# Comparaison de méthodes de caractérisation du rythme des langues

*Jean-Luc Rouas, Jérôme Farinas*

Institut de Recherche en Informatique de Toulouse
UMR 5505, Université Toulouse III,
118, route de Narbonne, 31062 Toulouse Cedex 4, France
Mél : jean-luc.rouas@irit.fr, jerome.farinas@irit.fr

## ABSTRACT

In this paper, several approaches for language identification by rhythm are described. The authors of these methods tried to find by easiest means if linguistic theories about rhythm can be confirmed. In fact, rhythmic gathering of languages is often hypothesized but quite few experiments were made to prove this hypothesis, and most of those experiments were achieved on quite small corpora. In this paper, we implement the recent methods described by Ramus and Grabe, and using automatic labeling of the speech signal, we perform experiments on a database larger than those usually employed, containing 7 languages with 10 speakers per language. Those approaches are then compared to our own method and the results are discussed.

## 1. INTRODUCTION

Dans cet article sont décrites les principales méthodes existantes de modélisation du rythme pour la classification des langues. Les auteurs de ces méthodes essaient de trouver des moyens efficaces et simples pour confirmer ou infirmer les théories linguistiques portant sur les regroupements rythmiques des langues. En s'appuyant sur ces travaux, nous allons décrire notre technique et la comparer aux précédentes en les testant sur un corpus de parole lue.

Le point de départ des travaux récents sur le rythme est la méthode de description du rythme proposée par Ramus à partir de 1999 [14]. Par la suite, d'autres études se sont inspirées de ces travaux et différentes méthodes de modélisation du rythme plus complexes ont été proposées [5, 4].

Le point faible commun à la plupart de ces méthodes est qu'elles n'ont pour le moment été testées qu'après une segmentation manuelle des voyelles dans le signal de parole, ce qui implique souvent des corpus de taille relativement réduite. Nous proposons ici de vérifier l'efficacité de ces méthodes sur un corpus de taille plus conséquente (le corpus MULTEXT [2] comportant 5 langues auxquelles ont été ajoutées récemment le japonais et le mandarin, enregistrés dans des conditions similaires), ce qui peut être réalisé grâce à une segmentation automatique du signal de parole en segments vocaliques.

Nous décrivons les paramétrisations initiées par Ramus (section 3) et par Grabe (section 4), ainsi que notre approche (section 5). La méthode employée pour détecter automatiquement les voyelles est ensuite décrite (section 6), ainsi que le corpus de test. Enfin, une série d'expériences permet de comparer les performances en identification automatique des langues de chacune des propositions.

## 2. MESURER LE RYTHME D'UNE LANGUE

La mesure du rythme d'une langue n'est pas une tâche aisée. La première étape consiste à donner une définition claire et précise de ce que l'on entend par rythme d'une langue. Malheureusement, un consensus sur une telle définition ne semble pas encore s'imposer.

Cependant, la plupart des chercheurs devraient s'accorder sur le fait que le rythme est lié à l'existence d'un phénomène détectable se répétant au cours d'une phrase.

Une définition plus précise est difficile à fournir. Le rythme peut être considéré comme l'alternance de "segments" dans le signal acoustique. D'après la définition de Crystal ("*Rhythm is the regular perception of prominent units in speech*"), il pourrait être l'alternance d'unités proéminentes avec des unités moins proéminentes, mais définir ces unités est loin d'être une tâche aisée. L'alternance de syllabes accentuées ou non peut résulter en un type de rythme, mais l'alternance de séquences de sons voisés ou non peut être un autre type de rythme, de même pour les alternances de consonnes et de voyelles, de sons longs et courts, etc.

Les approches considérées dans cet article, dont la nôtre fait partie, se focalisent sur la recherche d'unités pertinentes dans le but de caractériser le rythme d'une langue.

L'approche de Ramus (section 3) est la plus simple : l'auteur cherche ici à différencier les langues en prenant comme unités les phonèmes et en regardant sur chaque phrase la proportion en durée des voyelles par rapport aux consonnes, ainsi que la régularité (en terme de variances) de ces durées.

L'approche de Grabe (section 4) considère les voyelles comme éléments caractérisant le rythme. Elle propose des mesures de variabilités des intervalles inter-vocaliques et des durées des voyelles.

Notre approche (section 5) s'inspire des deux précédentes en ajoutant une notion perceptuelle puisqu'elle s'appuie sur des motifs ressemblant à des syllabes. Les mesures proposées sont basées sur la régularité de la durée des voyelles, des durées inter-vocaliques de l'ensemble des consonnes et du nombre de consonnes constituant chaque syllabe.

## 3. APPROCHE DE RAMUS

Dans son article [14], Ramus fait référence à sa méthode de classification des langues selon le rythme. Cette approche est basée sur une conception du rythme de parole comme étant la conséquence de propriétés phonologiques liées à l'identité des langues : la complexité des syllabes, la corrélation entre poids syllabique et accent, la présence ou non de réduction vocalique. Ramus propose une analyse de la complexité syllabique d'une langue afin de déterminer sa classe rythmique. La complexité est mesurée à l'aide d'une segmentation manuelle en consonnes et voyelles. Les paramètres sont :

– %V la proportion (en durée) d'intervalles vocaliques dans la phrase,
– $\Delta$V l'écart-type des durées d'intervalles vocaliques par phrase,
– $\Delta$C l'écart-type des durées d'intervalles consonantiques.

Ces paramètres ont été employés sur un corpus composé de huit langues (anglais, néerlandais, polonais, français, espagnol, italien, catalan et japonais). Quatre locutrices sont enregistrées par langue, chacune lisant cinq phrases.

Sur ces données, les paramètres font apparaître clairement des regroupements entre les langues.

– Le plan (%V,$\Delta$C) fait ressortir trois groupes qui correspondent aux classes rythmiques décrites dans la littérature : anglais, néerlandais et polonais pour les langues accentuelles, espagnol, italien et français pour les langues syllabiques, et japonais pour les langues moraïques. Les langues accentuelles admettent plus de syllabes complexes, donc des groupes de consonnes de taille importante. En conséquence, les langues accentuelles ont un faible %V. Les langues admettant les syllabes complexes admettent aussi les syllabes plus simples, donc les groupes consonantiques sont plus variés. Le $\Delta$C des langues accentuelles est donc plus élevé que celui des langues syllabiques.
– En considérant le plan (%V,$\Delta$V), la variable $\Delta$V est moins directement liée aux classes de rythme. Cependant $\Delta$V apporte une information supplémentaire puisqu'elle suggère que le polonais a des différences importantes avec les autres langues accentuelles.

## 4. APPROCHE DE GRABE

Dans ses articles [5] et [6], Grabe propose une méthode de prise en compte de la durée pour l'identification des langues. L'approche de Grabe est justifiée par le fait que les méthodes statistiques développées récemment ont fourni un support pour les classifications des phonéticiens, mais la catégorisation sans ambiguïté des langues en groupes distincts accentuelles et syllabiques n'a pas émergé. Grabe affirme que le rythme n'est pas relié à des unités phonologiques telles que les intervalles interstress ou les durées des syllabes. Elle propose de mesurer la variabilité de la durée d'intervalles acoustico-phonétiques successifs en employant un paramètre appelé "Pairwise Variability Indice" (PVI). Cette approche est novatrice car le PVI est différent du %V et $\Delta$C de Ramus [14] puisqu'il prend en compte le niveau de variabilité entre les intervalles vocaliques et intervocaliques successifs (normalisé pour les variations de débit).

Le Pairwise Variability Indice (PVI) est défini selon :

$$rPVI = \sum_{k=1}^{m-1} |d_k - d_{k+1}|/(m-1)$$

avec $d_k$ la durée de la voyelle à l'instant $k$ et $m$ le nombre de voyelles dans l'extrait.

Pour s'affranchir des variations liées au débit, une version normalisée de cet indice est définie :

$$nPVI = \sum_{k=1}^{m-1} |\frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2}|/(m-1)$$

Enfin, cet indice peut être calculé en prenant en compte les durées des voyelles ou les durées inter-vocaliques, on notera alors respectivement intra-PVI et inter-PVI.

Les expériences sont menées sur un corpus de 18 langues ou dialectes (thai, hollandais, allemand, anglais britannique, tamoul, malais, anglais de Singapour, estonien, roumain, gallois, grec, polonais, français, catalan, japonais, luxembourgeois, espagnol et mandarin), avec un locuteur par langue sauf pour le français et l'espagnol (7 locuteurs).

Les langues accentuelles (anglais, allemand, hollandais) sont bien séparées des langues syllabiques (français, espagnol). Le Pairwise Variability Indice ne donne cependant pas une séparation des 18 langues en groupes syllabiques et accentuels, mais une répartition suivant un continuum.

## 5. NOTRE APPROCHE

Notre approche pour la modélisation du rythme est basée sur une segmentation en pseudo-syllabes [15]. A partir d'un étiquetage que nous avons rendu automatique, les segments vocaliques sont notés V, les silences #, les autres segments de parole - considérés comme des segments consonantiques - sont étiquetés C. Le signal de parole est segmenté en motifs correspondant à la structure CC...CV. Par exemple, si la séquence CCVVCCVCVCCCVCVCCC est obtenue, elle sera partitionnée en 7 pseudo-syllabes : CCV|V|CCV|CV|CCCV|CV|CCC. Les paramètres utilisés pour caractériser les pseudo-syllabes sont extraits automatiquement. Pour chaque pseudo-syllabe, trois paramètres sont calculés, correspondant respectivement à la durée totale des segments consonantiques (Dc), à la durée totale du segment vocalique (Dv) et à la complexité (Nc) de la pseudo-syllabe qui s'exprime en terme de nombre de segments consonantiques. Les durées sont exprimées en millisecondes. Une illustration de la segmentation et de la paramétrisation des pseudo-syllabes est donnée sur la figure 1.

## 6. DÉTECTION AUTOMATIQUE DES VOYELLES

Nous avons implémenté toutes les méthodes décrites cidessus d'une manière automatique grâce aux outils disponibles à l'IRIT :

– Un découpage du signal de parole en une suite de zones quasi stationnaires, appelées segments, est opéré [1].
– Parmi ces segments sont détectés ceux contenant de la parole.
– Sur l'ensemble des segments de parole, une détection de segments vocaliques [11] permet de détecter les segments contenant des parties vocaliques.

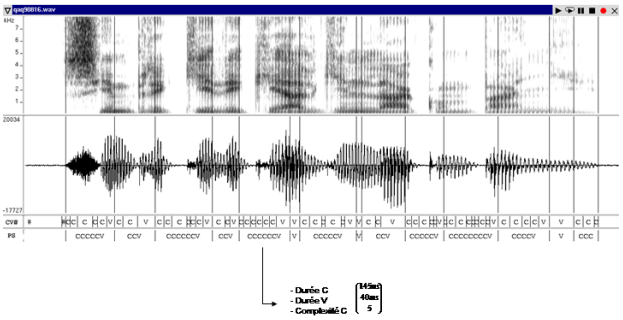

**FIG. 1:** Exemple de caractérisation d'une pseudo-syllabe sur la phrase « Siempre quiere que alguien la acompañe ».

On obtient ainsi un étiquetage automatique et indépendant des langues du signal de parole en segments vocaliques, consonantiques et en pauses.

Les performances de la détection automatique des voyelles sont comparées à celles de différentes approches dans le tableau 1.

**TAB. 1:** Comparaison de différents algorithmes de détection automatique de voyelles.

| Référence | Corpus | Langue | % Err |
|---|---|---|---|
| Pfitzinger & al., 1996 [13] | PhonDatII | Allemand | 12,9% |
| | Verbmobil | Allemand | 21,0% |
| Fakotakis & al., 1997 [3] | TIMIT | Anglais | 32,0% |
| Pfau & Ruske, 1998 [12] | Verbmobil | Allemand | 22,7% |
| Howitt, 2000 [7] | TIMIT | Anglais | 29,5% |
| Pellegrino & André-Obrecht, 1998 [10] | OGI MLTS | Coréen | 28,5% |
| | | Espagnol | 19,2% |
| | | Français | 19,5% |
| | | Japonais | 16,3% |
| | | Vietnamien | 21,1% |
| | | Moyenne | 22,9% |

## 7. EXPÉRIENCES

Les expériences sont toutes effectuées sur le même corpus et avec le même protocole expérimental quels que soient les paramètres étudiés. Nous avons procédé à une représentation graphique des paramètres extraits afin d'appréhender leur pouvoir discriminant. Ensuite, des expériences en identification des langues sont menées, avec l'emploi de modèles de mélange de lois gaussiennes. Les résultats des expériences avec les différents paramètres sont comparés, afin de discuter des avantages et des inconvénients de chaque méthode.

### *7.1. Corpus*

Les expériences ont été effectuées sur la base de parole lue MULTEXT [2] qui comporte cinq langues européennes avec dix locuteurs par langue et quinze phrases par locuteur, auxquelles ont été ajouté le japonais [8] et le mandarin [9]. Les enregistrements ont été effectués dans de bonnes conditions (chambre sourde, échantillonnage à 20 kHz). D'un point de vue rythmique, ce corpus peut se décomposer en trois classes :

– Langues accentuelles : Anglais, Allemand et Mandarin,
– Langues syllabiques : Français, Italien et Espagnol,
– Langues moraïques : Japonais.

Ce corpus a été divisé par nos soins en deux sous-corpus : un corpus d'apprentissage (tableau 2) comportant 8 locuteurs par langue (quatre hommes et quatre femmes) et un corpus de test (tableau 3) comportant 2 locuteurs par langue (un homme et une femme). La durée moyenne d'un fichier est de 22 secondes. Les décisions d'identification des langues sont prises sur chaque fichier.

Les corpus d'apprentissage et de test sont totalement indépendants, les locuteurs et le contenu des textes de ces deux ensembles sont différents.

**TAB. 2:** Description de l'ensemble d'apprentissage

| Langue | Nombre de locuteurs | Nombre de fichiers | Durée Totale | Durée Moy |
|---|---|---|---|---|
| Anglais | 8 | 80 | 24 mn | 18 s |
| Allemand | 8 | 80 | 29 mn | 22 s |
| Mandarin | 8 | 80 | 26 mn | 20 s |
| Français | 8 | 80 | 29 mn | 22 s |
| Italien | 8 | 80 | 30 mn | 23 s |
| Espagnol | 8 | 80 | 27 mn | 20 s |
| Japonais | 4 | 80 | 39 mn | 29 s |
| Total | 52 | 560 | 204 mn | 22 s |

**TAB. 3:** Description de l'ensemble de test

| Langue | Nombre de locuteurs | Nombre de fichiers | Durée Totale | Durée Moy |
|---|---|---|---|---|
| Anglais | 2 | 20 | 6 mn | 18 s |
| Allemand | 2 | 20 | 7 mn | 21 s |
| Mandarin | 2 | 20 | 6 mn | 18 s |
| Français | 2 | 19 | 7 mn | 22 s |
| Italien | 2 | 20 | 7 mn | 21 s |
| Espagnol | 2 | 20 | 8 mn | 24 s |
| Japonais | 2 | 20 | 11 mn | 33 s |
| Total | 14 | 139 | 52 mn | 22 s |

### *7.2. Graphiques*

Les graphiques sont effectués en considérant l'ensemble d'apprentissage. Étant donné le nombre de points à représenter (80 par langue), nous avons décidé par souci de lisibilité de nous limiter à visualiser la moyenne des paramètres pour chaque langue autour de laquelle nous avons dessiné une barre d'erreur ayant pour longueur l'écart-type.

**Paramètres de Ramus** Les distributions des paramètres de Ramus sont représentés sur les figures 2 et 3. Dans le plan (%V, $\Delta$C), on voit apparaître des différences entre certains groupes de langues : le japonais est très à l'écart de toutes les autres langues, avec un $\Delta$C très élevé. Les autres langues considérées ici ont un $\Delta$C assez proche. On voit en revanche apparaître des regroupements entre l'allemand et l'anglais et le mandarin au niveau du %V. De manière générale, on peut distinguer trois groupes, un correspondant uniquement au japonais (langue moraïque), un autre regroupant l'anglais, l'allemand, le mandarin (langues accentuelles) et l'italien (langue syllabique), et un groupe français-espagnol (langues syllabiques).

Le plan (%V, $\Delta$V) ne fait pas apparaître aussi clairement

de tels regroupements. On notera toutefois que le paramètre $\Delta$V permet de séparer le français de l'espagnol.

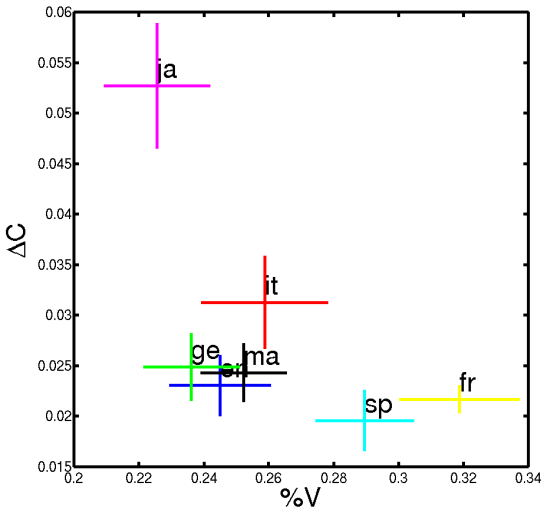


**FIG. 2:** Paramètres %V et $\Delta$C.

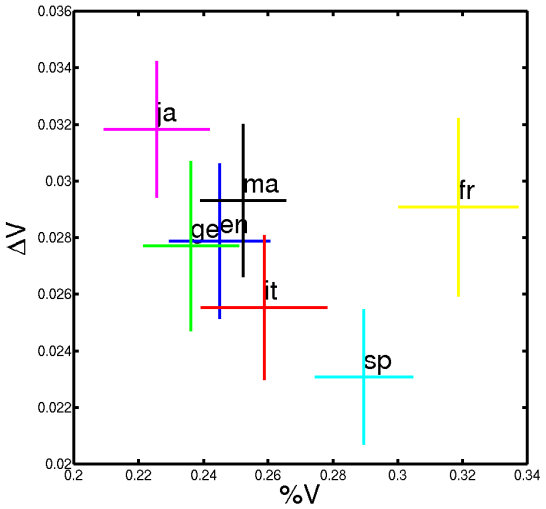


**FIG. 3:** Paramètres %V et $\Delta$V.

Ces résultats sont concordants avec ceux trouvés par Ramus dans ses expériences. Les groupes de langues sont bien retrouvés dans le plan (%V, $\Delta$C), et le plan (%V,$\Delta$V) ne donne pas d'informations liées aux groupes rythmiques, mais il permet de différencier certaines langues à l'intérieur de ces groupes rythmiques.

**Paramètres de Grabe** Les distributions des paramètres de Grabe sont représentés sur les figures 4 et 5.

Dans le plan (Intra-nPVI, Inter-nPVI), on remarque une nette séparation de l'anglais et de l'allemand par rapport aux autres langues, qui sont par contre bien regroupées, avec le mandarin en position intermédiaire.

Les paramètres (Intra-rPVI, Inter-rPVI) montrent une nette séparation du japonais par rapport aux autres langues, principalement au niveau du paramètre *intra raw PVI*.

Ces expériences montrent des résultats moins clairs que ceux obtenus avec le système de Ramus. Les groupes de

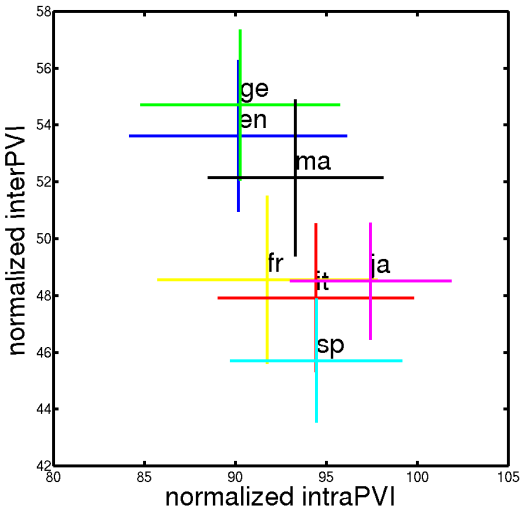


**FIG. 4:** Paramètres *normalized PVI*.

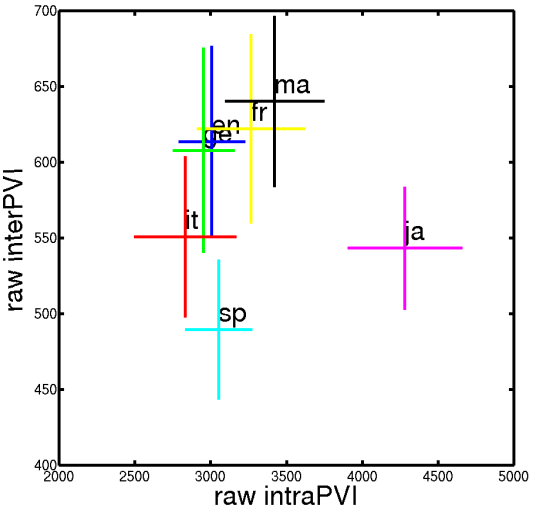


**FIG. 5:** Paramètres *raw PVI*.

langues ne sont pas aussi bien séparés les uns des autres que dans les expériences réalisées dans [5].

Cela est certainement dû à un manque de robustesse vis-à-vis de la variabilité des paramètres employés : les expériences proposées dans [5] n'emploient qu'un seul locuteur par langue.

**Paramètres Pseudo-Syllabes** Les distributions des différents paramètres pour chaque langue sont représentées sur les figures 6 et 7.

La différence principale entre notre approche et les précédentes est que nos paramètres ne sont pas calculés sur la phrase entière, mais sur chaque pseudo-syllabe. Afin de comparer les différents graphiques, nous avons donc gardé uniquement la moyenne de chaque paramètre sur chaque phrase.

On peut remarquer que le paramètre $D_v$ permet de séparer le groupe français-espagnol d'un groupe formé par l'ensemble des autres langues. Le paramètre $D_c$ permet la distinction entre le français et l'espagnol dans le premier groupe, ainsi que celle du mandarin dans le deuxième groupe.

Avec le couple ($D_c$,$N_c$), nous pouvons effectuer un regroupement en classes rythmiques, avec un groupe de langues accentuelles (anglais, allemand et mandarin), un groupe de langues syllabiques (français, espagnol), et un groupe intermédiaire (japonais, italien).

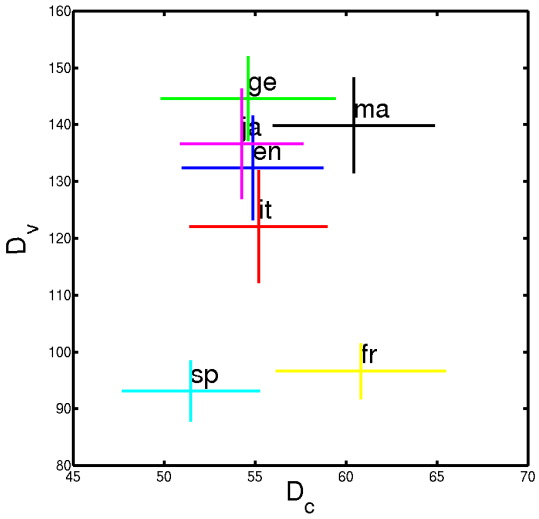


**FIG. 6:** Paramètres $D_c$ $D_v$.

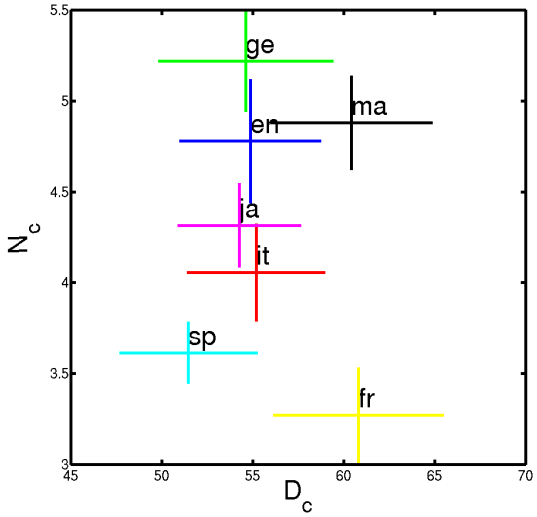


**FIG. 7:** Paramètres $D_c$ $N_c$.

### *7.3. Expériences en identification des langues*

Nous avons ensuite procédé à des expériences en identification des langues pour chacune des méthodes sus-citées. Comme pour les graphiques précédents, le corpus employé est le corpus MULTEXT.

– Tout d'abord les modèles sont estimés à partir des données de l'ensemble d'apprentissage. Il s'agit ici de Modèles de Mélanges de lois Gaussiennes (MMG). Ces modèles sont appris pour différents nombres de lois gaussiennes (2, 4, 8, 16, 32, 64) dans le MMG.
– Pour chaque dimension des MMG, des expériences sont effectuées en utilisant l'ensemble d'apprentissage comme ensemble de développement. Cela permet de déterminer le nombre de lois gaussiennes optimal.
– Une fois que le nombre optimal de lois gaussiennes est déterminé, les expériences d'identification sont effectuées sur l'ensemble de test et les matrices de confusion correspondantes sont données. Des regroupements à l'intérieur des matrices de confusion permettent de visualiser les différents groupes rythmiques.

**Paramètres de Ramus** Les expériences ont montré que le meilleur taux d'identification pour l'ensemble d'apprentissage est de 50,2 % soit 281 identifications correctes sur 560 fichiers, ce résultat est obtenu avec 4 gaussiennes par MMG.

En reconnaissance, l'expérience est menée pour les paramètres donnant le meilleur résultat sur l'ensemble d'apprentissage, c'est-à-dire 4 gaussiennes. Le taux d'identification correcte est de 43,9 % (61 identifications correctes sur 139 fichiers). La matrice de confusion est représentée tableau 4.

**TAB. 4:** Matrice de confusion pour l'ensemble de test ; méthode "Ramus", correct : 43,9±8,3% (61/139).

| | Ang | All | Man | Fra | Ita | Esp | Jap |
|---|---|---|---|---|---|---|---|
| Ang | 1 | 9 | 8 | - | 1 | 1 | - |
| All | 8 | 5 | 7 | - | - | - | - |
| Man | - | 8 | 4 | 1 | - | 6 | 1 |
| Fra | - | - | 1 | 13 | - | 5 | - |
| Ita | - | 3 | 6 | - | 7 | - | 4 |
| Esp | - | - | 1 | 5 | 1 | 13 | - |
| Jap | - | 2 | - | - | - | - | 18 |

Les résultats en identification des langues sont corrects, tout en reflétant les regroupements que nous avons pu mettre en évidence sur les graphiques :

– les langues accentuelles sont bien séparées des autres groupes de langues, mais elles sont confondues entre elles (l'anglais est totalement confondu avec l'allemand et le mandarin). L'italien est également en partie confondu avec ces langues,
– les langues syllabiques sont également bien identifiées (sauf l'italien), on notera aussi des confusions entre le français et l'espagnol,
– le japonais, seule langue moraïque du corpus, est la langue la mieux reconnue.

**Paramètres de Grabe** Les expériences ont montré que le meilleur taux d'identification pour l'ensemble d'apprentissage est de 67,0 % soit 375 identifications correctes sur 560 fichiers. Ce résultat est obtenu avec 8 gaussiennes par MMG.

En reconnaissance, le taux d'identification correcte est de 36,7 % (51 identifications correctes sur 139 fichiers). La matrice de confusion est représentée tableau 5.

**TAB. 5:** Matrice de confusion pour l'ensemble de test ; méthode "Grabe", correct : 36,7±8,0% (51/139)

| | Ang | All | Man | Fra | Ita | Esp | Jap |
|---|---|---|---|---|---|---|---|
| Ang | 5 | 3 | 2 | 4 | 2 | - | 4 |
| All | 6 | 5 | - | 1 | 3 | - | 1 |
| Man | 3 | 1 | 2 | 5 | 1 | 3 | 5 |
| Fra | 1 | 1 | 3 | 10 | 1 | - | 3 |
| Ita | 4 | 3 | - | - | 6 | 4 | 3 |
| Esp | 1 | 1 | 1 | 3 | 5 | 5 | 4 |
| Jap | - | - | 1 | - | 1 | - | 18 |

D'après les graphiques, les paramètres *PVI* ne semblent pas suffisamment robustes pour discriminer les langues

sur des corpus conséquents. Cela est confirmé par les expériences en identification des langues qui ne permettent pas de distinguer quelque regroupement que ce soit. On notera toutefois la bonne performance en identification du système pour le japonais.

**Paramètres Pseudo-Syllabes** Les expériences ont montré que le meilleur taux d'identification pour l'ensemble d'apprentissage est de 68,7 % soit 388 identifications correctes sur 565 fichiers. Ce résultat est obtenu pour 4 gaussiennes par MMG.

En reconnaissance, le taux d'identification correcte est de 66,9 % (93 identifications correctes sur 139 fichiers). La matrice de confusion est représentée tableau 6.

**TAB. 6:** Matrice de confusion pour l'ensemble de test ; Méthode "pseuso-syllabe", correct : 66,9±7,8% (93/139)

| | Ang | All | Man | Fra | Ita | Esp | Jap |
|---|---|---|---|---|---|---|---|
| Ang | 11 | - | 5 | - | 1 | - | 3 |
| All | 1 | 19 | - | - | - | - | - |
| Man | 2 | 5 | 11 | - | 1 | - | 1 |
| Fra | - | - | - | 18 | - | 1 | - |
| Ita | 3 | 1 | 1 | - | 11 | - | 4 |
| Esp | - | - | - | 1 | 10 | 6 | 3 |
| Jap | 1 | - | - | - | 2 | - | 17 |

Ces résultats confirment la pertinence de la modélisation du rythme par les pseudo-syllabes. Toutes les langues sont correctement reconnues, sauf l'espagnol qui est principalement confondu avec une autre langue syllabique, l'italien. L'allemand et le mandarin sont légèrement confondus avec l'anglais, une autre langue accentuelle.

Afin de valider notre approche, un regroupement en fonction des typologies rythmiques des langues est effectué. Le tableau 7 reprend ces résultats. Le taux d'identification correcte des groupes rythmiques est de 84,9 %. Ce résultat confirme les théories linguistiques sur les propriétés rythmiques des langues.

**TAB. 7:** Confusion entre les groupes rythmiques ; Méthode "pseudo-syllabe", correct : 84,9±6,0% (118/139)

| | L. Accent. | L. Syllab. | L. Mora. |
|---|---|---|---|
| L. Accent. | 54 | 2 | 4 |
| L. Syllab. | 5 | 47 | 7 |
| L. Mora. | 1 | 2 | 17 |

## 8. DISCUSSION

Le système proposé par Ramus [14] est le plus simple. Les graphiques permettent de faire apparaître quelques différences entre les langues, mais les expériences d'identification ne montrent pas de résultats très concluants. Cela est peut être dû à une inadéquation de la modélisation employée.

Le système de Grabe [5] se comporte un peu différemment, les graphiques montrant peu de choses intéressantes, on notera tout de même les différences entre les indices normalisés ou non.

Le système que nous avons proposé semble le plus efficace au regard des résultats graphiques. On peut clairement voir apparaître des groupes rythmiques "classiques". Les expériences en identification des langues confirment ces résultats.

L'utilisation de la prosodie pour l'identification des langues reste une tâche difficile. Nous avons toutefois montré qu'il est possible de regrouper de manière automatique les langues en groupes rythmiques correspondants aux théories linguistiques. D'autres expériences devront être effectuées, en prenant en compte de plus nombreuses langues et différentes qualités d'enregistrement.

## RÉFÉRENCES


[1] R. André-Obrecht. « A New Statistical Approach For Automatic Speech Segmentation ». *IEEE Transanctions on ASSP*, 36(1) :29–40, 1988.

[2] E. Campione and J. Véronis. « A multilingual prosodic database ». In *ICSLP*, Sidney, 1998. http ://www.lpl.univ-aix.fr/projects/multext.

[3] N. Fakotakis, K. Georgila and A. Tsopanoglou. « A Continuous HMM Text-Independent Speaker Recognition System Based on Vowel Spotting ». In *Eurospeech*, Rhodes, Greece, September 1997.

[4] A. Galves, J. Garcia, D. Duarte and C. Galves. « Sonority as a Basis for Rhythmic Class Discrimination ». In *Speech Prosody*, Aix en Provence, France, April 2002.

[5] E. Grabe and E. L. Low. « Durational Variability in Speech and the Rhythm Class Hypothesis ». *Papers in Laboratory Phonology 7*, 2002.

[6] E. Grabe, B. Post, F. Nolan and K. Farrar. « Pitch accent realisation in four varieties of British English ». *Journal of Phonetics*, 28 :161–185, 2000.

[7] A. W. Howitt. « Vowel Landmark Detection ». In *ICSLP*, Beijing, China, 2000.

[8] S. Kitazawa. « Periodicity of japanese accent in continuous speech ». In *Speech Prosody*, Aix en Provence, France, April 2002.

[9] M. Komatsu, T. Arai and T Sugawara. « Perceptual discrimination of prosodic types ». In *Speech Prosody*, pages 725–728, Nara, Japan, 2004.

[10] F. Pellegrino. « *Une approche phonétique en identification automatique des langues : la modélisation acoustique des systèmes vocaliques* ». Thèse de doctorat, Université Paul Sabatier, Toulouse, France, December 1998.

[11] F. Pellegrino and R. André-Obrecht. « Vocalic System Modeling : A VQ Approach ». In *IEEE Digital Signal Processing*, Santorini, July 1997.

[12] T. Pfau and G. Ruske. « Estimating the speaking rate by vowel detection ». In *ICASSP*, Seattle, 1998.

[13] H. R. Pfitzinger, S. Burger and S. Heid. « Syllable Detection in Read and Spontaneous Speech ». In *ICSLP*, volume 2, pages 1261–1264, Philadelphia, October 1996.

[14] F. Ramus, M. Nespor and J. Mehler. « Correlates of Linguistic Rhythm in the Speech Signal ». *Cognition*, 73(3) :265–292, 1999.

[15] J-L. Rouas, J. Farinas, F. Pellegrino and R. André-Obrecht. « Modeling Prosody for Language Identification on Read and Spontaneous Speech ». In *ICASSP'2003, Hong Kong, China*, pages 40–43, Vol. I. IEEE, 6-10 avril 2003.