

Analyse spectrale des sons musicaux et de la parole

par Étienne SICARD
Maître de Conférences
et Anne MENIN
Orthophoniste

INSA - Département Génie Électrique
Avenue de Rangueil - 31077 Toulouse Cedex

RÉSUMÉ

Cet article présente une série d'illustrations des sons musicaux et des sons de la parole d'un point de vue temporel et fréquentiel, avec un système d'acquisition sur PC et un logiciel simple de visualisation du signal sonore et de calcul du spectre de fréquence associé. La structure des sons musicaux est décrite d'un point de vue harmonique en mettant en valeur le rôle du son fondamental, du timbre et de l'enveloppe. Après un rappel sur la physiologie des organes phonateurs, la structure des sons de la parole est présentée par les notions de formants, de triangle vocaliques illustré par les voyelles /a/i/ou/ et de traits consonantiques par les consonnes /p/b/m/.

1. INTRODUCTION

Le but de cet article est de présenter différents sons musicaux et sons de la parole sous leurs aspects temporels tels qu'ils peuvent être numérisés à l'aide d'interfaces d'acquisition de données, et sous leurs aspects fréquentiel grâce à la transformation en série de Fourier.

Les sons présentés dans cet article ont tous été numérisés à 10 kHz, par l'intermédiaire de la carte **Candiplus**, avec une résolution ramenée à huit bits. Le logiciel de numérisation, de présentation des résultats et de calcul du spectre est le didacticiel **Audio** de Langage et Informatique [1]. La fenêtre de calcul du spectre est de type rectangulaire, et inclut cent vingt-huit points de mesure. Le spectre est calculé selon l'algorithme de Transformée de Fourier Rapide (ou *Fast Fourier Transform* en Anglais, souvent noté FFT). Il donne une évaluation de soixante-

quatre composantes du spectre, allant de 78 Hz à 5 000 Hz, c'est-à-dire l'essentiel de l'étendue fréquentielle des sons de la musique et de la parole. Le calcul répétitif de la transformée de Fourier par tranche de cent vingt-huit échantillons permet de représenter l'évolution du spectre sonore dans le temps. Cette représentation à trois dimensions (fréquence, amplitude et temps) se révèle être un outil de grande portée pédagogique, en plus de son aspect visuel particulièrement attrayant.

Il a déjà été présenté dans des articles précédents [3], [4] un certain nombre de commentaires, suggestions et recommandations concernant l'utilisation à bon escient de la Transformée de Fourier. Dans cet article, le spectre de Fourier bien qu'étant l'outil de base de nos illustrations ne constitue pas une fin en soi. Les erreurs numériques liées à l'utilisation de la FFT ainsi que le choix de la fenêtre de calcul ne seront pas discutées ici.

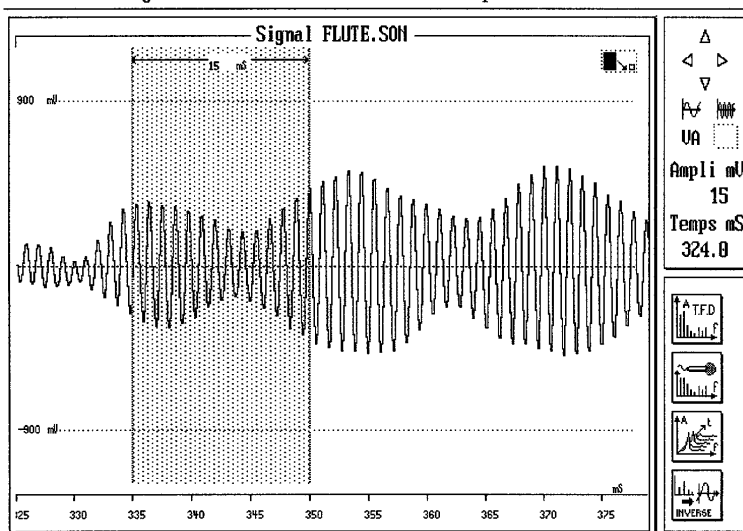
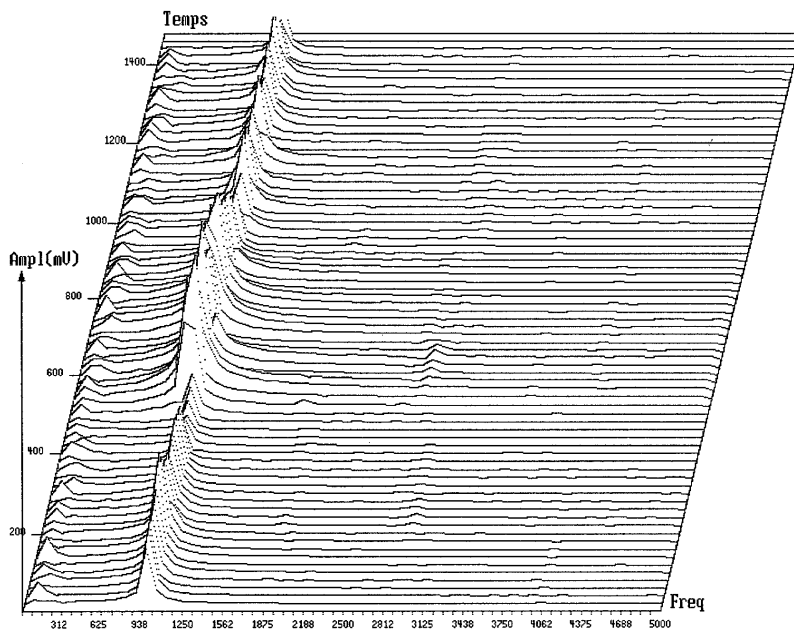
Au chapitre 2. de l'article, nous abordons les sons musicaux. Les notions de fréquence fondamentale, d'harmonique, et d'enveloppe sont introduites par des exemples représentatifs de sons musicaux. Les sons de la flûte, de la cloche, du violon et de la guitare sont commentés dans une approche voisine de [2], où le sonographe est remplacé par la visualisation en trois dimensions, plus explicite et plus détaillée en fréquence.

Les sons de la parole, mettant en jeu des mécanismes assez subtils, et de classification plus complexe font l'objet du chapitre 3. Celui-ci commence par un bref exposé des données physiologiques des organes phonateurs. La notion de formant est ensuite abordée avec la correspondance physiologique et la représentation spectrale associée. Le positionnement des voyelles est proposé suivant le principe du triangle vocalique. Enfin, pour clore le chapitre, un choix de consonnes permet d'illustrer leurs principes articulatoires de base. Le choix a porté sur les consonnes /s/p/m/, qui illustrent respectivement les caractères fricatif, explosif et nasal des phonèmes consonantiques. En ce qui concerne les notations, f_0 est la fréquence fondamentale et $n \times f_0$, les harmoniques de rang n .

2. LES SONS MUSICAUX

La forme d'onde d'un son musical est de complexité très variable. Certains sons tels que la flûte ont une évolution spectrale simple, avec une composante principale correspondant à la fréquence fondamentale

Fichier Enregistrer Créer Editer Jouer Imprimer

**a** - Allure temporelle**b** - Spectre**Figure 1** : Allure temporelle et spectrale d'un son joué à la flûte.

suivi d'une faible série d'harmoniques. La figure 1 illustre l'allure du son de la flûte dans le temps numérisé à 10 kHz ainsi que son spectre calculé par Transformée de Fourier Rapide répétée dans le temps. Le calcul du spectre porte sur cent vingt-huit points de mesure et permet de visualiser les composantes allant de 78 Hz à 5 000 Hz.

De manière générale, le son musical est caractérisé par sa hauteur, son intensité et son timbre. La hauteur est donnée par la fréquence du son fondamental c'est-à-dire le son le plus grave. L'intensité est fonction de la pression acoustique, matérialisée par l'amplitude de l'onde. La distribution des harmoniques donnent au son une coloration particulière, le timbre. L'appréciation du timbre est plus subtile car celui-ci dépend de la structure harmonique du son.

La distribution peut être construite sur la base de séries simples. Dans le cas de la flûte (figure 1), le spectre se compose principalement de la fondamentale f_0 évaluée à 930 Hz. On devine une composante harmonique à $3 \times f_0$ (2 800 Hz).

Il existe aussi des séries plus complexes, voire même aléatoires, qui ne sont pas seulement explicitées en termes de rapport simple avec la fréquence fondamentale. Le son produit par une cloche dont le spectre est reporté figure 2 illustre bien ce phénomène. Il existe une multitude de composantes d'énergie, certaines présentant des phénomènes de battement au cours du temps.

Le son d'un violon est analysé figure 3. Il correspond à une gamme composée de huit notes jouées rapidement. La structure harmonique se compose de la fondamentale f_0 que l'on voit se décaler vers les aigus avec le temps, et d'harmoniques principales $2 \times f_0$ et $3 \times f_0$ dont les rapports se modifient avec la hauteur de la note. Les composantes harmoniques du son d'un violon tirent une partie de leur richesse dans les phénomènes de résonance des cordes voisines : au temps $t = 1\,000$, l'essentiel de l'énergie est sur f_0 (corde de LA accordée à environ 440 Hz), au temps $t = 1\,400$ l'énergie la plus importante est sur $2 \times f_0$ (note SI qui fait vibrer par «sympathie» la corde voisine).

Une note jouée à la guitare diffère d'une note jouée au violon par la non uniformité du son : la corde pincée crée une sorte d'explosion suivie d'une relaxation, par opposition au son continu et homogène du violon. L'enveloppe de la note jouée à la guitare est donc non constante. La figure 4 représente le spectre de quelques notes jouées à la guitare.

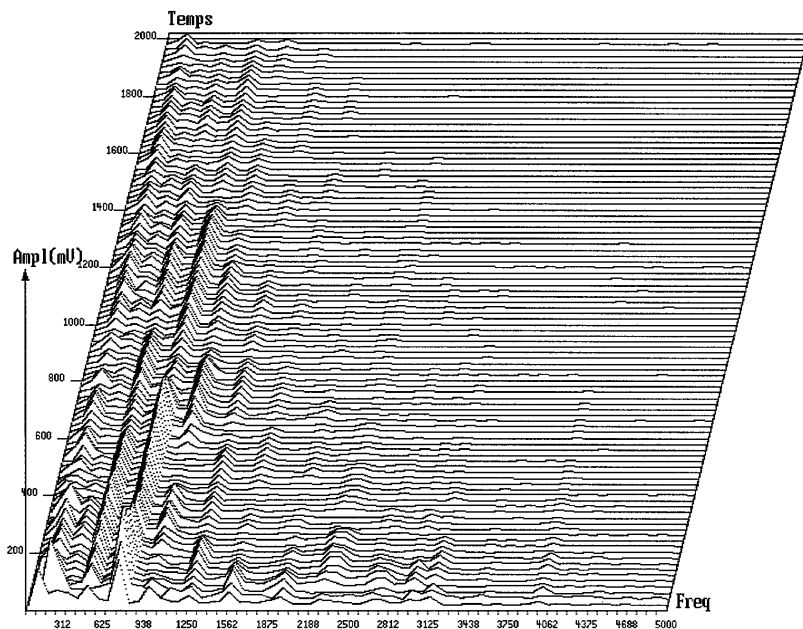


Figure 2 : Allure temporelle du son d'une cloche.

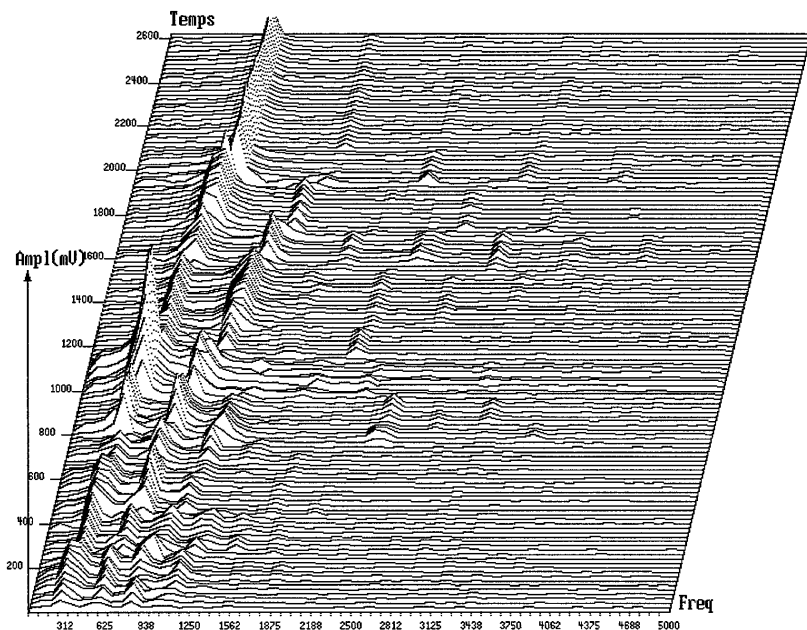


Figure 3 : Une gamme jouée au violon.

Les attaques sont situées aux temps $t = 0$, $t = 500$ ms, $t = 1\,000$ ms et $t = 1\,500$ ms environ.

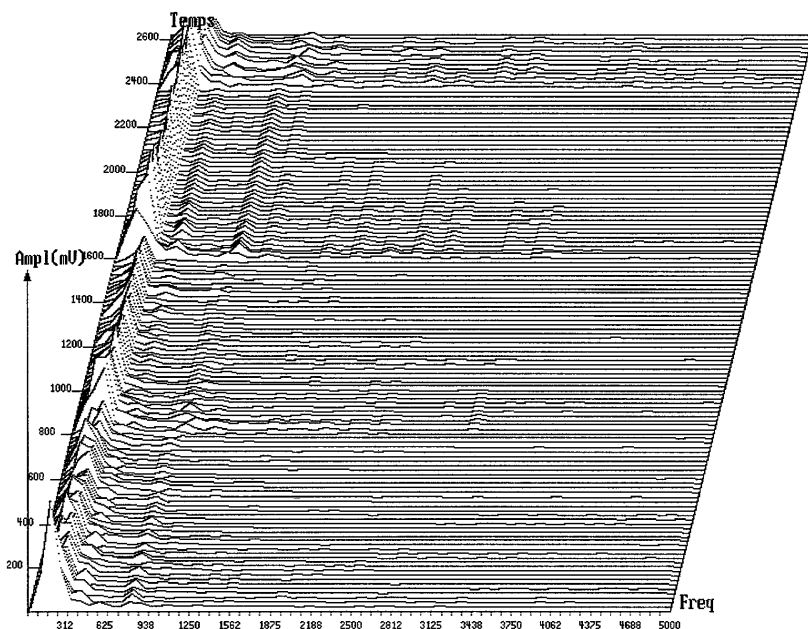


Figure 4 : Quelques notes à la guitare.

3. LES SONS DE LA PAROLE

3.1. Données physiologiques

La distinction des sons de la parole en deux catégories bien distinctes, les voyelles et les consonnes, est principalement liée à leur caractère périodique ou apériodique. Les sons vocaliques sont périodiques et classifiés en terme de répartition des énergies dans le spectre, tandis que les sons consonantiques sont assimilables à des bruits, avec une structure moins ordonnée et apériodique.

Les sons vocaliques résultent de deux phénomènes : la mise en vibration des cordes vocales situées dans le larynx, et la mise en résonance du pharynx, de la cavité buccale et de la cavité nasale. L'ouvrage de Le Huche [5] fait référence pour sa description claire et complète de l'anatomie des organes de la voix. La figure 5 schématise

l'emplacement des différentes cavités mises en jeu. Les cordes vocales sont situées dans le larynx à l'intérieur d'un cartilage qui fait saillie dans la gorge et que l'on appelle communément la «pomme d'adam». Le larynx représente l'ensemble des cartilages et muscles constituant le support des cordes vocales. On peut aisément percevoir la vibration des cordes vocales en plaçant deux doigts de part et d'autre de la pomme d'adam. Le pharynx est le conduit situé au-dessus du larynx prolongé par les cavités nasale et buccale.

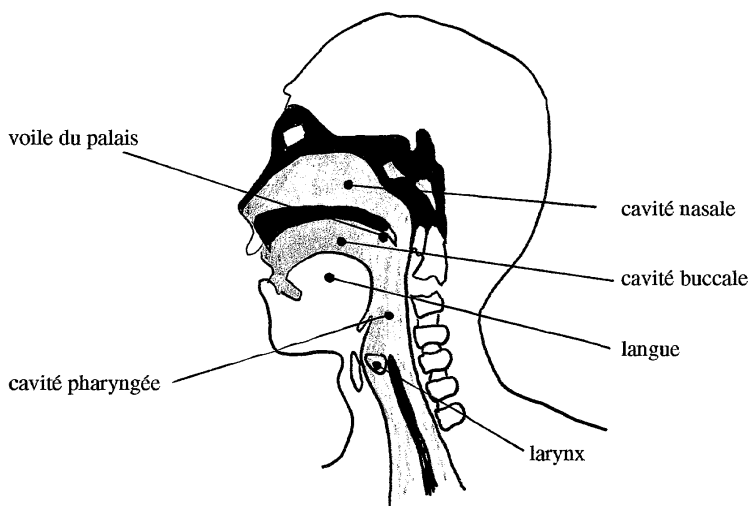


Figure 5 : Emplacement du larynx et des cavités laryngée, pharyngée, buccale.

La vibration des cordes vocales dépend de deux mécanismes : un mécanisme purement musculaire qui a pour rôle de rapprocher les deux faisceaux de cordes vocales, et le flux d'air provenant des poumons qui va initier la vibration. La hauteur de la note produite, c'est-à-dire sa fréquence est liée à la tension des cordes vocales. Celles-ci ont les mêmes propriétés physiques qu'une corde d'un violon. Plus elle sont tendues plus le son résultant est aigu, et inversement. Lorsque le chanteur veut produire un son, il met en jeu un ensemble complexe de muscles reliés à des cartilages articulés entre eux, afin de rapprocher les cordes vocales et les tendre en fonction de la hauteur du son désiré. La beauté du son produit dépend aussi de la qualité d'accolement c'est-à-dire de la fermeture des cordes vocales. Lorsque la voix est rauque, c'est souvent que les cordes vocales se ferment mal.

Le son produit par le larynx est déjà complexe et riche en composantes spectrales. Il traverse ensuite les conduits et cavités qui se trouvent sur son passage : la cavité pharyngale, la cavité buccale et la cavité nasale. Chaque cavité possède une fréquence de résonance propre, liée à sa morphologie particulière et sa position par rapport aux autres. C'est par effet de résonance que certaines composantes du son issu du larynx sont amplifiées. L'ensemble de ces composantes spectrales amplifiées par la disposition et la forme de ces différentes cavités forment des «paquets» d'énergie appelés *formants*. La localisation précise des formants dans le spectre fréquentiel est l'une des bases de la reconnaissance automatique de la parole [6]. Les formants du son vocalique /a/ sont extraits du spectre de la figure 6.

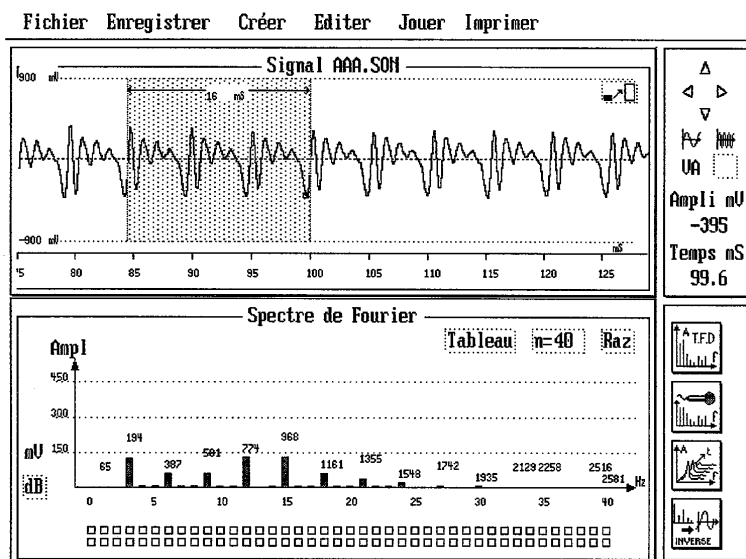


Figure 6 : Son vocalique : le son /a/ et son spectre.

Toutes les harmoniques de F_0 jusqu'à $8 \times f_0$ sont représentées, ce qui est caractéristique de la voyelle /a/. Le formant F_0 se compose de la fréquence fondamentale f_0 située à 194 Hz et de l'harmonique $2 \times f_0$. Le dôme de point culminant à 968 Hz est le formant F_1 . Le formant F_2 , accolé à F_1 , est constitué des harmoniques 6, 7, $8 \times f_0$.

Il existe une correspondance physiologique entre les formants et les cavités. En effet, les harmoniques du son se trouvent donc amplifiés

à plusieurs niveaux : principalement, au niveau de la cavité pharyngale, ce qui correspondra au premier formant F1 et au niveau de la cavité buccale, ce qui donnera le deuxième formant F2. D'autres cavités telles que les sinus ou les lèvres projetées en avant génèrent un troisième formant F3, d'amplitude nettement plus faible que F1 et F2. Ce sont donc essentiellement les deux premiers formants qui caractérisent les sons vocaliques.

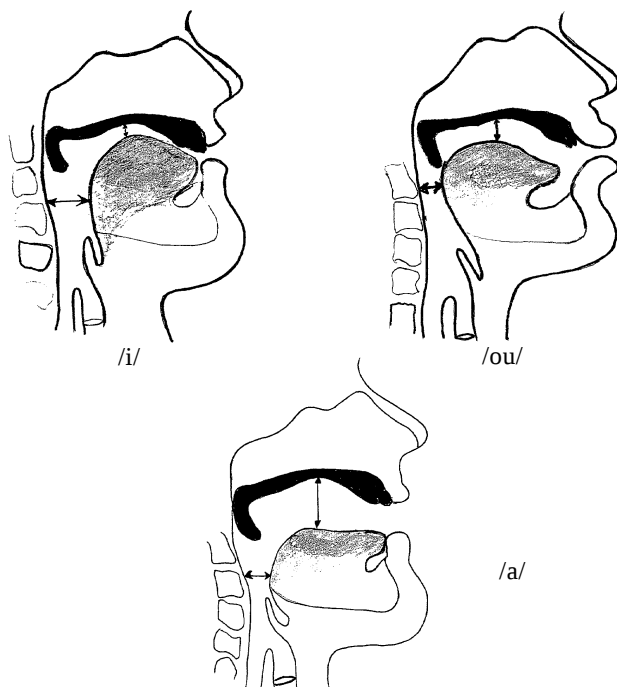


Figure 7 : Position de la langue des voyelles /a/i/ou/.

3.2. Les voyelles

La position de la langue joue un rôle fondamental dans la production des voyelles. Par exemple lorsque l'on produit le son /ou/ (figure 7) le volume de la cavité pharyngée est petit car la langue adopte une position très reculée dans la gorge obstruant partiellement celle-ci. En revanche, elle dégage alors un long volume dans la cavité buccale, prolongé par la projection des lèvres vers l'avant créant ainsi une sonorité grave, particulièrement sonore. Dans le cas du /i/, le volume

de la cavité buccale est réduit par la fermeture partielle de celle-ci, et par la position antérieure de la langue qui dégage alors une cavité pharyngée volumineuse. Le phonème /i/ est globalement plus aigu et peu sonore. Dans le cas du /a/ les cavités buccale et pharyngée sont relativement grandes grâce à l'ouverture de la bouche et à la position intermédiaire de la langue.

Le spectre des voyelles /a/i/ou/ est reporté figure 8. On note une structure harmonique bien différenciée. La variation de la fréquence fondamentale f_0 associée à chaque voyelle réalise l'intonation de la voix.

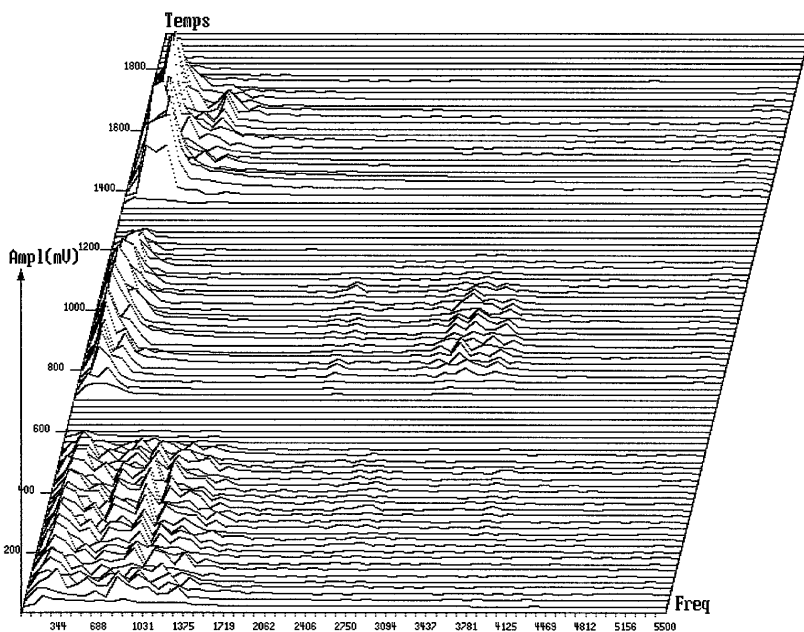


Figure 8 : Spectre des voyelles /a/i/ou/.

Une représentation des voyelles les unes par rapport aux autres, consiste au positionnement de chaque voyelle dans un plan où X représente la fréquence du formant F1 et Y la fréquence du formant F2. La représentation prend la forme d'un triangle (figure 9) dont l'aspect et les valeurs numériques évoluent légèrement selon qu'il s'agit d'une voix d'homme ou une voix de femme. La définition précise de F1 et F2 n'est pas chose aisée comme cela a été le cas pour le /a/. Cette

représentation bien que séduisante par sa simplicité n'en pose pas moins de sérieux problèmes quant à l'extraction des valeurs numériques des formants dans les cas pratiques.

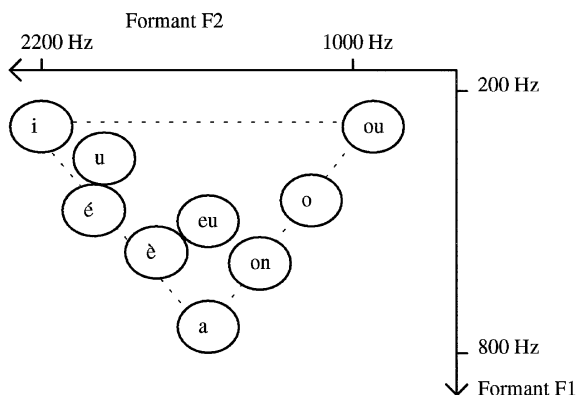


Figure 9 : Visualisation des formants F1 et F2 de la langue française sous la forme d'un triangle vocalique (voix d'homme).

Il existe une autre catégorie de phonèmes vocaliques particulièrement nombreux dans la langue française, les voyelles *nasales*. Ces sons résultent de la mise en vibration de la cavité du nez. Cette vibration est obtenue grâce au relâchement volontaire de ce que l'on appelle le voile du palais (figure 10). Il s'agit d'une petite paroi mobile terminée par la luette qui en s'abaissant, libère la cavité de résonance nasale. Celle-ci était obturée dans la production des voyelles non nasalisées.

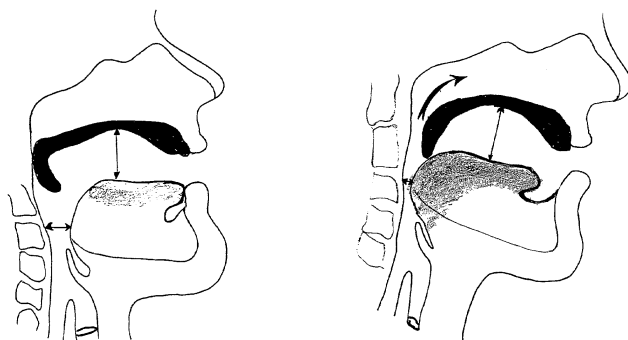


Figure 10 : Position du voile du palais dans le cas des sons /a/ et /an/.

Il suffit d'appliquer légèrement deux doigts sur le nez en prononçant alternativement /a/ et /an/ pour sentir la vibration nasale consécutive à l'ouverture de la cavité du nez. Ce phénomène peut être visualisé facilement devant une glace : on aperçoit la luette qui s'abaisse jusqu'à disparaître derrière la base de la langue. La même expérience peut être répétée pour le /o/ et le /on/ ainsi que le /è/ et le /in/ différenciés selon le même mécanisme.

Le spectre des voyelles /a/ et /an/ est reporté figure 11. La structure harmonique du /a/ est riche (0 à 1 100 ms) tandis que celle du /an/ bien qu'étant un son voisin est plus pauvre (1 400 ms à 2 600 ms). L'atténuation des énergies dans la zone 600-1 200 Hz est due [7] à la diminution de l'ouverture de la cavité buccale et de sa mise en communication avec les fosses nasales, provoquant un effet «d'anti-formant». Cette atténuation crée à elle seule l'effet de nasalisation, que l'on retrouverait de la même manière dans le couple /o/on/.

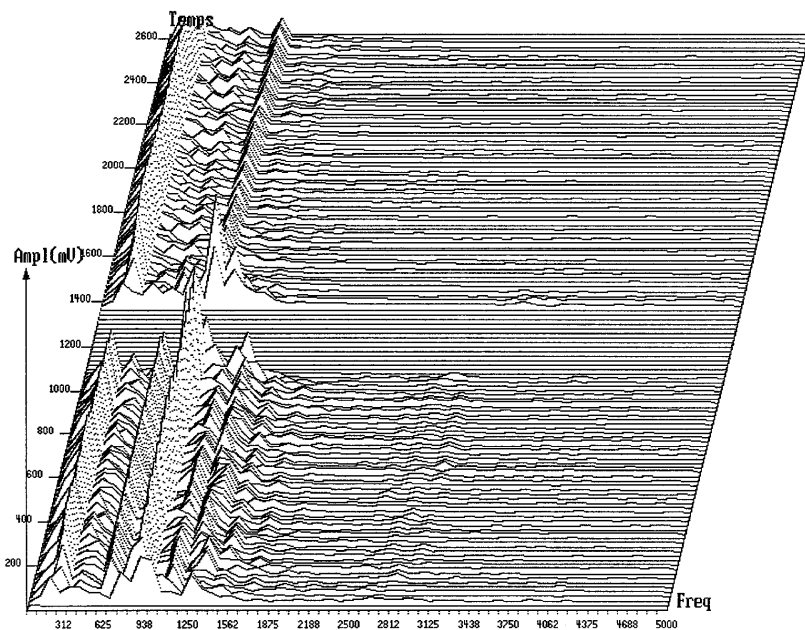


Figure 11 : Spectre des voyelles /a/ et /an/.

3.3. Les consonnes

Les consonnes sont présentées dans ce chapitre suivant trois traits distinctifs : frictions, explosions et nasalisations. Les *frictions* sont des vibrations acoustiques erratiques, intermittentes ou aléatoires, constituées généralement d'une structure harmonique complexe. La figure 12 donne l'exemple de la consonne /f/. Le spectre comporte de nombreuses composantes spectrales de faible amplitude et parsemées. On note l'aspect aléatoire des composantes spectrales, caractéristique du bruit.

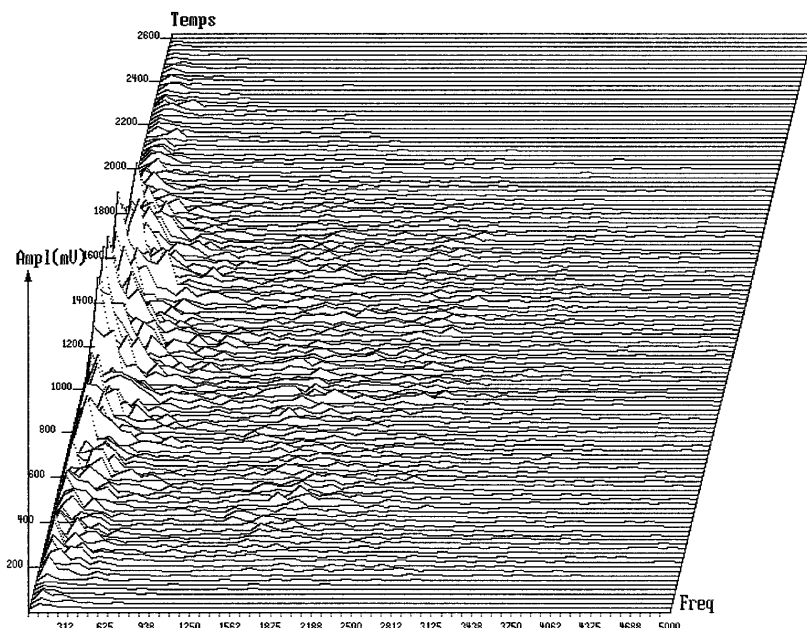


Figure 12 : Spectre d'une consonne /f/.

Les consonnes dites fricatives ou encore constructives, résultent du passage du souffle d'air dans un rétrécissement réalisé soit par la langue qui se rapproche du palais (cas des sons /ch/, /j/, /s/, /z/) soit par un rapprochement des lèvres et des dents (cas des sons /f/, /v/).

Les consonnes dites *explosives* sont le résultat d'une fermeture très brève des organes phonateurs. Cette occlusion réalisée soit par les deux lèvres /p/, /b/, soit par l'extrémité de la langue sur le palais /t/, /d/ ou encore par la base de la langue sur le fond du palais /k/, /g/, est

immédiatement suivie d'un relâchement d'air réalisant ainsi une légère explosion de souffle. Ces consonnes se reconnaissent aisément sur le spectre car elles se caractérisent par de soudaines énergies spectrales consécutives à la phase d'occlusion. Cette dernière est à l'origine des interruptions de son bien visibles sur la figure 13 au temps 500 ms.

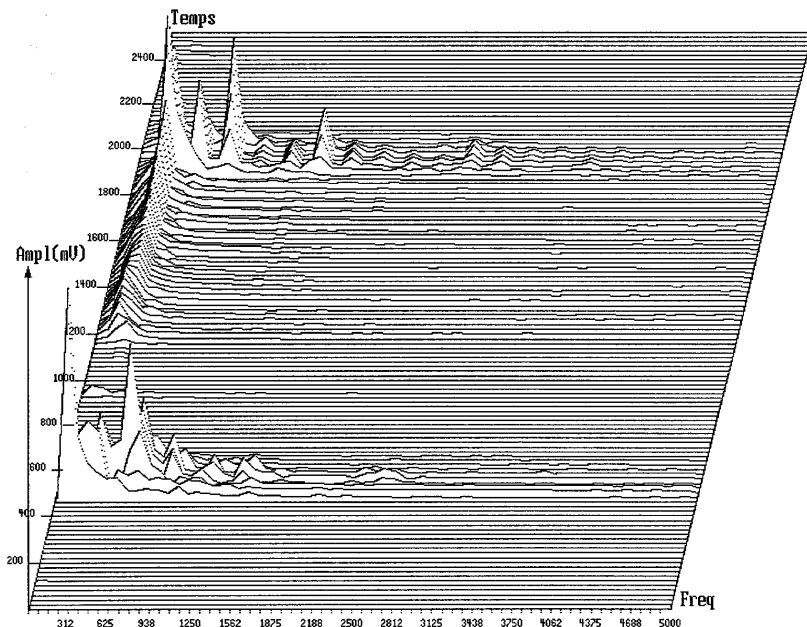


Figure 13 : Spectre des sons /p/ suivi de /m/.

Les consonnes, tout comme les voyelles, peuvent aussi bénéficier de la cavité de résonance du nez. Ce trait est appelé *nasalisation*. C'est le cas des phonèmes /m/ et /n/ qui se caractérisent sur le spectre par une composante spectrale continue, comme l'indique la figure 13 du temps 1 200 ms au temps 1 900 ms.

4. CONCLUSION

La découverte de la structure des sons musicaux et des sons de la parole grâce à l'analyse spectrale est une aventure passionnante. L'illustration des notions de fréquence fondamentale, d'harmoniques, d'enveloppes et de formants est grandement facilitée par les outils

d'analyse du signal par Transformée de Fourier Discrète, et par une visualisation de l'évolution du spectre en trois dimensions. Les études proposées dans cet article s'adressent aux lycéens dans le cadre de l'étude des sons, aux étudiants des universités et grandes écoles en illustration des cours sur le traitement du signal et d'acoustique, ainsi qu'aux étudiants en orthophonie pour l'illustration de la phonétique. Les exemples qui illustrent cet article traitent une grande variété des sons, donnant des représentations spectrales des plus simple au plus complexe, du plus ordonné au plus aléatoire. L'accent a été mis sur la classification des voyelles et consonnes, l'explication des notions de formants, de triangle vocalique qui sont à l'origine des systèmes de reconnaissance automatique de la parole.

BIBLIOGRAPHIE

- [1] AUDIO II, édité par Langage et Informatique, Toulouse, ISBN 2-84131-001-9, 1993.
- [2] M. CASTELLENGO : «*Perception du signal musical*», B.U.P. n° 761, p. 197.
- [3] C. CANCE : «*Mise en œuvre pratique de l'analyse spectrale numérique*», B.U.P. n° 761, p. 211.
- [4] J. ESQUIEU : «*Traitement numérique du signal*», B.U.P. n° 754.
- [5] LE HUCHE : «*Anatomie et physiologie des organes de la voix et de la parole*», édité par Masson.
- [6] HATON et Coll. : «*Reconnaissance automatique de la parole*», édité par Dunod.
- [7] F. CARTON : «*Introduction à la phonétique du français*», édité par Bordas.
- [8] P. DELATTRE : «*Studies in French and Comparative Phonetics*», Mouton, La Haye, 1966.