

Analyse de signaux sonores par les lois de Zipf et Zipf Inverse

Emmanuel DELLANDRÉA¹, Pascal MAKRIS¹, Michelle BOIRON², Nicole VINCENT¹

¹Laboratoire d'Informatique - Université de Tours
64 avenue Jean Portalis, 37200 Tours, France

²Laboratoire de Physiologie et de Motricité Digestive
2 boulevard Tonnellé, 37000 Tours, France

dellandrea@univ-tours.fr, makris@univ-tours.fr
boiron@med.univ-tours.fr, vincent@univ-tours.fr

Résumé – Nous présentons dans cet article un ensemble de codages de signaux sonores que nous avons développés afin d'adapter à ce type de signaux, l'analyse par les lois de Zipf et Zipf Inverse. L'efficacité de ces lois à décrire les phénomènes physiques n'est plus à démontrer, et à motiver nos investigations concernant le problème de la caractérisation de signaux sonores. Afin de valider notre approche, la méthode a été évaluée sur des signaux sonores médicaux, correspondant à des bruits xiphoidiens.

Abstract – We present in this paper a set of audio signal codings that we have developed in order to adapt to this kind of signals, studies by Zipf and Inverse Zipf laws. The efficiency of these laws to model numerous physical phenomena has been shown and has motivated our investigations concerning the problem of audio signals characterization. In order to validate our approach, the method has been evaluated on medical audio signals, corresponding to xiphoidal sounds.

1 Introduction

L'identification et la caractérisation de signaux est un vaste domaine de recherche du fait des nombreuses difficultés rencontrées, ainsi que du nombre très important des applications possibles (indexage multimédia, diagnostique médical, analyse de la parole, ...). Ainsi les ordinateurs sont d'une grande aide pour les humains, leur permettant d'analyser des signaux de plus en plus nombreux dans des domaines variés. Cependant, si la tâche est réalisée naturellement par les humains d'une manière apparemment aisée, il est nécessaire de déterminer une méthodologie implicite afin de développer des systèmes d'identification automatique. La classification de signaux sonores (ASC pour *Audio Signal Classification*) [1] est une branche spécifique, destinée à l'analyse de signaux sonores.

Généralement, le principe d'un ASC est d'extraire des primitives à partir des signaux sonores et ensuite d'identifier les classes en fonction de la plus grande probabilité d'appartenance.

Différents types de primitives peuvent être envisagés [1], en particulier des primitives de type physique et de type perceptuel. Les premières sont basées sur les propriétés statistiques et mathématiques des signaux, alors que les deuxièmes sont liées à la manière dont les humains entendent les signaux.

Lorsque les primitives ont été extraites, elles sont assemblées dans un vecteur de primitives. Une méthode de classification est alors utilisée afin de déterminer la classe à laquelle le vecteur a la plus grande probabilité d'appartenir.

Nous présentons dans cet article de nouvelles primitives permettant la quantification de la structure des signaux grâce aux lois de Zipf et Zipf Inverse [2]. L'efficacité de ces lois à décrire les phénomènes physiques n'est plus à démontrer. Cependant, la nature de la structure mise en évidence est liée au choix d'un

codage du signal.

Dans la première partie de cet article, nous présentons les lois de Zipf et Zipf Inverse. Les codages permettant l'adaptation de ces lois aux signaux sonores seront ensuite décrits. Les primitives issues de l'analyse des signaux par les lois de Zipf et Zipf Inverse, ainsi que les méthodes de classification utilisées seront décrites dans une troisième partie. Enfin, une dernière partie sera consacrée à l'application de nos codages à des signaux médicaux, afin de valider notre méthodologie.

2 Les lois de Zipf

G. K. Zipf proposa dans [3] une loi empirique connue sous le nom de loi de Zipf, qui a été observée dans de nombreux domaines. Cette loi énonce que la fréquence $f(p)$ d'un événement p et son rang $r(p)$ par rapport à sa fréquence d'apparition (du plus fréquent au moins fréquent) sont liés par une loi de puissance: $f(p) = A_1 r(p)^{-\beta}$ où A_1 et β sont des constantes réelles.

L'application la plus célèbre de la loi de Zipf est la linguistique statistique. En effet, il a été montré [3] que cette loi est valide dans le cas de textes écrits en langage naturel, β étant proche de 1. Cette régularité a ensuite été observée pour de très nombreux phénomènes dans divers domaines. Citons par exemple la distribution de la population des villes [3, 4], la distribution du revenu des employés, la distribution de la taille des entreprises aux Etats-Unis [5] ainsi que les séquences d'ADN [6].

Grâce à la loi de Zipf, une caractérisation de certains phénomènes peut être réalisée par l'étude des paramètres liés à cette loi. Il est ainsi possible, par exemple, de distinguer les différents auteurs de textes [7].

Un autre comportement observé pour les textes écrits en lan-

gage naturel a été modélisé par une loi, appelée loi de Zipf Inverse. Elle s'intéresse à la distribution de la fréquence des mots d'un texte. Zipf a découvert une loi de puissance qui est vérifiée uniquement pour les mots de faible fréquence : le nombre de mots distincts $I(f)$ de fréquence d'apparition f et la fréquence f sont liés par la relation : $I(f) = A_2 f^{-\gamma}$ où A_2 et γ sont des constantes réelles.

Les perspectives très intéressantes offertes par les lois de Zipf nous ont amené à évaluer leur pertinence et leur efficacité pour le problème de classification de signaux sonores. Ainsi, nous avons choisi d'utiliser une analogie entre les signaux sonores et les textes, considérés comme des séquences ordonnées de motifs. La première étape du processus est donc le codage des signaux sonores afin de les transformer en textes. Les lois de Zipf pourront alors être utilisées pour l'analyse de ces signaux à valeurs discrètes.

3 Codage des signaux sonores

Un signal sonore contient au moins deux types d'informations : une information temporelle et une information fréquentielle. Ainsi, deux types de codages vont être présentés, permettant d'exploiter l'information pertinente contenue dans les signaux. Le premier type repose sur la représentation temporelle des signaux sonores et le second repose sur leur représentation temps-fréquence.

3.1 Codages temporels

Les codages temporels des signaux sonores exploitent l'information apportée par l'évolution temporelle de l'amplitude des signaux. Deux codages TC (Temporal coding) ont été développés : les codages TC1 et TC2.

3.1.1 Le codage TC1 : codage d'évolution

Comme pour tous les codages proposés, le principe du codage TC1 est de permettre la construction d'une séquence de motifs basée sur le codage du signal sonore d'origine.

Ici, les motifs sont eux-mêmes des séquences de trois caractères. Ces derniers, pour chaque échantillon sonore, peuvent prendre trois valeurs : la lettre U (Up), F (Flat) et D (Down), qui remplacent les valeurs du signal. La lettre U est utilisée lorsque la différence d'amplitude entre deux échantillons successifs du signal est positive. Si cette différence est proche de 0, la lettre F est utilisée, et sinon la lettre D est utilisée.

Ainsi, un motif représente l'évolution temporelle locale sur quatre échantillons du signal. Le motif suivant est obtenu en translatant d'un échantillon vers la droite, la fenêtre d'analyse de taille 3 échantillons. Une séquence de motifs est finalement construite à partir du signal sonore. Il y a $3^3 = 27$ possibilités différentes de motifs lorsqu'ils sont de taille 3. Chacun d'eux est associé à une lettre de l'alphabet. Le nombre de lettres nécessaires est égal au cardinal de l'ensemble des motifs.

A ce stade, une séquence de lettres est obtenue, ces dernières prenant en compte une information de variation d'amplitude très locale. Afin d'adapter la longueur de la fenêtre d'observation à des échelles plus larges, les lettres sont regroupées afin de créer des mots de longueur n qui vont constituer un texte. n

est un paramètre qui doit être fixé et qui dépend de la longueur de la fenêtre d'observation à utiliser.

L'intérêt du codage TC1 est que chaque lettre correspond à la forme locale de l'enveloppe du signal temporel. Ainsi, les mots représentent les variations d'amplitude du signal.

3.1.2 Le codage TC2 : codage d'amplitude adaptatif

L'objectif du codage TC2 est d'utiliser l'information relative à l'amplitude du signal sonore. La notion de grande ou petite amplitude est bien sûr liée au signal à étudier car pouvant dépendre notamment des conditions d'acquisition du signal. Une méthode de définition adaptative de l'importance des signaux est donc utilisée. Pour cela, le signal doit être lu entièrement avant le début du processus de codage.

Le principe du codage TC2 est de considérer la valeur d'amplitude des échantillons sonores du signal. Afin de réduire la dimension des données, ces valeurs doivent être discrétisées selon K valeurs. Ainsi, à partir de ces données initiales, une classification par l'algorithme des K -mean [8] est réalisée, permettant d'obtenir K classes. Chacune d'elles correspond à un intervalle d'amplitude. Ce processus de classification permet de s'assurer que le code n'est pas absolu mais plus sensible aux variations internes des signaux. Une lettre de l'alphabet est ensuite associée à chaque classe.

A partir du signal sonore, une séquence de lettres est obtenue, celles-ci correspondant à la classe de chaque échantillon.

Enfin, les lettres sont regroupées afin de construire un texte constitué de mots de longueur n .

3.2 Codages temps-fréquence

Dans cette section, deux codages basés sur la représentation temps-échelle des signaux sonores sont présentés, permettant de prendre en compte simultanément l'information temporelle et fréquentielle.

3.2.1 Codage TSC1 : transformée en ondelettes continue

Le codage TSC1 (Time Scale Coding) repose sur une représentation temps-échelle des signaux sonores, obtenus grâce à la transformée en ondelettes continue [9].

Le résultat de cette transformée correspond aux coefficients d'ondelettes qui sont fonction de l'échelle a liée à la fréquence, et de la position b . Les positions correspondent à l'indice des échantillons du signal, et les valeurs d'échelle doivent être fixées.

La transformée en ondelettes continue des signaux sonores est réalisée en utilisant k différentes valeurs d'échelle. Ainsi, chaque position b est associée à un ensemble de k coefficients appelé *profil* et organisé dans le vecteur suivant :

$C_b = (c_1^b, c_2^b, \dots, c_k^b)^t$ où M^t désigne la transposée de la matrice M et c_i^b est le $i^{ième}$ coefficient d'ondelette calculé à la position b .

L'objectif, à ce stade où la nature de l'information à considérer à chaque instant a été choisie, est de créer un code en utilisant un nombre limité de symboles qui seront regroupés afin de former de courtes séquences significatives correspondant à des "mots".

Ainsi, les profils C_b calculés à chaque instant b sont constitués de k coefficients numériques tels que les premiers soient

liés aux hautes fréquences du signal alors que les suivants soient liés aux fréquences plus basses. Cette variété de l'information doit être préservée dans le codage, et les mots vont donc représenter cette unité d'observation. Afin de permettre l'étude de l'évolution des paramètres dans le temps, chaque valeur de coefficient est comparée à la valeur moyenne des coefficients correspondants calculée sur l'ensemble du signal. Seule une information binaire est alors conservée, relative à l'importance du coefficient considéré par rapport à la valeur moyenne. Cette information est contenue dans une représentation binaire de k coefficients. Néanmoins, afin d'obtenir un codage plus compact, ces coefficients sont regroupés de manière à former une séquence de longueur m interprétée comme le code d'un nombre exprimé dans une base binaire. Un symbole explicite est ainsi défini parmi 2^m symboles possibles.

Les différentes étapes de ce processus sont exprimées par les formules suivantes.

Soit le vecteur \bar{C} égal à la valeur moyenne des vecteurs C_b : $\bar{C} = \frac{1}{n} \sum_{r=1}^n C_b = (\bar{c}_1, \bar{c}_2, \dots, \bar{c}_k)^t$ où n est égal au nombre d'échantillons considérés.

Pour chaque position b , un profil binaire est ensuite calculé $B_b = (b_1^b, b_2^b, \dots, b_k^b)^t$ avec :

$$b_i^b = \begin{cases} 1 & \text{si } c_i^b > \bar{c}_i \\ 0 & \text{sinon} \end{cases} \quad \forall i \in [1, k] \quad (1)$$

Les coefficients b_i^b sont ensuite regroupés en l ensembles de m coefficients tels que l et m soient des entiers strictement positifs et $l * m = k$.

Le vecteur S_b est finalement calculé : $S_b = (s_1^b, s_2^b, \dots, s_l^b)^t$ avec :

$$s_j^b = \sum_{i=1}^m 2^{i-1} b_{(j-1)*m+i}^b \quad \forall j \in [1, l] \quad (2)$$

Les entiers s_j^b sont interprétés comme l'index de caractères d'un alphabet contenant 2^m caractères. Ainsi, chaque vecteur S_b , et donc chaque position b , est codé par un mot de longueur l lettres, sélectionné dans un ensemble de mots de cardinal 2^{lm} .

A partir de l'ensemble de profils constitués des coefficients de la transformée en ondelettes continue du signal, un texte composé d'une séquence ordonnée de mots est donc obtenu.

3.2.2 Le codage TSC2 : multirésolution par ondelettes

Le codage TSC2 (Time Scale Coding) est basé sur une analyse multirésolution par ondelettes des signaux sonores [9] permettant de décomposer un signal d'origine en un ensemble de signaux d'approximation comportant les informations basses fréquences du signal, et de signaux de détail comportant les informations hautes fréquences.

Puisque les signaux auxquels nous nous intéressons ont un comportement non stationnaire avec de fortes discontinuités, notre étude porte sur les signaux de détail permettant de mettre en évidence ces informations.

Ainsi, à partir de la représentation du signal obtenue par une analyse multirésolution par ondelettes, des profils C_b sont calculés, tout comme dans le cas du codage TSC1. Ces profils sont obtenus en considérant k signaux de détail de niveaux successifs. Pour chaque position b , les profils C_b sont donc constitués de k coefficients numériques correspondant aux valeurs des signaux de détail à la position b .

Un texte est ensuite obtenu de la même manière que pour le codage TSC1.

Cette phase de codage peut être considérée comme la transformation d'un signal dans un nouvel espace de représentation, et dans chaque cas, est obtenue une séquence de mots construits sur un alphabet relativement petit. Chaque codage a pour rôle de mettre en valeur différentes caractéristiques du signal sonore. Néanmoins la quantité de données est toujours importante et un nombre limité de primitives doit être extrait afin de permettre la classification des signaux.

4 Extraction des primitives et classification

Comme indiqué dans la section 2, la loi de Zipf implique le nombre d'occurrences d'un événement et permet de modéliser la relation entre le nombre d'occurrences de ces événements et leur rang relatif à leur fréquence d'apparition (du plus fréquent au moins fréquent). Cette relation s'exprime sous la forme d'une loi de puissance qui, lorsqu'elle est représentée dans un espace logarithmique (appelé courbe de Zipf) devient linéaire.

La forme de cette courbe est liée à la structure du texte correspondant. Afin de l'étudier, la loi de Zipf n'étant pas toujours parfaitement vérifiée, la fonction correspondante est approximée par un polynôme dont les coefficients feront partie de l'ensemble des primitives associées à l'ensemble Zipf.

La loi de Zipf Inverse correspond à l'étude de la distribution de la fréquence des mots d'un texte. Cette loi étant également une loi de puissance, elle est représentée dans un espace logarithmique, la courbe correspondante étant appelée courbe de Zipf Inverse. Seules les fréquences les plus faibles sont considérées (par exemple les 10 plus faibles). Il apparaît dans les cas étudiés que cette courbe est généralement correctement approximée par une droite. Une seule valeur est donc retenue afin de caractériser la courbe de Zipf Inverse : le coefficient directeur de la droite de régression linéaire.

Enfin, les paramètres suivants sont calculés : les entropies liées aux lois de Zipf et Zipf Inverse, l'aire sous les courbes de Zipf et de Zipf rééchantillonnée, le nombre de mots distincts, le nombre total de mots, le nombre de mots dont la fréquence d'apparition est de 1, le nombre total d'occurrences différentes et la fréquence la plus élevée.

Ces primitives vont ensuite être assemblées dans un vecteur entité qui caractérisera chaque signal sonore.

L'étape suivante correspond à la classification des signaux afin de déterminer les classes auxquelles les signaux ont la plus grande probabilité d'appartenir.

Nous avons considéré quatre méthodes de classification supervisées : l'analyse discriminante (AD), un réseau de neurones (RN) de type perceptron à une couche cachée, un algorithme K-mean (Kmean) et une distance aux classes utilisant la distance de Mahalanobis (Mahal). De plus, afin de combiner les résultats de ces méthodes, un méthode de fusion des classifieurs basée sur l'utilisation des matrices de confusion [10] a été appliquée.

5 Application à l'analyse de signaux médicaux

Afin d'étudier l'efficacité des codages développés, et la pertinence des primitives extraites, nous avons appliqué notre méthode à l'étude de signaux médicaux. Ces signaux sont des signaux de déglutition enregistrés par un microphone placé près de l'appendice xiphoïde du patient, lors du passage d'un bolus à travers l'œsophage et le sphincter inférieur de l'œsophage (SIO). Notre objectif est d'étudier les bruits xiphoïdiens produits par le SIO dont le dysfonctionnement est responsable du phénomène de reflux gastro-œsophagien. Différents ensembles de classes liés aux caractéristiques de la pathologie, ont été considérés. L'ensemble ba est composé de deux classes : les patients avant et après opération. L'ensemble hiat1 est composé de la classe des patients souffrant d'une hernie hiatale et de celle des patients n'en souffrant pas. L'ensemble hiat2 est composé de la classe des patients souffrant d'une hernie hiatale avec une faible pression du SIO, et de celle des patients n'en souffrant pas. Enfin, l'ensemble oo est composé de la classe des patients ayant une faible ouverture œsophagienne, et de celle des patients en ayant une forte.

111 signaux ont été traités. Ils ont été enregistrés à partir de 11 patients souffrant de reflux gastro-œsophagien. Pour chacun d'eux, les signaux ont été enregistrés avant et après opération. 2/3 des signaux ont été utilisés pour l'apprentissage, et 1/3 pour l'ensemble de test.

Les résultats obtenus pour la classification des bruits xiphoïdiens de l'ensemble de test sont donnés TAB. 1, 2, 3, 4 respectivement pour les codages TC1, TC2, TSC1 et TSC2. Les taux de reconnaissance obtenus par AD, RN, Kmean, et Mahal ainsi que la fusion de ces classificateurs sont indiqués pour les quatre ensembles de classes.

Ensemble	AD	RN	Kmean	Mahal	Fusion	Max
ba	80.00	83.33	73.33	73.33	90.00	90.00
hiat1	80.00	90.00	80.00	60.00	90.00	90.00
hiat2	70.00	60.00	60.00	30.00	80.00	80.00
oo	92.86	92.86	50.00	57.14	92.86	92.86

TAB. 1: Résultat de la classification pour le codage TC1.

Ensemble	AD	RN	Kmean	Mahal	Fusion	Max
ba	78.13	75.00	56.25	65.63	75.00	78.13
hiat1	81.82	81.82	72.73	63.64	90.91	90.91
hiat2	54.55	63.64	54.55	45.45	72.73	72.73
oo	85.71	64.29	71.43	57.14	85.71	85.71

TAB. 2: Résultat de la classification pour le codage TC2.

Ensemble	AD	RN	Kmean	Mahal	Fusion	Max
ba	84.38	62.50	65.63	56.25	81.25	84.38
hiat1	54.55	72.73	63.64	63.64	72.73	72.73
hiat2	63.64	72.73	63.64	27.27	81.82	81.82
oo	64.29	64.29	57.14	57.14	64.29	64.29

TAB. 3: Résultat de la classification pour le codage TSC1.

Les résultats montrent que pour les différents ensembles de classes, un bon taux de reconnaissance est généralement obtenu. En effet, le taux maximum est de 90,00 % pour l'ensemble ba avec le codage TC1, de 90,91% pour hiat1 avec TC2

Ensemble	AD	RN	Kmean	Mahal	Fusion	Max
ba	53.13	43.75	68.75	43.75	68.75	68.75
hiat1	81.82	72.73	63.64	72.73	90.91	90.91
hiat2	72.73	63.64	63.64	45.45	63.64	72.73
oo	71.43	57.14	50.00	57.14	57.14	71.43

TAB. 4: Résultat de la classification pour le codage TSC2.

et TSC2, de 81,82% pour hiat2 avec TSC1 et de 92,86% pour oo avec TC1. Ces résultats indiquent que les codages sont complémentaires, dépendant de l'information à considérer.

La bonne classification des bruits xiphoïdiens prouvent que d'une part, l'information sur la pathologie obtenue à partir des données médicales peut être extraite des signaux sonores, et que d'autre part les primitives de Zipf sont bien adaptées à ces signaux.

6 Conclusion

Nous avons présenté dans cet article de nouvelles primitives permettant la caractérisation de signaux sonores. Elles reposent sur l'application des lois de Zipf et Zipf Inverse, grâce au codage des signaux sonores permettant de les transformer en textes.

Afin de valider notre approche, nous avons analysé des signaux médicaux pour lesquels des méthodes classiques ne nous ont pas permis d'obtenir des résultats satisfaisants. Les résultats obtenus par notre méthode montrent que les codages sont complémentaires, et particulièrement bien adaptés à la nature de ces signaux, permettant de les classer avec succès en des classes liées à la pathologie des patients.

Références

- [1] D. GERHARD, "Audio signal classification: an overview," *Canadian Artificial Intelligence*, vol. 45, pp. 4–6, 2000.
- [2] A. COHEN, R. N. MANTEGNA, and S. HAVLIN, "Numerical analysis of word frequencies in artificial and natural language texts," *Fractals*, vol. 5, no. 1, pp. 95–104, 1997.
- [3] G. K. Zipf, *Human Behavior and the Principle of Least Effort*, Addison-Wesley Press, 1949.
- [4] M. MARSILI and Y.-C. ZHANG, "Interacting individuals leading to zipf law," *PHYSICAL REVIEW LETTERS*, vol. 80, no. 12, pp. 2741–2744, July 1998.
- [5] R. L. AXTELL, "Zipf distribution of u.s. firm size," *SCIENCE*, vol. 293, pp. 1818–1820, September 2001.
- [6] C.K. PENG and AL, "Statistical properties of dna sequences," *Physica A 221*, pp. 180–192, 1995.
- [7] S. HAVLIN, "The distance between zipf plots," *Physica A216*, pp. 148–150, 1995.
- [8] JOHN A. HARTIGAN, *Clustering algorithms*, John Wiley & Sons, Inc, 1975.
- [9] G. STRANG and T. NGUYEN, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, 1997.
- [10] C. Y. SUEN L. XU, A. KRYZAK, "Methods of combining multiple classifiers and their application to handwriting recognition," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 22, no. 3, pp. 418–435, June 1992.