMSI 90-10.

[2] d'ALESSANDRO, C. (1990), "Time-frequency speech transformation based on an elementary waveform representation", Speech Comm. Vol. 9. Nos 5/6, pp. 419-431.

[3] BLADON, A. (1982), "Arguments against formants in the auditory representation of speech", in The Representation of Speech in the Peripheral Auditory System, R. Carlson and B. Granström eds, Elsevier Biomedical Press (North-Holland). pp. 95-102.

[4] CARLSON, R. and GRANSTROM, B. (1982), "Towards an auditory spectrograph", in The Representation of Speech in the Peripheral Auditory System, R. Carlson and B. Granström eds, Elsevier Biomedical Press (North-Holland). pp. 95-102.

[5] COOKE, M.P. (1986), "A computer model of peripheral auditory processing uncorporating phase-locking, suppression and adaptation effects", Speech Comm., Vol. 5. Nos 3/4, pp. 261-281.