



HAL
open science

Rôle du rythme dans l'identification du locuteur

Dan Zhang

► **To cite this version:**

Dan Zhang. Rôle du rythme dans l'identification du locuteur. Sciences de l'Homme et Société. 2016. dumas-01494403

HAL Id: dumas-01494403

<https://dumas.ccsd.cnrs.fr/dumas-01494403>

Submitted on 23 Mar 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNIVERSITÉ
Grenoble
Alpes

Rôle du rythme dans l'identification du locuteur

ZHANG
Dan

Sous la direction de **Solange ROSSATO**

Laboratoire : Laboratoire d'Informatique de Grenoble

UFR LASSIC

Mémoire de master 2 recherche - 30 crédits – Mention Sciences du langage

Spécialité : Industries de la langue

Année universitaire 2015-2016



UNIVERSITÉ
Grenoble
Alpes

Rôle du rythme dans l'identification du locuteur

ZHANG
Dan

Sous la direction de **Solange ROSSATO**

Laboratoire : Laboratoire d'Informatique de Grenoble

UFR LASSIC

Mémoire de master 2 recherche - 30 crédits – Mention Sciences du langage

Spécialité : Industries de la langue

Année universitaire 2015-2016

Remerciements

Je tiens à remercier ma directrice de mémoire, madame Solange ROSSATO, pour la confiance qu'elle m'a accordée en acceptant d'encadrer ce travail de master, pour ses multiples conseils et pour toutes les heures qu'elle a consacrées à diriger cette recherche. J'ai été extrêmement sensible à ses qualités humaines d'écoute et de compréhension tout au long de ce travail.

Sommaire

Remerciements.....	4
Introduction.....	6
Chapitre 1 : État de l'art et le contexte de recherche.....	7
1.1 Styles de parole.....	7
1.2 Le rythme.....	17
1.3 Les systèmes de RAL.....	20
Conclusion Chapitre 1.....	22
Problématique.....	23
Chapitre 2 : Méthodologies et expérimentations.....	24
2.1 Introduction base de données.....	24
2.2 Mesures des paramètres rythmiques.....	27
2.3 Méthodologies d'analyse d'influence des facteurs.....	32
2.4 Classification en locuteur à partir des données rythmiques.....	33
2.5 Influence du rythme sur l'identification du locuteur.....	36
Chapitre 3 : Résultats.....	39
3.1 Influences des facteurs sur le rythme.....	39
3.2 Validation croisée en locuteur et le test post-hoc.....	45
3.3 Les macro-classes rythmiques et son impact sur le système.....	53
Conclusion.....	55
Bibliographie.....	56
ANNEXE 1.....	59
ANNEXE 2.....	63
ANNEXE 3.....	66
ANNEXE 4.....	69

Introduction

Ce mémoire de recherche fait une étude sur les paramètres rythmiques de la parole, qui permettent de caractériser les locuteurs et qui influencent la performance de système de comparaison de voix pour le forensic dans le cadre d'identification du locuteur.

Nous nous intéressons à la notion de styles de parole, qui est une base théorique pour cette étude. Les styles de parole peuvent être influencés par plusieurs facteurs tels que le contexte social, le rôle du locuteur, etc. Ils relèvent aussi l'idiosyncrasie du locuteur. En particulier, nous avons focalisé la recherche aux paramètres rythmiques de la parole tandis qu'ils sont largement étudiés par les linguistes, les sociolinguistes et les chercheurs du domaine lorsqu'ils font leur étude sur les styles de parole.

D'abord, nous faisons une étude sur la notion de styles de parole. Les facteurs qui influencent les variations des styles de parole et les indicateurs de ces variations sont discutées dans au 1.1. Le rythme de la parole en tant qu'une partie des styles de parole, et sa caractérisation sont discutés au 1.2. Une présentation sur le système de RAL et notre système de base sont introduits au 1.3.

Pour pouvoir obtenir des paramètres rythmiques, nous avons implémenté 7 mesures. La méthode semi-automatique a été utilisé pour faciliter les tâches. Toutes ces méthodes et conceptions d'expérimentations sont présentées dans le Chapitre 2. Une nouvelle notion : macro-classes rythmiques est également introduite. Toutes nos mesures sont basées sur le corpus FABIOLÉ.

Une série d'analyses sont effectuées pour vérifier l'impact des paramètres rythmiques en fonction des différents facteurs. Une classification en locuteur et un test post-hoc ont été faits d'après les 7 paramètres. Nous avons testé l'impact des paramètres rythmiques sur le système comparaison de voix en faisant lien avec les macro-classes rythmiques. Tous ces résultats sont présentés dans le Chapitre 3.

Chapitre 1 : État de l'art et le contexte de recherche

Dans ce chapitre, nous allons faire une description de la notion du style de parole et les paramètres utilisés pour le caractériser, notamment le rythme de parole. Nous nous intéressons aussi à l'impact porté par le rythme sur les systèmes de reconnaissance automatique du locuteur. Donc ces systèmes vont aussi être présentés, car ils servent en tant que la base et le contexte de la recherche.

1.1 Styles de parole

1.1.1 Les définitions du style de parole

Le style de parole est un concept qui est traité en linguistique ainsi qu'en sociolinguistique. Au niveau linguistique, la définition de style de parole est très générale et large. Chaque linguiste a son explication et compréhension pour le style de parole. Eskenazi (1993) a cité deux linguistes dans son travail pour expliquer la définition du style de parole : « For Joos (1968) and Zwicky (1972) , styles may be defined along an axis of casualness. For Zwicky, four degrees of casualness can be defined whereas Joos found 5 (intimate, casual, consultative, formal, frozen)¹ The one-dimensional aspect and fixed levels of these two authors were a first effort at organising various speaking styles»². La définition pour le style de parole de Joos et Zwicky se fondent tous sur l'échelle de familiarité. Dans le même travail, Eskenazi a aussi développé sa propre définition pour le style de parole. D'après lui, « Style reflects the action of the environment upon the individual and the individual upon the environment. It is the perception of the various status levels of his listener and of the type of situation in which he finds himself. It is also a projection of himself, his background, and is a setting of the type and tone of conversation he wishes to have. All of this is a mixture of conscious and unconscious (voluntary/involuntary) effort on his part and is not always perceived in the same way as it was intended»³. Compatible à cette définition, elle a aussi souligné la dimension d'après laquelle le style de parole peut varier :

1.« intelligibility-oriented »⁴ : the degree of clarity that the speaker intends his message to have is the first dimension.

1 Ma traduction : intime, quotidien, consultatif, formel, figé.

2 Ma traduction : Pour Joos (1968) et Zwicky (1972) , styles pourraient être définis le long d'un axe de familiarité. D'après Zwicky, 4 degrés de familiarité peuvent être définis alors que Joos en a trouvés 5. (intime, quotidien, consultatif, formel, figé) . L'aspect unidimensionnel et les niveaux fixés de ces deux auteurs sont des premiers essais en organisation des styles de parole variés.

3 Ma traduction: Le style reflète l'action de l'environnement sur l'individu et celle de l'individu sur l'environnement. C'est la perception des niveaux de statuts variés de son interlocuteur et le type de situation où il trouve celui du sien. Il est également une projection de lui-même, de son contexte. Le style est aussi le ton de conversation que le locuteur souhaite à avoir. tous ces éléments sont une mixture de l'effort volontaire et involontaire de la part du locuteur. Par contre, ces styles ne sont pas toujours perçus et compris de la même manière par les interlocuteurs tels qu'ils étaient produits.

4 Intelligibilité-orienté

2. « familiarity »⁵ : the expression of style may change greatly according to the speaker's familiarity with the listener.

3. « social strata »⁶ : the degree of cultivation that the speaker and listener wish to accord their conversation.

Cependant, quand Iván Fónagy donne sa définition sur le « phonostyle », il l'a décrit comme : « une science qui étudie des phénomènes sonores particuliers, des variantes associées soit à la parole de différents groupes sociaux majeurs (classes sociales, sexes, groupes d'âge) ou mineurs (groupes professionnels, groupuscules), soit à certaines situations en fonction de leur valeur stylistique et du message supplémentaire qu'ils ajoutent au message proprement dit ». (Fónagy, I. 1977). Nous constatons que Fónagy donne sa définition plutôt en fonction des facteurs sociaux et situationnels. De plus, un autre linguiste, Pierre Léon, lorsqu'il définit les phonostyles, tient compte de ces différents facteurs : les phonostyles sont des styles sonores « tels qu'ils sont perçus en tant que caractéristiques d'un individu (jeune, vieux, homme, femme), d'un groupe social (prolétaire, bourgeois), ou d'une circonstance particulière (discours politique, sermon, etc.) » (1993: 3).

Dans le domaine sociolinguistique, la notion « style de parole » a été aussi beaucoup discutée. Pour, William Labov (1972), la parole est un fait individuel qui ne relève que de l'individualité du locuteur : « chacun a son style ». Il retient trois ordres de facteurs pour distinguer les différents styles : la relation entre le locuteur et son interlocuteur, le contexte social et la thématique de la communication. William Labov est parmi les premiers sociolinguistes qui étudient la notion de style de parole.

Le style de parole dépend en grande partie du type de discours ou genre de l'oral. "La problématique des « genres » de l'oral, pour remonter à l'Antiquité, appartient également au langage courant, qui dispose de nombreux termes pour caractériser « une conversation, une discussion ou un débat, du bavardage ou du marchandage, une interview, un entretien ou une consultation, un cours ou un discours, une conférence ou une plaidoirie, un récit ou un rapport, une confidence ou une dispute, etc.»" (Kerbrat-Orecchioni, 2003).

De Mareüil (2014) a fait un résumé de l'étude des autres auteurs du domaine sur la définition du style de parole. Il a souligné le rôle du locuteur comme un facteur très important dans son article « Mais un éventail de styles bien plus large existe : celui du prêtre, celui du politicien, celui du journaliste sportif, celui du gendarme

5 familiarité

6 Couches sociales

(stéréotype naguère conventionnellement caricaturé avec un accent méridional), celui de l'acteur de théâtre, celui du DJ, etc. Des travaux dans ce domaine, en traitement automatique, ont été menés : observations à l'appui, ils montrent en particulier des débits de parole (taux d'articulation et de phonation) très différents entre un journal parlé et un discours officiel ». Il a aussi donné fait une remarque sur l'importance de la situation de communication : « Nous nous intéresserons davantage au style vu comme un ensemble d'ajustements à une situation, dépendant entre autres que la familiarité que l'on a avec son interlocuteur et le sujet abordé, qui font que l'on ne parle pas de la même façon dans toutes les circonstances de la vie (Bolinger, 1989; Lindblom, 1990; Eskénazi, 1993). » (Philippe Boula de Mareüil, 2014).

Cependant, avant de chercher les paramètres intéressants pour la caractérisation du style de parole, il existe quelques défauts à prendre en compte. Joaquim Llisterri (1992) considère que : "un but plus ambitieux devrait être réfléchi, la standardisation de la définition des terminologies conventionnelles du style de parole,...au moins chaque chercheur doit utiliser le même terme pour définir la même entité." En 2016, la même remarque peut être énoncée.

Ainsi, il est probable que les terminologies utilisées ci-dessous pourraient être les synonymes. Tel que le « contexte social » de Labov et « groupes sociaux » de Fónagy.

Notons que le degré de la familiarité, les circonstances de communication et le contexte social, l'âge et le rôle du locuteur sont des facteurs plus utilisés pour définir la notion de style de parole.

En conclusion, le style de parole est une notion complexe, il relève les caractéristique des circonstances dans lesquelles la parole est produite, et les difficultés à définir. Cette notion illustre sans doute son omniprésence.

1.1.2 Caractériser les styles de parole

Selon les études faites, le style de parole dépend beaucoup des facteurs extérieurs comme les circonstances de communication. Alors, d'après quels traits de la parole peut-on caractériser un style de parole? Les différents paramètres pourraient être retrouvés dans les études faites sur le style de parole. Nous allons les montrer en donnant des exemples.

Simon et al.(2010) ont fait une étude sur « comment les paramètres prosodique varient-ils en fonction des styles de parole ». Leur étude vise à découvrir si certains paramètres prosodiques (le débit, l'intonation, les accents) varient systématiquement en fonction de

facteurs situationnels et si cette variation permet de définir des styles de parole spécifiques en analysant un corpus de la langue française de 70 minutes comprenant 7 activités de communication. Limité par leur corpus, ils ont établi une catégorisation à priori des échantillons, sur la base de facteurs situationnels. Ils ont classifié les échantillons en fonction des différents paramètres observables de la situation de communication. Trois juges ont noté chaque style selon les dix critères de e Koch & Oesterreicher (2001). Pour chaque critère, ils attribuent une valeur entière entre 1 et 5 selon l'échelle de la familiarité de la parole, c'est-à-dire, de « la langue de la proximité » à « la langue de distance ». Ils ont retenue les variables prosodiques suivants : « variables temporelles (taux d'articulation, débit de parole et débit d'articulation) ; variables mélodiques (étendue du registre et agitation mélodique F0) ; accentuation (proportion de syllabes proéminentes) ; pauses pleines (marquant des hésitations). »

Donc, classifier les styles de parole par échelle ou degré sera un des critères utiles pour la recherche suivante. Ce type de paramètres correspondent bien à la définition du style de parole proposée par Labov et Eskenazi. Tous les deux auteurs ont marqués l'importance de facteurs de situations communicationnelles pour distinguer les styles de parole.

De plus, d'après Goldman et al.(2009) ont fait une recherche sur la discrimination de phonostyles, « c'est-à-dire, styles de parole perçu comme identifiant une situation de communication via un genre, une image acoustique typifiée». Ils ont fait une analyse sur un corpus C-PROM comprenant 6 genres de parole : lecture, conférence scientifique, interview radiophonique, journal parlé, récit conversationnel et discours politique. L'annotation phonétique et grammatical sont semi-automatique. La proportion des syllabes proéminente et les macro-unités (séparé par des longue pauses) sont mesurées en tant qu'indicateur de changement du style. Eskénazi a également fait remarque sur les indicateurs dans la parole pour caractériser les changements de style. Il a conclu au niveau acoustique-phonétique : consonant, formant et réduction de voyelles ; au niveau prosodique : la durée des syllabes, des voyelles, des silences, le débit de parole ; la fréquence fondamentale;amplitude, rythme, pauses, spectraux maximum ; ou bien, de niveau encore plus élevé : insertion de phonème, assimilation et élision, classement de mots ; redondance ou intelligibilité. Tous ces indicateurs peut démontrer un changement de style de parole. Ils peuvent être mesurés pour une caractérisation de style.

Dans leur analyse phonostylistique du discours radiophonique, les deux auteurs, Hupin et Simon (2009) ont mesuré le changement du phonostyle dans les discours de radio en fonction des situations de communication . « Notre étude vise principalement à comparer

différents phonostyles, qu'il s'agisse des styles d'un même journaliste dans des situations différentes ou des styles de journalistes qui officient sur des chaînes radio différentes. » À travers l'analyse sur le débit de parole, l'étendue de la fréquence fondamentale, du registre moyen (F0 moyen), de la proportion de syllabes statiques par rapport aux syllabes dynamiques, de l'agitation mélodique (Il s'agit du chemin moyen parcouru par la f0 en une seconde. Elle se mesure en demi-tons) et du pourcentage de syllabes proéminentes. Leur conclusion est qu'il y a une mise en fonction professionnelle et qu'elle tend vers une standardisation du ton adopté durant les journaux.

Par ailleurs, Brognaux et Drugman (2014), ont fait une étude sur l'impact de la situation communicationnelle sur la variation phonétique. Ils ont effectué une analyse sur un corpus français qui a une durée de 300 minutes et contient dix situations communicatives différentes (commentaires de sports, journal télévisé, discours politique, etc.), avec trois types de discours: lecture/spontané, média/non-média et expressif/non-expressif. Les aspects analysés sont l'élision de schwa, épenthétique de schwa, élision de la consonne finale et la liaison. De plus, ils ont aussi étudié la corrélation entre les caractéristiques phonétiques et rythmiques. Pour mesurer le rythme : « It focusses on three rhythmic measures: the articulation rate, the mean duration of inter-pausal units (IPU) and the proportion of prominent syllables ». Ils ont pu montrer que la parole spontanée et non-média expose significativement un plus haut pourcentage au niveau de la variation phonétique que la lecture standard. La parole de média occupe un taux similaire que la lecture standard. Par contre, l'analyse de la corrélation entre les caractéristiques phonétiques et rythmiques montre que la variation phonétique dépend essentiellement des types de discours différents.

Tableau de résumé des paramètres pour la caractérisation des styles de parole
Le débit de parole, l'intonation, l'accentuation ;
Taux d'articulation, débit d'articulation, étendue du registre et agitation mélodique F0, pauses pleines marquant des hésitations ;
Consonant, formant et réduction de voyelle ;
La durée des syllabes, des voyelles, des silences ;
Amplitude, rythme, spectraux maximum...;

Tableau 1 : résumé des paramètres pour la caractérisation des styles de parole

Les paramètres qui caractérisent le style de parole pourraient être en même temps les

indicateurs qui le représentent. Encore une fois, nous constatons que les facteurs de ces paramètres sont pas uniques : ils peuvent être les situations communicationnelles différentes tel que professionnel ou familial ainsi que les types de parole tel qu'un entretien ou un discours devant le publique.

Classification de styles	Auteurs
proximité au public	Baptiste Hupin et al. (2009)
lecture, interprété, parole spontanée	Judit Bóna et al. (2008)
Degré de familiarité et les relations entres locuteurs	Nick Campell (2007)
Messe catholique, politique, journalistique, commentaires sportifs	Nicolas Obin et al. (2011)
Type d'activité de parole et circonstances situationnelles	Jean-Philippe Goldman et al. (2009)
Discours monologue et dialogue (en fonction de nombre de locuteurs)	Caroline Émond (2010)
Intime Familier Consultatif Formel Figé	Joos Martin (1972)
Variation de situations communicatives (définie d'après 3 type de traits « binaire ») : Média/non-média Expressif/non-expressif Lecture/spontanée	Sandrine Brognaux et al. (2014)
Parole spontanée lecture oralisée	Philippe Boula de Mareüil (2014)
(PhonoGenre) DIDA : discours pédagogique NARR : narration spontanée PARL : (au parlement français) questions et réponses READ : lecture RELG : sermons RPRW : radio SPOR : commentaires sportifs WISH : vœux de nouvel aux présidentiels WFOR : météo	Jean-Philippe Goldman et al.(2014)
Bulletin d'informations radiophoniques Conférences radiophoniques Conversations radiophoniques	Rena Nemoto (2012)

Tableau 2 : tableau de résumé des styles de parole d'après différents critères

1.1.3 Les paramètres prosodiques utilisés

Nous avons discuté la définition du style de parole au 1.1.1 et comment les caractériser au 1.1.2 dans lequel nous avons remarqué l'importance de la prosodie en discrimination de styles de parole. Donc, dans cette partie nous allons mettre en évidence les paramètres utilisés pour la notion de style de parole en donnant quelques études comme exemples.

Deux autres auteurs, Émond et Ménard (2010), ont fait une recherche sur les marques

prosodiques des styles de parole dans les télé-journaux québécois. Leur objectif est de « décrire les marques prosodiques des situations de discours monologue et dialogue sur les plans intonatifs (hauteur de la voix, contours intonatifs), rythmique temporel (débit) et rythmique accentuel (nombre et types d'accents, hauteur des syllabes accentuées) ». Dans leur travaux, contrairement à la classification traditionnelle « lue » ou « spontanée », ils ont introduit les types de discours en monologue ou dialogue en fonction de nombre de locuteur. Corpus issu de Société Radio-Canada, une chaîne publique de l'état, dure 180 minutes, comprends trois locuteurs différents. Au niveau de l'analyse acoustique, le modèle de regroupement prosodique de Cedergren et Perreault (1994) a été employé.. « Ce modèle comporte trois niveaux enchâssés dont l'unité minimale est la syllabe. »

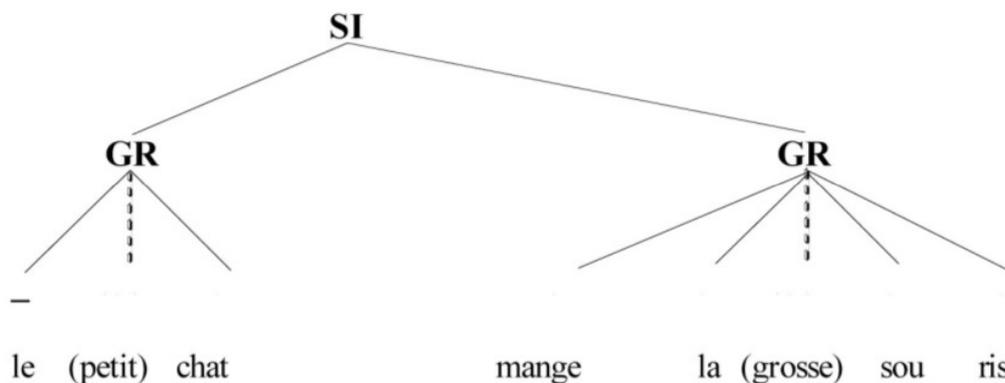


Figure 1 : modèle de regroupement prosodique de Cedergren et Perreault (1994)

En lien avec ce modèle, les séquences retenues sont analysés à l'aide de Praat, sous cinq paliers : phones, syllabes, rythmes, accents et autres. « Les propriétés acoustiques ont été examinées : tel que le registre de F0 , l'étendue de F0 , le nombre de syntagmes intonatifs ainsi que le nombre et le type d'accents ». Généralement, un changement de prosodie a été trouvé en fonction de situations communicationnelles différentes, en excluant un cas particulier. Ceci peut être dû à une manifestation de caractéristiques individuelles du locuteur dans les deux discours à cause des raisons environnementales, par exemple un mauvais fonctionnement d'équipement pendant la production de la parole.

Une étude récente, faite par Goldman et al. (2014), sur la variation phonétique de la parole basant sur un corpus français d'une durée huit heures, a montré que les situations de communication peuvent être utilisées pour classifier les styles de parole. De plus, une manifestation des effets des types de parole/situations ont été trouvé dans la prosodie. Les mesures appliquées ce sont : « temporal measures (e.g. articulation rate); pitch measures (e.g. pitch register and movement); syllabic prominence measures (e.g. percentage of prominent syllables in various positions); correlational measures (e.g. percentage of

accentual-group-initial prominent syllables) ». Leur analyse affirme que la prosodie varie en fonction des situations communicationnelles alors les indicateurs de ces variantes sont visibles au niveau phonétique. Leur étude illustre aussi que le type de l'oral est un critère intéressant pour l'étude des styles de parole.

Nick Campbell (2007), a effectué une étude pour trouver les différences dans les styles de parole d'un homme japonais en fonction de son interlocuteur en analysant la prosodie. L'étude a été menée sur un corpus conversationnel de la langue japonaise, montrant que les quatre caractéristiques de la prosodie : la durée, le hauteur, l'intensité et le F0 varient tous selon les différents interlocuteurs et les différences au niveau de la familiarité sur une période de temps déterminée avec le même interlocuteur. Il a employé comme mesure : le débit de parole, le Pitch, la valeur moyenne de l'énergie, les caractéristiques de spectrum. « By factoring the analysis according to differences in interlocutor as well as by differences in time, or sequence of the conversations, we were able to show that the changes are not a result of time-related changes, such as tiredness or ill-health, but that they correlate more with differences in interlocutor and with development of the individual relationships ». Dans la partie discussion et conclusion, il a souligné qu'il a pu trouver une corrélation entre le changement de style de parole de locuteur et son interlocuteur. En prenant compte de cette étude, nous pouvons bien noter que l'interlocuteur est aussi un facteur très important qui provoque le changement de style de parole au niveau prosodique. À partir de tous ces exemples d'études portés sur les paramètres prosodiques du style de parole, nous avons souligné que la recherche sur la notion du style de parole est bien étendue. Depuis 1950s, les sociolinguistes commençaient déjà à chercher les facteurs situationnels qui ont des effets sur la production de la parole. Pour conclure, nous constatons que les chercheurs du domaine ont mis énormément attention au lien entre la prosodie et les facteurs définis par les linguistes, sociolinguistes ou bien eux-même. Quelques mesures en commun sur la prosodie pour but de distinguer les styles de parole sont appliquées : le débit de parole, la fréquence fondamentale, le rythme, les syllabes proéminentes, le pitch, la durée de la parole, le hauteur (représenté par la valeur moyenne de registre) et l'énergie.

Les recherches sur la notion du style de parole sont faites avec méthodes diverses, sur les facteurs divergents, sur les langues différentes et pour buts différents. Il existe des linguistes qui s'intéressent au niveau de la phonétique, par exemple Fónagy, des sociolinguistes qui s'intéressent aux facteurs situationnels, sociaux, comme Labov. Il existe aussi des auteurs qui établissent le lien entre la production de la parole et un fait social

concret, comme Romain(2014). Les styles de parole pourraient être définis par des facteurs distincts : tel que le genre, l'âge du locuteur ainsi que les situations communicationnelles, le contexte social ; de plus, il est également indissociable avec le type de l'oral.

Les indicateurs pour caractériser les styles de parole sont pareillement hétéroclites : la prosodie a tenu une position prédominante (cf : Tableau 1 ; cf : 1.1.3). Comme la prosodie est une notion complexe parce qu'elle comprends tous les traits suprasegmentaux de la parole, les chercheurs la mesurent souvent de manière différentes. Nous constatons que le débit de parole est largement utilisé pour caractériser le rythme de la parole pour but de la caractérisations des styles de parole en fonction de la prosodie. Par contre, outre que le débit de parole (représenté par le taux d'articulation des fois), peu d'autres mesures sont appliquées pour découvrir les paramètres rythmiques. Alors, nous disons que l'importance du rythme a été remarqué par les auteurs (Goldman et al.; Émond; Campbell), les mesures sont pas suffisantes.

1.2 Le rythme

Dans la partie 1.1, nous avons introduit la définition du style de parole au niveau linguistique et sociolinguistique ainsi que ses indicateurs dans la parole en mettant évidence des traits prosodiques. Par contre, comme nous avons indiqué à la Conclusion 1.1, l'attention a mis sur le rythme de la parole alors que les mesures appliquées sont monotones. « What does seem to be clear is that rhythm is useful to us in communicating: it helps us to find our way through the confusing stream of continuous speech, enabling us to divide speech into words or other units, to signal changes between topic or speaker, and to spot which items in the message are the most important. » (Peter Roach, 2001). Alors, nous allons introduire par la suite la notion complète du rythme de la parole en linguistique et les études relatives.

1.2.1 Définition du rythme de la parole

Le rythme est un concept qui porte différents sens pour les domaines différents. Nous allons chercher sa définition spécifiquement pour le domaine linguistique puis que nous faisons la recherche sur la parole.

« Every language may be assigned one of three rhythmical types: stress-timed (where the durations of the intervals between stressed syllables is relatively constant), syllable-timed (where the durations of successive syllables are relatively constant) and mora-timed (where the durations of successive morae are relatively constant) »⁷.

« In phonetics, *rhythm* is the sense of movement in speech, marked by the stress, timing, and quantity of syllables ». ([Glossary of Grammatical and Rhetorical Terms](#)). Ceci est la définition du rythme en phonétique.

D'une autre façon, le rythme est défini comme « Impressions analogues que créent différents éléments prosodiques qui reviennent régulièrement dans la chaîne parlée ». (Office québécois de la langue française, 2000).

En conséquence, nous constatons que la définition du rythme en linguistique est aussi complexe : il implique plusieurs paramètres de la parole, tel que la durée, le débit, les syllabes. Par la suite, nous allons discuter des caractéristiques rythmiques de la parole en montrant quelques études.

⁷ Ma traduction: ce qui nous semble intéressant est que le rythme est un fait utile à la communication, il nous donne les traits pour distinguer les confusion dans la parole spontanée, nous permet de segmenter la parole en syntagme ou les autres unités en mettant évidence aux sens importants dans la parole.

1.2.2 Les caractéristiques rythmiques de la parole

Comme nous avons donné la définition du rythme de la parole, nous allons parler ici des paramètres qui le caractérisent et les facteurs qui l'influencent. D'abord, nous sommes conscients que les caractéristiques rythmiques de la parole sont observables à partir de ces traits de la parole : le débit, les pauses lors du discours, le ralentissement ou l'accélération lors de la production de la parole, l'interaction entre le voisement et la pause.

Stanley et Bond (1981) ont effectué une expérimentation en perception des fréquence et intensité de la parole pour montrer leur influence sur la perception de débit de parole. Un morceau d'enregistrement de 30 secondes, qui est segmenté en 9 stimuli est utilisé comme corpus. 88 experts sont demandé d'annoter les stimuli par « lent » ou « rapide ». Les résultats montrent que les deux paramètres étudiés ont tous l'influence sur la perception de débit de parole. Cela est parmi les premiers à la découverte des caractéristiques rythmiques.

Suite au développement des approches automatiques, le rythme de la parole peut être utilisé pour distinguer les langues. Dellwo et al. (2007) ont implémentés des mesures rythmiques pour classifier des langues : contrairement aux mesures traditionnelles, ils ont mesuré le %VO (le pourcentage sur lequel un signal de parole est voisé) au lieu de %V (le pourcentage sur lequel le signal est vocalique) ; À la place de ΔC (écart-type des durées des intervalles consonantes), ils ont mesuré le ΔUV (écart-type des durées des intervalles non-voisés). Ils ont pu démontrer l'efficacité des mesures basées sur les segments voisés et non voisés pour une classification des langues en groupes rythmiques. De plus, leur étude illustre également l'importance des durées des segments voisés et non voisés en tant que paramètres rythmiques. Une autre étude du même auteur, Dellwo(2012) se concentre sur les caractéristiques rythmiques dans les signaux de parole. L'objectif de l'étude est de démontrer comment les caractéristiques de la parole sur le domaine temporel expliquent-elles l'idiosyncrasie du locuteur. « ...we study how temporal characteristics of human speech (e.g. segmental or prosodic timing patterns, speech rhythmic characteristics and durational patterns of voicing) contribute to speaker individuality ». Similaire à son étude en 2007, le %V et %VO sont mesurés. Ils ont montré que les mesures rythmiques basant sur les caractéristiques de la parole sur le domaine temporel relèvent de très haut niveau les différences entre locuteur.

Certaines études du rythme sont faites sur les pauses dans la parole. Ces auteurs sont bien

d'accord sur le fait qu'une pause transmette aussi des informations importantes comme l'attitude ou l'émotion du locuteur. Duez (1999) a prouvé que les pauses dans la parole des hommes politiques ont des effets sémantiques.

Par ailleurs, Kolly et al. (2015) ont appliqué des mesures de pauses pour investiguer les variabilités inter et intra locuteurs lors de leur production de L1 (Suisse allemand) et L2 (anglais/français). Ils ont montré que le nombre de pauses et les durées de pauses sont spécifiques au locuteur.

Leemann et al. (2013) a étudié comment les caractéristiques suprasegmentales du domaine temporel de la parole relève-elles l'idiosyncrasie du locuteur dans un contexte de forensic sur la langue Suisse allemand. Ils ont amené les expérimentations sur deux types discours : parole lu et spontanée. D'après leur conclusion, les variabilités inter-locuteur sont trouvées en fonctions des deux types.

Pour conclure, le rythme est un composant de la prosodie au niveau linguistique et, en même temps, un composant très important du style de parole. De plus, il relève aussi l'idiosyncrasie du locuteur.

1.3 Les systèmes de RAL

La reconnaissance automatique du locuteur (RAL), thématique de recherche apparue dans les années 1960s, connaît maintenant des grands succès. Ces technologies sont largement utilisées dans le domaine Forensic pour un but judiciaire ou le domaine commercial tel que l'identification du locuteur via le téléphone dans les services de banque.

Au début, les études d'identification du locuteur sont faites en grande partie par des approches manuelles. Au fur et à mesure du développement des techniques, cette procédure est déjà devenue automatique.

Les systèmes de RAL, développés à travers plusieurs années dans le domaine académique et industriel, ont maintenant des performances assez stables et de haut niveau. U

Pour évaluer la puissance des technologies en RAL, l'Institut national des normes et de la technologie⁸ (National Institut of Standards and Technology en anglais) propose une campagne d'évaluation qui a lieu tous les deux ans.

NIST propose d'exprimer le score de sortie du système RAL, qui est une estimation du log likelihood ratio, LLR. La formule est :

$$LLR = \log (\text{pdf} (\text{data} | H1) / \text{pdf} (\text{data} | H2)) ;$$

Où data est une paire d'enregistrement (pièce question et pièce de comparaison en Forensic. H1 étant hypothèse que les deux enregistrements sont issues du même locuteur. H2, de deux locuteurs).

Le système que nous allons utiliser dans cette étude est : LIA-ALIZE/SpkDet.

Voici une description du système.

« Acoustic features are composed of 19 LFCC parameters (cepstral parameters using a linear scale) issued from a frequency window restricted to 300-3400 Hz, its derivatives, and 11 second order derivatives. A (file-based) normalization process is applied, so that the distribution of each coefficient is 0-mean and 1-variance for a given utterance. The Universal Background Model (UBM) has 512 components and is trained by EM/ML. The UBM and the total variability matrix, T , are trained on Ester 1&2, REPERE and ETAPE databases on male speakers that do not appear in FABIOLÉ database. They are estimated using “7, 690” sessions from “2, 906” speakers whereas the inter-session matrix W is estimated on a subset (selected by keeping only the speakers who have pronounced at least

8 <http://nist.gov/itl/iad/mig/sre.cfm>

two sessions) using “3, 798” sessions from “672” speakers. The dimension of the I-Vectors in the total factor space is 400. »⁹(Ajili et al, 2016).

Dans cette partie, nous avons donné une vue sur les systèmes RAL et son évaluation, qui sert à justifier sa performance. De plus, le système de base pour notre recherche est aussi présenté ainsi que ses méthodes techniques. Nous avons aussi fait une étude sur les paramètres en remarquant la manque d'attention sur les traits rythmiques de la parole dont nous avons noté importance à la première partie en introduisant la notion style de parole.

9 En général, 19 LFCC sont pris comme paramètres, normalisation impliquée.

Conclusion Chapitre 1

En conclusion, la notion de styles de parole est largement utilisée. Le rythme est un composant de la prosodie de la parole. Il est un paramètre important des styles de parole. Il montre en même temps l'idiosyncrasie du locuteur et influencé par les situations communicationnelles. Par contre, l'étude portée sur les paramètres rythmiques sont plus ou moins limités. Nous allons, par la suite, introduire notre problématique qui centralise particulièrement aux paramètres rythmiques et ses effets relatifs.

Problématique

D'après les trois dernières parties 1.1, 1.2 et 1.3, nous avons comme problématique : quelle information sur le locuteur apportent les paramètres rythmiques ? De plus, dans quelle mesure le style de parole interagit-il avec l'identification du locuteur ?

Chapitre 2 : Méthodologies et expérimentations

2.1 Introduction base de données

2.1.1 Base de données : Le projet et corpus FABIOLÉ

La recherche est faite dans le cadre du projet FABIOLÉ : Fiabilité en biométrie vocale, subventionné par l'ANR. Ce projet s'inscrit dans le domaine de l'authentification biométrie vocale et s'attache plus particulièrement à explorer l'apport des « connaissances phonétiques » dans ce cadre applicatif.

L'objet de ce projet est dans un premier temps, à chercher les influences des variations inter et intra locuteurs sur la performance du système de la comparaison de voix, soit la comparaison de deux signaux, surtout dans le contexte de forensic.

Pour effectuer la recherche sur les variabilités inter et intra locuteur, il est nécessaire d'avoir un corpus qui contient un grand nombre de locuteurs.

Le corpus FABIOLÉ, créé sous le projet FABIOLÉ, est un corpus oral de la langue française qui peut satisfaire au maximum à cette exigence.

FABIOLÉ est un nouveau corpus oral du français qui a collecté 130 locuteurs masculins de métiers divers. Ce corpus a été extrait des programmes de télévision ou de radio français. La construction de corpus est principalement pour but d'investiguer les variabilités de la parole inter et intra locuteur dans un contexte de comparaison de voix. Pour cela, les autres paramètres sont strictement contrôlés : les variabilités de canal de communication sont réduites au maximum, car les extraits viennent que des programmes télévisés ou radiophoniques. Donc l'effet de canal est suffisamment faible pour être négligé. La qualité de signaux est à un très haut niveau pour limiter les influences de bruits même si certains signaux contiennent toujours la musique de fond. Par contre, ceci est inévitable pour toutes les collections de la langue parlées dans une situation communicationnelle de la vie quotidienne. Chaque extrait a une durée au minimum de 30 secondes de parole pour que les signaux soient comparables. 30 locuteurs sont les locuteurs target (Série T : target) et 100 sont les imposteurs (Série I : impostors). Chaque locuteur dans la série T occupe au moins 100 fichiers sonores. La série I dont chaque locuteur a un seul extrait sonore est préparée pour les expérimentations imposteurs, donc chaque locuteur dans la série I occupe un seul fichier sonore.

Le corpus FABIOLÉ contient au total 30 heures de parole avec une dispersion d'informations de locuteurs avec une étendue assez vaste. Comparé avec les autres corpus

oraux de la langue française, FABIOLÉ a un atout au niveau de taille et de diversité.

2.1.2 Émissions et rôle du locuteur

Les locuteurs de FABIOLÉ ont des métiers différents : ils sont, soit des journalistes, comme « Benjamin Sportouch » et « Olivier Truchot », ou des hommes politiques comme « Manuel Valls » et « Jean-Marc Ayrault », etc. Nous supposons que les rôles joués par les locuteurs ont une influence importante sur le style de parole. D'après ces informations de métiers des locuteurs, un effet de rôle va être étudié par la suite.

Voici le diagramme pour la dispersion des métiers des locuteurs :

Statistique des métiers de locuteurs - Série Target

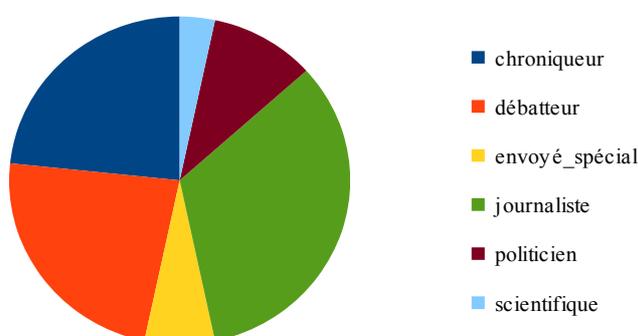


Figure 2 : diagramme de la dispersion des métiers des locuteurs

Outre que l'effet de « rôles » des locuteurs, le corpus FABIOLÉ nous a aussi donné la possibilité à étudier les effets d'émissions, car toutes les informations d'émissions peuvent être retrouvées. Cependant, à cause des limites de la répartition des information d'émissions, tel qu'un même locuteur se présente que dans une émission, les études des effets de situations communicationnelle en coordination avec le facteur locuteur seront impossible. De plus, la distribution des émissions est aussi in-équilibrée.

Distribution de données sous chaque émission : 29 locuteurs

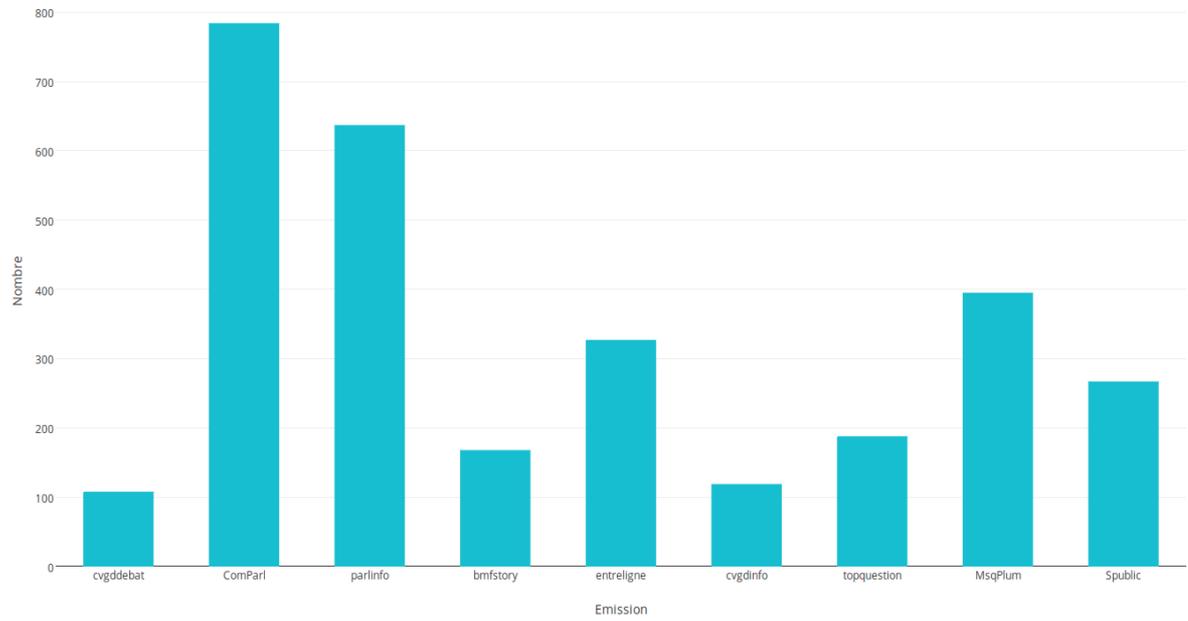


Figure 3 : distribution des extraits des 29 locuteurs étudiés sous chaque émission

2.2 Mesures des paramètres rythmiques

2.2.1 Mesures des paramètres rythmiques de la parole

Dans cette partie, nous allons présenter les différentes mesures appliquées pour obtenir les paramètres rythmiques de la parole.

Nous allons d'abord introduire l'élément de base pour nos calculs : les segments voisés et non voisés.

Un segment voisé se concerne du voisement dans la production de la parole. Un segment est dit voisé si sa production s'accompagne d'une vibration des cordes vocales et non voisé sinon.

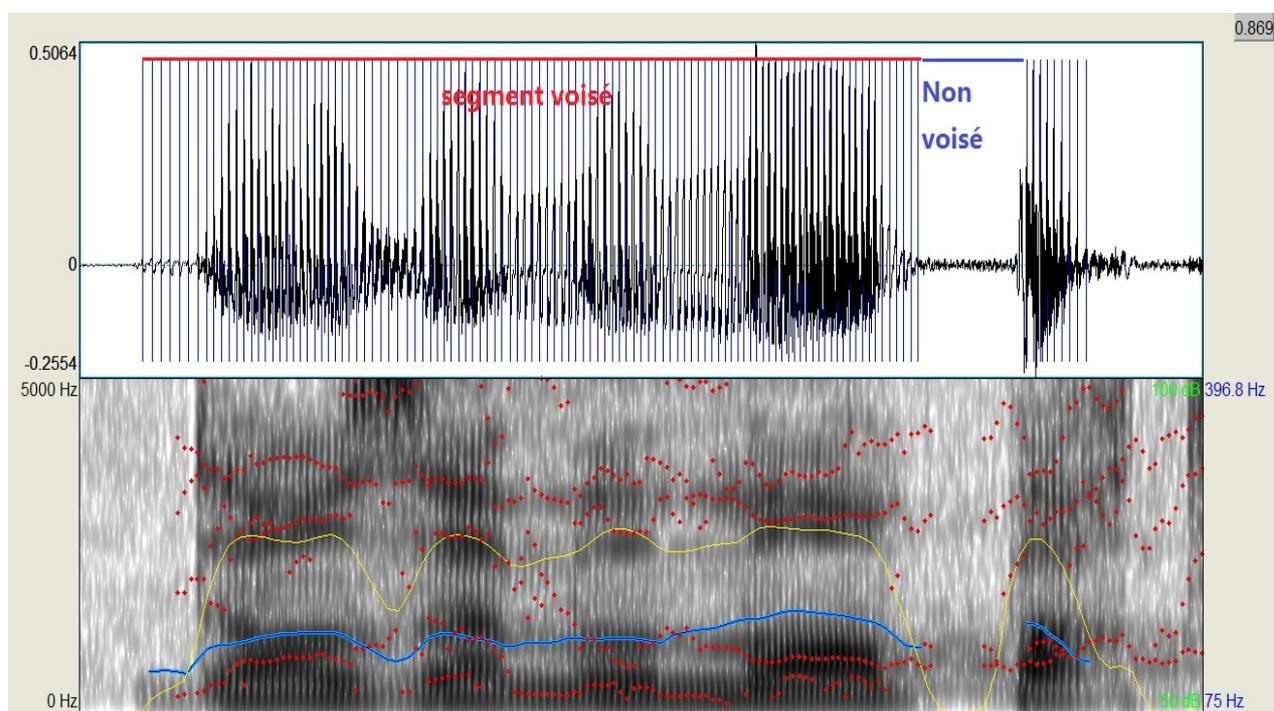


Figure 4 : Un segment voisé et un segment non voisé d'un signal

Toutes les mesures sont faites sur la base des segments voisés et non voisés. Contraire aux méthodes traditionnelles appliquées sur les voyelles et consonnes, les segments voisés et non voisés sont de plusieurs avantages : le travail manuel pour l'annotation est réduit parce que les segments voisés et non voisés sont accessibles directement par le domaine temporel par les informations acoustiques (Dellwo et Fourcin, 2013). Les consonnes voisées sont comprises dans les segments voisés, par exemple le « v », « b » en français.

Dans cette partie, nous allons introduire de façon explicite les mesures effectuées et les calculs pour obtenir des paramètres rythmiques. Sept paramètres seront implémentés,

parmi lesquels quatre sont exploités par les autres chercheurs du domaine, trois sont développés lors de cette étude en compléments.

Nous avons ainsi mesuré : le pourcentage global de segments voisés dans l'extrait sonore, nommée %VO; la durée moyenne des segments voisés, nommée *Moyenne(DuréeVO)*; le coefficient de la variation de la durée des segments voisés, nommée *VarcoVO*; le coefficient de la variation de la durée des segments non voisés, nommée *VarcoUV*; le pourcentage que la durée d'un segment non voisé est supérieure que celle du segment voisé qui précède, nommée $\%(UV_{i+1} > VO_i)$; la durée moyenne de chaque paire de segment voisé et non voisé nommé *Moyenne(paire)*; le coefficient de la variation de la durée des paires de segments, nommée *VarcoPaire*;

Voici les formules.

Le pourcentage global de segments voisés dans l'extrait sonore :

$$\%VO = \sum_1^N \text{DuréeVO} / \text{DuréeSignal} * 100 ;$$

Le %VO représente le pourcentage de l'ensemble des segments voisés dans un signal (Dellwo et Fourcin, 2013). Des auteurs utilisent cette mesure de rechercher des caractéristiques rythmiques de la parole intra et inter langue. D'après notre problématique, nous décidons de ré-appliquer cette méthode pour découvrir les variabilités du rythme intra et inter locuteur. De plus, selon l'étude des deux auteurs, les pauses sont prises en compte en tant que segments non voisés. Donc, un extrait de parole qui contient des pauses importantes aurait un %VO légèrement plus faible que celui qui contient moins de longues pauses, et un coefficient de variance plus grande.

La durée moyenne des segments voisés :

$$\text{Moyenne(DuréeVO)} = \sum_1^N \text{duréeVO} / N ;$$

La *Moyenne(DuréeVO)* représente la durée moyenne des segments voisés d'un signal.

Le *VarcoVO* :

$$\text{VarcoVO} = 100 * \frac{\Delta \text{VO}}{\bar{\text{VO}}}, \text{ (Leemann et al., 2014) ;}$$

Le VarcoVO, le coefficient de la variation de la durée des segments voisés, où le ΔVO est l'écart-type de la durée des segments voisés et le $\bar{\text{VO}}$ la durée moyenne des segments voisés d'un extrait sonore. Le VarcoVO est une valeur normalisée. Il représente le niveau de variation de la durée des segments voisés dont une valeur grande indique une variation large.

Le VarcoUV :

$$\text{VarcoUV} = 100 * \frac{\Delta \text{UV}}{\bar{\text{UV}}}, \text{ (Dellwo et al., 2007) ;}$$

Le VarcoUV, le coefficient de la variation de la durée des segments non voisés. Analysé une formulation identique, VarcoUV est différent de VarcoVO, car il représente la variabilité des segments non voisés où pauses incluses. D'après Duez, les pauses dans la parole des hommes politiques portent du sens. Nous avons trois hommes politiques dans le corpus de la série Target, donc, appliquer les mesures sur les pauses sont utilisé pour cette étude.

Le $\%(UV_{i+1} > VO_i)$:

$$\%(UV_{i+1} > VO_i) = \frac{\sum_1^m j}{i}, \text{ (} j = 0 ; j = j+1 \text{ si } UV_{i+1} > VO_i \text{) ;}$$

Le $\%(UV_{i+1} > VO_i)$ est une valeur en pourcentage, qui indique la prédominance dans le signal, entre 2 segments successifs. Un segment voisé et un segment non voisé se situe en alternance dans le signal. Considérons un segment voisé qui est suivi d'un segment non voisé. Nous comparons la durée du segment voisé VO_i et celle du segment non voisé qui lui succède UV_{i+1} . Le dernier segment voisé ne serapas pris en compte par cette mesure. Pour être explicite, nous avons déterminé la proportion de comparaison pour lesquelles $\%(UV_{i+1} > VO_i) > 50\%$, c'est-à-dire, les segments non voisés sont à la place prédominante. Le signal mesuré contiendrait par exemple des pauses longues, un débit de parole faible. Sinon, les segments voisés sont à la place prédominante.

La Moyenne(paire) :

$$\text{Moyenne(paire)} = \overline{VO_k + UV_{k+1}}, (k \geq 0);$$

La durée moyenne de paire, c'est-à-dire la durée moyenne d'un segment voisé plus celle du segment non voisé qui le suit. Cette mesure issue des études de Fant et al. (1989). Ils ont amené une recherche sur le style de lecture. La « paire » est de l'origine de la notion « inter-stress intervals ». Le « inter-stress intervals » est définie comme : une intervalle entre le début d'un segment voisé et le début du prochain. Cette mesure va donner les traits de regroupements des segments sur le domaine temporel, qui sont estimés spécifiques aux locuteurs.

Le VarcoPair :

$$\text{VarcoPaire} = \frac{\Delta \text{DuréePair}}{\text{Moyenne}(\text{paire})} * 100 ;$$

Le VarcoPair, le coefficient de la variation de la paire de segments.

2.2.3 Une méthode semi-automatique pour les mesures des paramètres rythmiques

Au 2.2.2, nous avons déjà présenté les mesures à appliquer pour obtenir des paramètres rythmiques de la parole. Alors, pour avoir les fichiers d'entrées nécessaires afin de réaliser de ces mesures, nous allons employer un outil semi-automatique.

L'étude de la parole implique souvent une grande quantité de données, nous avons choisi un outil qui est automatisé au maximum. Donc, c'est le script Praat : ProsodyPro¹⁰, développé par Xu Yi (2013). Il permet de mesurer automatiquement tous les fichiers sons sous le même répertoire.

¹⁰ <http://www.homepages.ucl.ac.uk/~uclyyix/ProsodyPro/>

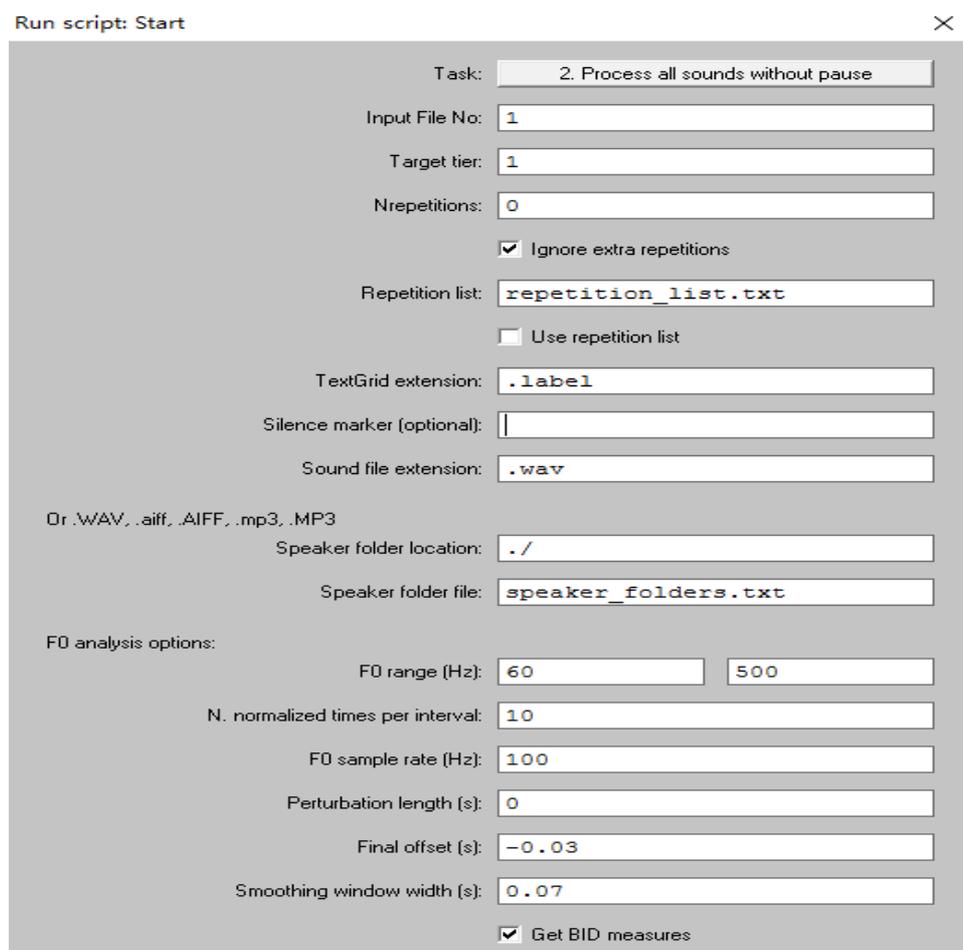


Figure 5 : Capture d'écran console ProsodyPro

ProsodyPro est principalement développé pour faciliter l'analyse prosodique. Ce script parcourt automatiquement le dossier du corpus qui contient les sons sous format .wav, mesure les valeurs de F0 de chaque enregistrement et indique en même temps l'instant qui correspond aux valeurs mesurées avec la fonctionnalité « Pulses » sous Praat. Donc, les fichiers de sortie NomWav.f0 sont directement exploitables. Plusieurs analyses sont appliquées sur un signal sonore, une dizaine fichiers de sorties sont générés pour un seul signal tel que la vélocité de F0, les segmentations et annotations des labels semi-automatiques, etc. Nous allons simplement utiliser les NomWav.f0 pour le calcul des paramètres rythmiques pour cette étude. Ces fichiers se présentent comme sur la figure 5, le temps est indiqué en seconde et le F0 en Hz.

2.2.4 Procédure de calcul

Dans cette partie, la procédure de calcul sera présentée explicitement. Nos entrées

originales sont les extraits d'enregistrement du corpus FABIOLÉ, en raison d'un défaut incontournable de qualité des signaux, nous avons choisi 29 locuteurs parmi les 30 dans la série Target. Hervé Pauchon est exclu pour cette cause.

Les signaux sont d'abord mesurés sous Praat avec le script ProsodyPro. Les fichiers NomWav.f0 sont générés. Un fichier NomWav.f0 avec deux colonnes « les traces sur le domaine de temps » et « la valeur F0 qui correspond à chaque point temporel ». Chaque extrait sonore aura un fichier NomWav.f0. Ces NomWav.f0 sont nécessaire pour le calcul des mesures rythmiques.

Voici un exemple de fichier NomWav.f0 de ProsodyPro :

<u>Time</u>	<u>F0</u>
0.3822624377468177	141.87542526791563
0.3891457147222782	145.78047823567823
0.39617552117625876	145.70534834427576
0.4031310651894413	145.66885392677634
0.41000796878044815	145.63277212062604
0.4168679944146769	145.40712600469902
0.42374522505210566	145.18091396637556
0.43067223220287304	143.5532415093289
0.43798017115764576	143.5432415093289
0.55433693637441	170.69842287488586
0.560147243905525	170.68842287488587
0.5660749869620754	166.75399133999994
0.57213996981844	163.0494750778298
0.5783291201581287	160.1231558514338
0.5846354275064958	157.04943630215536
0.5910190333414478	156.2551296155502
0.59745263429684	154.62135364180494
0.6039774566424172	151.92413408746552

Figure 6 : exemple fichier NomWav.f0

Après la préparation de données, nous allons les passer sous MATLAB. Toutes les formules sont implémentées dans un script MATLAB (cf : annexe1). Les fichiers NomLocuteur.csv vont être générés à partir des NomWav.f0. Le NomLocuteur.csv contient sur chaque ligne : le nom du fichier .wav et les sept valeurs obtenues à partir du signal. Comme chaque locuteur de la série Target a au minimum 100 fichiers de son, le NomLocuteur.csv contient environ 100 lignes chacun. Nous avons au total 29 fichiers NomLocuteur.csv pour les étapes suivantes.

Tous les fichiers de sortie sont pré-traités via un script Perl pour faire une annotation des informations de l'émission et du locuteur. Cette manipulation a été faite pour faciliter l'étape de l'analyse statistique.

Après l'annotation, les données sont prêtes pour l'analyse et classification.

2.3 Méthodologies d'analyse d'influence des facteurs

2.3.1 Les facteurs étudiés

D'après nos études, nous nous intéressons aux effets des facteurs suivants :

Premièrement, un effet de locuteur. Nous supposons que le rythme soit graduellement spécifique aux locuteurs : tel que le rythme de la parole d'un certain locuteur pourrait être proche que celui de l'autre et en même temps loin que celui des autres.

Deuxièmement, un effet de situation communicationnelle. Dans la partie d'état de l'art, nous avons fait référence aux théories linguistiques et sociolinguistes et la plupart des études montrent que les styles de parole varie en fonction des situations de communication.

Les deux premiers occupent principalement la partie d'analyse, néanmoins, le troisième facteur est aussi à grand intérêt pour nos études : le rôle de locuteur. Donc, le rôle du locuteur sera le troisième facteur. Il est mis à la troisième place à cause des limites de nos données : la dispersion de cette information n'est pas équilibrée. Donc nous allons faire une analyse sur ce facteur qu'avec une partie de nos données au lieu de l'ensemble.

Limité par l'organisation du corpus, il n'est pas possible de voir l'interaction de ces trois facteurs. Parce que les données sont pas suffisantes pour effectuer une ANOVA à deux facteurs : soit un locuteur est présenté dans une seule émission, soit le nombre de locuteurs sous chaque émission n'est pas équilibré. Donc, nous ne pouvons pas croiser ces deux facteurs.

2.3.2 ANOVA à un facteur

Dans les chapitres précédent, nous avons présenté les méthodes impliquées par nos études et la réalisation du processus de calcul. Pour démontrer nos hypothèses, que le rythme est un composant des paramètres de style de parole qui est spécifique aux locuteurs et aussi sensible aux situations communicationnelles, nous allons effectuer une série d'analyses statistiques sur les résultats obtenus.

La méthode d'analyse employée est principalement l'ANOVA à un facteur.

Toutes les analyses statistiques sont effectuées sous MATLAB avec l'outillage « Statistics and Machine Learning ».

La taille de l'effet de chaque mesure sera calculé pour montrer la puissance de chaque paramètre.

2.4 Classification en locuteur à partir des données rythmiques

Pour voir l'impact du facteur locuteur sur les paramètres rythmiques, nous avons effectué

une classification en locuteur.

Weka¹¹ est employé lors de l'étape de classification. Il est une suite de logiciels d'apprentissage automatique écrit en Java, outil informatique développé par des chercheurs de l'Université de Waikato en Nouvelle-Zélande. Plusieurs algorithmes d'apprentissage automatique et de fouille de données sont implémentés sous Weka, la classification et clustering sont réalisables.

Nous allons passer une série de tests avec différents algorithmes sur nos données pour trouver une méthode de classification adéquate. Un test post-hoc sera aussi fait pour vérifier l'exactitude de la classification. Il va être discuté dans le chapitre résultats.

Nous allons aussi générer les macro-classes rythmiques à partir des résultats de classification dont nous discuterons dans le sous-chapitre suivant.

La classification est l'un des buts principaux d'apprentissage automatique. Plusieurs modèles sont implémentés, tel que les règles, les arbres. Nous avons passé une série de test de classification sous Weka pour trouver une méthode appropriée. Parmi tous les algorithmes, nous avons trouvé deux modèles robustes: le J48 et le LMT. Le J48 est une version de C4.5, le plus connu du domaine apprentissage automatique, implémenté sous Weka. Le J48 est beaucoup plus rapide que le LMT lors de l'apprentissage et de la construction de modèle. Par contre, le LMT donne une valeur de taux de bonne classification plus élevée que le J48. Ci-dessous le tableau de comparaison d'un test avec la même série de données :

	J48	LMT
Vitesse (construction/test sur l'apprentissage)	0.34/0.13 secondes	87.89/0.09 secondes
Taille de l'arbre	54 feuilles/107 hauteur	3 feuilles/5 hauteur
Taux de bonne classification : apprentissage	30,84%	33,06%
Taux de bonne classification : test	21,71%	29,67%

Tableau 3 : comparaison d'algorithme pour la classification

Nous avons pu constater que le LMT est généralement plus robuste que le J48. Il fait moins de sur-apprentissage, car il donne un écart de taux de bonne classification égale à 3.39%, plus petit que celui de J48, qui est égale à 9.13%. De plus, c'est aussi important au

¹¹ <http://www.cs.waikato.ac.nz/ml/index.html>

stade de la classification, LMT est plus robuste que le J48 parce qu'il donne un taux de bonne classification plus élevé. Ce test a été fait sous les mêmes conditions et avec la même série de données : tous les paramètres sont initialisés avec les mêmes valeurs, le même seuil de classification en fonction de locuteur, pour que les résultats soient comparables. Le pourcentage de découpage est de 66%, c'est-à-dire, 66% de nos données sont employées par l'apprentissage et 34% par la classification. 1018 instances sont utilisées pour faire ce test de classification.

D'après nos premiers tests, il n'est pas difficile de noter que le LMT convient plus à notre tâche que le J48.

Le LMT, abréviation de *Logistic Model Trees* en anglais, est un modèle de classification développé par Landwehr et al. (2003). LMT respecte la structure d'un arbre et effectue une régression logistique aux feuilles de cet arbre. Comme la structure de l'arbre est facile à interpréter et rapide à construire, elle est largement utilisée dans le domaine.

Le LMT sépare l'espace d'instances en plusieurs parties grâce à un arbre classique. Dans chaque partie, les fonctions de régression logistiques sont appliquées pour construire les sous-modèles dans cet espace. Suivant cette étape, l'estimation du maximum de vraisemblance est faite (Landwehr et al., 2003). Un arbre traditionnel pourrait être considéré comme une suite de règles qui indiquent « comment classifier les instances ». Son chemin correspond à une conjonction d'une série d'expressions booléennes. Alors, comme le LMT a calculé la probabilité pour chaque instance, il est plus flexible et robuste. Pour conclure, LMT apparaît plus adéquat pour une tâche de classification dans le cadre de reconnaissance automatique du locuteur par rapport aux autres méthodes. Donc, la procédure de classification et la définition de macro-classe seront réalisées sous LMT. Par contre, un test post-hoc de la classification sera fait avec le réseau neurone, nommé *MultilayerPerception*, implémenté sous Weka dans la gamme *Fonction* comme le *classifier*.

Nous effectuons la classification avec les sept valeurs obtenues à partir des mesures des segments voisés et non voisés. Comme nous avons présenté, chaque enregistrement correspond à sept attributs numériques, soient les sept valeurs, et trois attributs nominaux, soient le nom de l'extrait sonore, nommé Fichier, l'information d'émission de chaque extrait, nommé Émission, le nom du locuteur, nommé Nom.

Voici un extrait de la structure de données pour la classification :

```

@relation z9_loc_toutes_mesures
@attribute Fichier {Arnaud_Ardoin---cavousregardeledebat_2013-10-24---3362---00,Arnaud_Ardoin
@attribute Emission {cvgddebat,ComParl,parlinfo,bfmstory,entreligne,cvgdinfo,topquestions,Ms
@attribute Nom {Arnaud_Ardoin,Benjamin_Sportouch,Bruno_Jeudy,Carl_Meeus,Christophe_Conte,Dan
@attribute VO numeric
@attribute moydVO numeric
@attribute VarcoVO numeric
@attribute VarcoUV numeric
@attribute dUVdVO numeric
@attribute moydVodUV numeric
@attribute VarcoPair numeric

@data
Arnaud_Ardoin---cavousregardeledebat_2013-10-24---3362---00,cvgddebat,Arnaud_Ardoin,65.73809
Arnaud_Ardoin---cavousregardeledebat_2013-10-24---3364---00,cvgddebat,Arnaud_Ardoin,59.86476

```

Figure 7 : extrait de la structure de données pour la classification

Nous avons au total 2993 instances, alors la valeur minimum de taux de bonne classification est de 3.35%, c'est-à-dire une classification complètement dûe au hasard. Comme décrit en dessus, 1975 instances sont utilisées à l'étape d'apprentissage et 1018 instances sont passées au test. Nous avons laissé 2/3 des données pour l'apprentissage car le LMT exige une quantité de données assez grande pour obtenir des paramètres adéquates pour ses fonctions de régression logistique au niveau de sous-modèles afin que l'exactitude de classification soit maximisée.

2.5 Influence du rythme sur l'identification du locuteur

2.5.1 Définition de macro-classes rythmiques basée sur LMT

Pour voir l'influence du rythme sur l'identification du locuteur, nous allons définir les macro-classes rythmiques à partir des résultats de la classification sous LMT, qui génère par nos données un arbre à 3 classes (cf : figure 8).

Nous allons faire notre premier essai avec ce chiffre là.

LM_1	VarcoUV <= 81,77739 & %VO <= 59,14389
LM_2	VarcoUV <= 81,77739 & %VO > 59,14389
LM_3	VarcoUV > 81,77739

Tableau 4 : seuils de LMT

Représentation graphique de la structure de l'arbre :

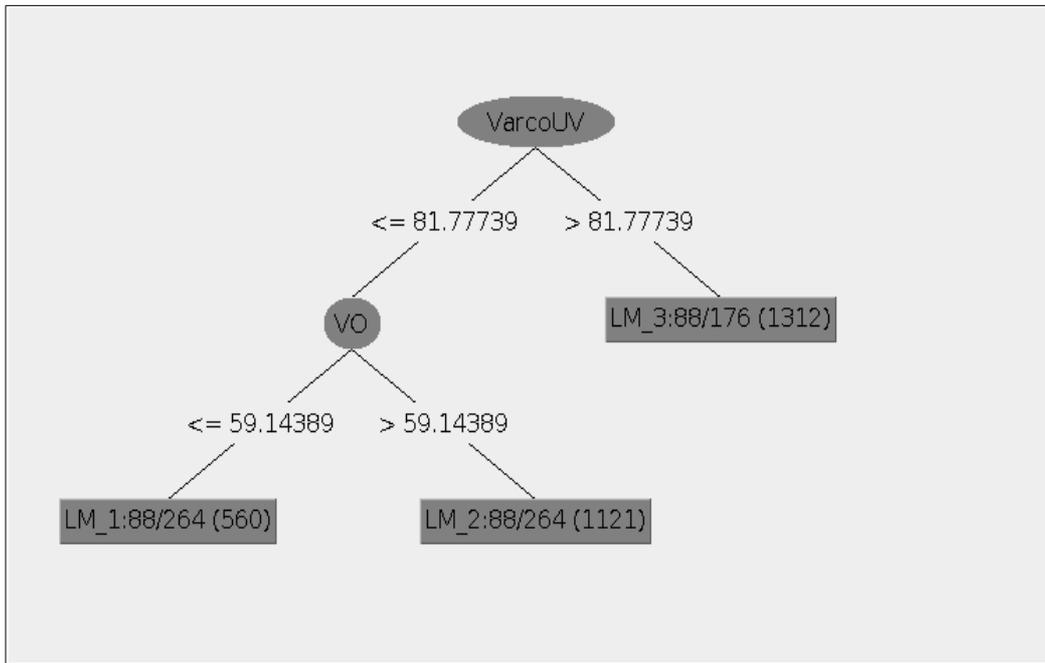


Figure 8 : Classification en fonction de locuteurs par LMT structure de l'arbre

2.5.2 Association de la macro-classe avec un extrait sonore

D'après la problématique, nous supposons que les traits rythmiques de la parole ont l'impact sur le système RAL. Grâce aux études faites par les autres chercheurs sur le projet FABIOLÉ, nous avons une série de données générée par le système LIA-ALIZE/SpkDet à partir des tests *Target* et *Non-target*. Pour trouver l'influence du rythme sur l'identification du locuteur, nous allons par la suite associer chaque paire d'enregistrement avec sa macro-classe correspondante.

Ce qui nous intéresse beaucoup c'est l'interaction entre les classes. Les tests *Target* ont pour but de comparer 2 signaux de parole produit par le même locuteur. La paire de signaux mis en test étant issus d'un même locuteur, nous attendons un score le plus élevé possible. (cf : 1.3).

Nous supposons qu'une paire de signaux qui sont dans la même classe rythmique aura un score plus élevé que ceux qui ne sont pas dans la même classe. Ceci est dit « Interaction entre les classes ».

Pour montrer l'impact du rythme de la parole sur le système RAL, nous allons donc chercher le lien entre les scores du système et les macro-classes en calculant les moyennes

de scores inter et intra classes rythmiques.

Chapitre 3 : Résultats

3.1 Influences des facteurs sur le rythme

Nous présentons dans ici les détails des analyses concernant des facteurs étudiés en montrant les résultats. Chaque paramètre rythmique est étudié séparément.

3.1.1 Le facteur locuteur

Nous discutons dans cette partie de l'effet du locuteur.

Le tableau ci-dessous regroupe toutes les analyses faites avec le facteur locuteur sur l'ensemble des paramètres rythmiques, chacun étant étudié séparément.

%VO	F(28,2964) = 76.83, p <.001
Moyenne(DuréeVO)	F(28,2964) = 64.13, p <.001
VarcoVO	F(28,2964) = 36.22, p <.001
VarcoUV	F(28,2964) = 88.32, p <.001
%(UV _{i+1} > VO _i)	F(28,2964) = 60.02, p <.001
Moyenne(paire)	F(28,2964) = 73.99, p <.001
VarcoPaire	F(28,2964) = 25.47, p <.001

Tableau 5 : tableau d'ANOVA, locuteurs comme facteur

La figure 9 montre ainsi les variations du %VO en fonction des locuteurs. Certains locuteurs, tels les locuteurs 6, 12 ou 14 ont des valeurs de %VO proches de 55% en moyenne tandis que d'autres utilisent beaucoup plus souvent le voisement, avec environ 65% de %VO, par exemple locuteur 9, 13, 16.

Nous constatons un niveau de très haut significativité, la valeur p est tout le temps très faible quand on effectue l'analyse. Donc, nous pouvons affirmer que tous les paramètres rythmiques sont significatifs, mais si le locuteur est significatif, cela ne nous indique pas la taille de l'effet.

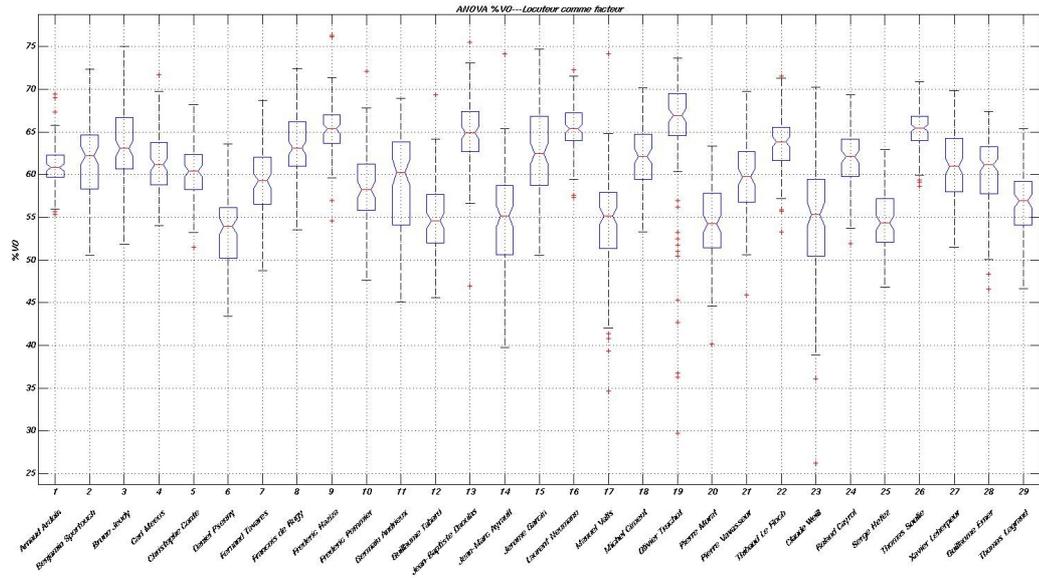


Figure 9 : boîte à moustache---%VO locuteur comme facteur

Les représentations pour les six autres paramètres en fonction du locuteur sont en l'annexe (cf : annexe 2).

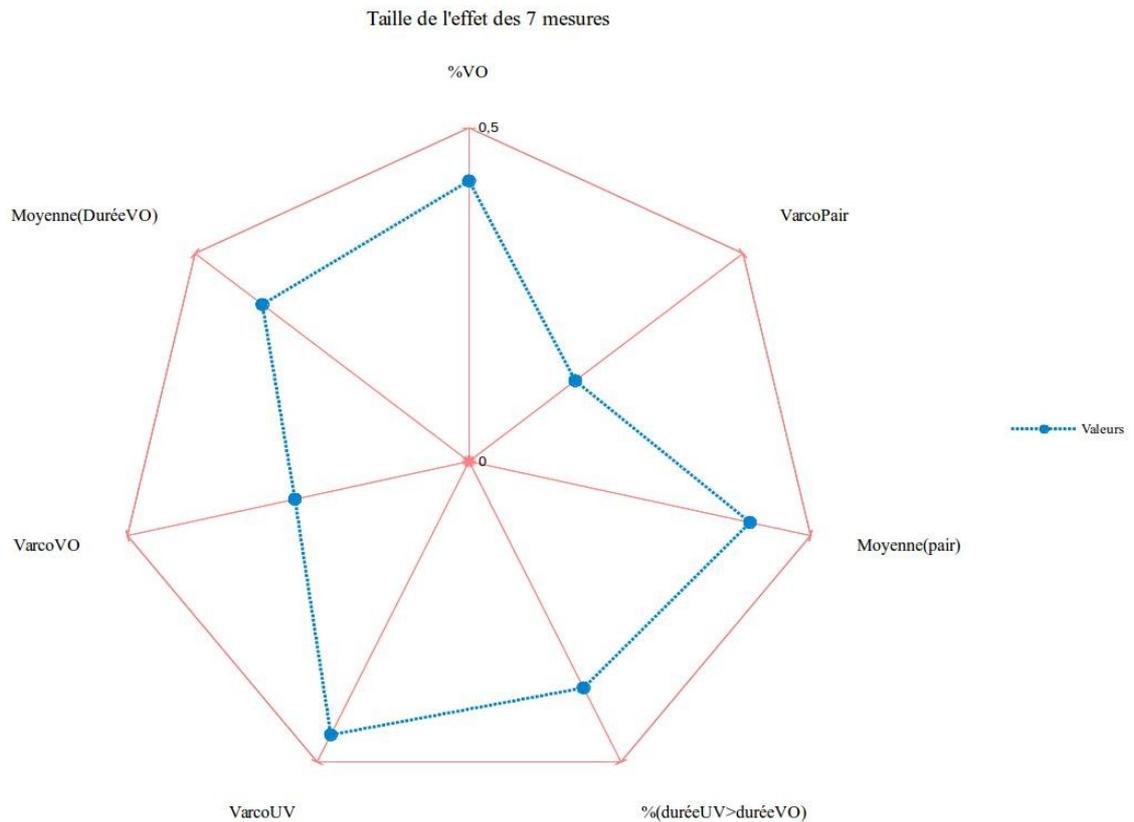


Figure 10 : Radar-chart d'êta-carré des paramètres---locuteur comme facteur

Pour évaluer la taille de l'effet pour les sept paramètres en fonction du locuteur, nous utilisons l'êta-carré représenté par le caractère grec : η^2 . L'êta-carré décrit la force de l'effet d'un paramètre et exprime le pourcentage de la variance totale expliquée par le facteur locuteur.

Ci-dessous la formule :

$$\eta^2 = \frac{SS_{between}}{SS_{total}} ;$$

D'après le tableau ci-dessous, le VarcoUV et le %VO ont des effets plus importants par rapport aux autres. Ceci indique que la variation des durées des segments non voisés en coordination avec le pourcentage du voisement d'un signal sont plus pertinents pour démontrer l'impact du facteur locuteur sur les paramètres rythmiques.

Mesures	η^2
%VO	0.42056
Moyenne(DuréeVO)	0.37727
VarcoVO	0.25492

VarcoUV	0.45485
$\%(UV_{i+1} > VO_i)$	0.37686
Moyenne(paire)	0.41142
VarcoPaire	0.19391

Tableau 6 : liste d'êta-carré des 7 paramètres en fonction de locuteurs

La représentation en radar-chart permet de bien visualiser les effets respectifs des paramètres (figure 10).

3.1.2 Influence de la situation communicationnelle : le facteur émission

Dans cette partie, nous avons étudié les résultats d'analyse sur le facteur émission, ce qui a un effet important sur les styles de parole d'après les sociolinguistes et les chercheurs.

Les émissions sont interprétées ici comme relevant de différentes situations communicationnelles.

Selon les informations du corpus, nous avons neuf émissions. Comme la dispersion des extraits sonores n'est pas équilibrée, nous avons d'abord tiré, parmi l'ensemble de nos données 900 lignes de données, soit 100 sous chaque émission, pour que l'analyse soit réalisable.

La répartition des %VO en fonction des 9 émissions est donné en figure 11 et celle de VarcoUV en figure 12. Les 5 restes sont en annexe. (annexe 3).

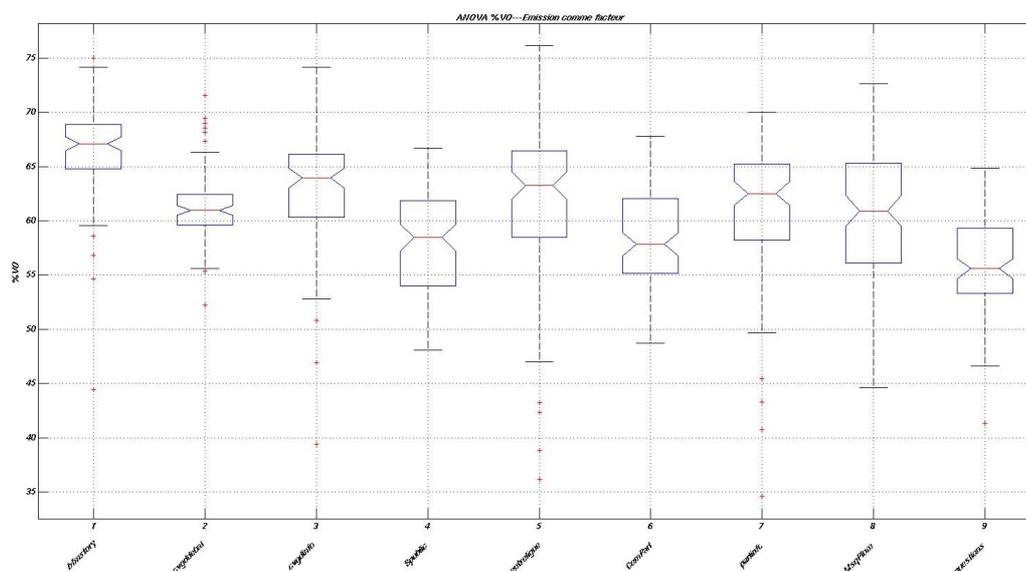


Figure 11 : boîte à moustache---%VO Émission comme facteur

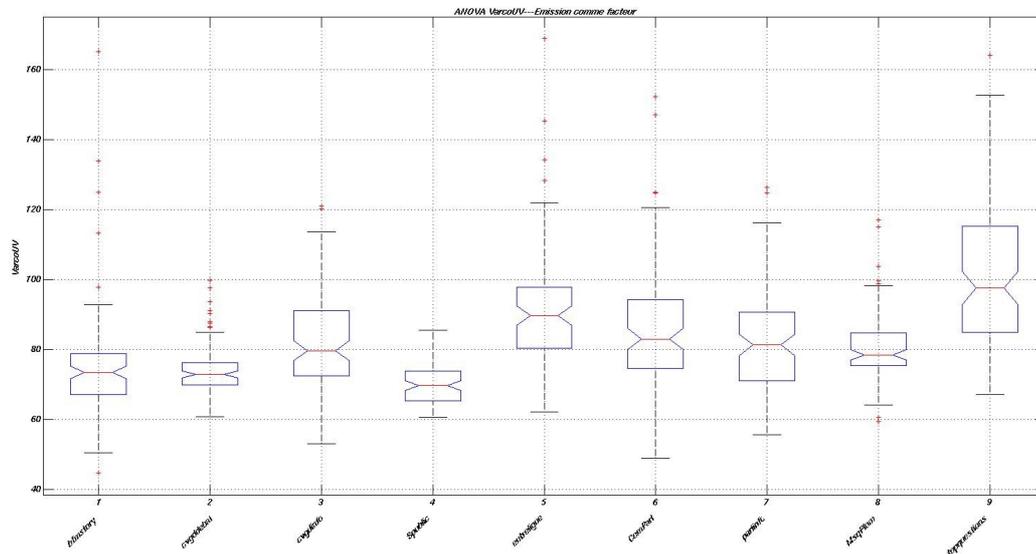


Figure 12 : boîte à moustache---VarcoUV émission comme facteur

Les résultats statistiques sont donnés dans le tableau suivant. (Tableau 7).

La taille des effets des émissions sur ces paramètres sont relativement faibles comparé au facteur locuteur. La valeur maximale d'êta-carré est celle du paramètre VarcoUV, égale à 0.2781. Cela indique que le facteur émission est moins important que le facteur locuteur sur les données FABIOLÉ.

Mesures	ANOVA	η^2
%VO	F(8,891) = 36.56, p<.001	0.2471
Moyenne(DuréeVO)	F(8,891) = 16.14, p<.001	0.1266
VarcoVO	F(8,891) = 30.77, p<.001	0.2156
VarcoUV	F(8,891) = 42.89, p<.001	0.2781
%($UV_{i+1} > VO_i$)	F(8,891) = 27.99, p<.001	0.2008
Moyenne(paire)	F(8,891) = 29.36, p<.001	0.2086
VarcoPaire	F(8,891) = 4.07, p<.001	0.0352

Tableau 7 : liste d'êta-carré des 7 paramètres en fonction d'émission

3.1.3 L'effet du rôle du locuteur sur le rythme de la parole

Le rôle social d'un locuteur est un facteur intéressant à analyser, surtout ses effets sur les

traits rythmiques. Par contre, l'analyse a été faite sur un sous-ensemble de données : deux hommes politiques, d'après l'étude de Duez (cf : 1.2.2). Nous sommes limités par la dispersion des données sur le rôle du locuteur parce qu'elle n'est pas équilibrée (cf : 2.1.2, figure 2).

D'après l'observation du corpus, nous trouvons que les données de Jean-Marc Ayrault et Manuel Valls sont appropriées. Ils sont tous les deux hommes politiques. De plus, ils ont parlé dans la même émission : « topquestions ».

La première valeur testée est le %VO (cf : figure 13), pour laquelle aucune différence n'est trouvée entre les deux locuteurs.

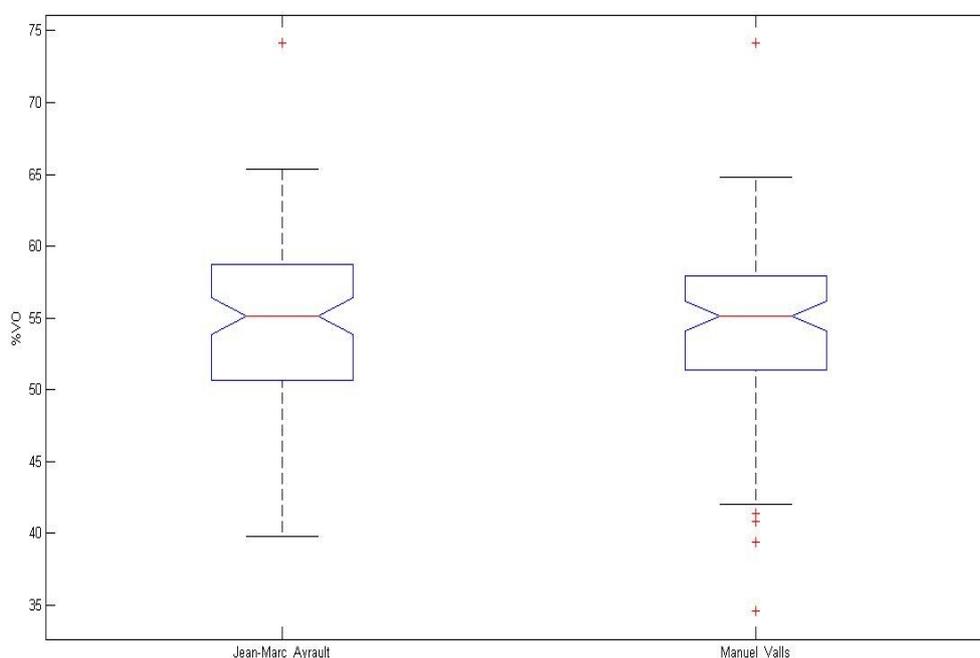


Figure 13 : boîte à moustache---%VO Jean-Marc Ayrault vs. Manuel Valls

$F(1,179) = 0.47$, $p = 0.4923$, ns. Jean-Marc Ayrault et Manuel Valls, sont intervenus dans la même situation communicationnelle avec le même rôle social et ne peuvent pas être distingués l'un de l'autre par le %VO dans ses signaux.

De plus, l'analyse est faite pour chaque paramètre, dont le %VO et le VarcoVO ne montrent pas de différence significative entre les deux hommes politiques. Cela illustre que la proportion de voisement d'un signal pourrait être utile pour regrouper ou distinguer la classe rythmique et que les facteurs intéressants pourraient être la situation

communicationnelle ou le rôle du locuteur.

Dans nos données, nous avons 3 hommes politiques, donc, pour vérifier les effets unifiés de l'émission et de rôle, nous avons effectué une comparaison multiple sur ces 3 sujets.

Il est intéressant de noter que François de Rugy est distingué des deux autres hommes politiques. Jean-Marc Ayrault et Manuel Valls sont présentés plusieurs fois dans la même émission « topquestions » avec le même rôle, cependant François de Rugy était présenté plusieurs fois dans les autres émissions : bmfstory, cvginfo, parlinfo, topquestions. En plus de l'impact du rôle du locuteur, un effet de situation communicationnelle est observé ici. Par contre, nous n'avons pas pu faire une analyse sur cela faute de manque données.

3.2 Validation croisée en locuteur et le test post-hoc

3.2.1 La classification en locuteur par LMT

Nous avons obtenu un taux de bonne classification égale à 32.1218% et celle de la validation croisée montré par la figure 13 ci-dessous, égale à 32.0748%. Tous les résultats présentés dessous sont basés sur la validation croisée car elle est effectuée sur l'ensemble des données.

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      960      32.0748 %
Incorrectly Classified Instances    2033     67.9252 %
Kappa statistic                    0.2957
Mean absolute error                 0.0553
Root mean squared error             0.1675
Relative absolute error             83.1505 %
Root relative squared error         91.8084 %
Total Number of Instances          2993
```

Figure 14 : statistique de validation croisée par LMT

LMT, comme décrit dans le chapitre précédent (cf : 2.5.1), divise dans un premier temps les données à classer en trois dimensions. Le paramètre primaire pris systématiquement par l'algorithme est « VarcoUV », qui a une taille de l'effet égal à 0.45485, la mesure la plus puissante parmi toutes les mesures appliquées. Et, celle qui le suit, le %VO, taille de l'effet égale à 0.42056. Ces deux mesures sont la première et la seconde puissantes méthodes parmi tous les paramètres mesurés alors que le LMT commence à montrer sa

robustesse lors du choix de ces paramètres.

En effets, certains locuteurs sont correctement classés à partir de ces seuls paramètres rythmiques, tels que Frédéric Pommier ou Carl Meeus. Par exemple, les locuteurs qui donnent une valeur de précision à l'extrémité, comme François de Rugy, Arnaud Ardoin, ou Jean-Marc Ayrault. Le premier est complètement confondu avec les autres locuteurs. En revanche, les deux derniers ont eu des taux de vrai positif relativement hauts. Le taux de vrai positif varie assez largement entre locuteurs.

Voici le tableau de statistiques détaillé de la validation croisée en locuteur sous LMT :

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.5	0.035	0.333	0.5	0.4	0.892	Arnaud_Ardoin
	0.38	0.029	0.309	0.38	0.341	0.893	Benjamin_Sportouch
	0.36	0.025	0.33	0.36	0.344	0.869	Bruno_Jeudy
	0.32	0.021	0.348	0.32	0.333	0.865	Carl_Meeus
	0.23	0.018	0.303	0.23	0.261	0.842	Christophe_Conte
	0.34	0.022	0.347	0.34	0.343	0.92	Daniel_Psenny
	0.13	0.021	0.173	0.13	0.149	0.822	Fernand_Tavares
	0.01	0.003	0.1	0.01	0.018	0.725	Francois_de_Rugy
	0.24	0.031	0.212	0.24	0.225	0.853	Frederic_Haziza
	0.58	0.023	0.468	0.58	0.518	0.953	Frederic_Pommier
	0.09	0.017	0.155	0.09	0.114	0.751	Germain_Andrieux
	0.42	0.034	0.3	0.42	0.35	0.894	Guillaume_Tabard
	0.25	0.025	0.258	0.25	0.254	0.853	Jean-Baptiste_Daoulas
	0.64	0.015	0.604	0.64	0.621	0.929	Jean-Marc_Ayrault
	0.301	0.024	0.33	0.301	0.315	0.897	Jerome_Garcin
	0.14	0.026	0.159	0.14	0.149	0.826	Laurent_Neumann
	0.39	0.021	0.39	0.39	0.39	0.938	Manuel_Valls
	0.19	0.023	0.224	0.19	0.205	0.843	Michel_Ciment
	0.36	0.02	0.387	0.36	0.373	0.887	Olivier_Truchot
	0.1	0.021	0.143	0.1	0.118	0.826	Pierre_Murat
	0.19	0.02	0.25	0.19	0.216	0.858	Pierre_Vavasseur
	0.4	0.031	0.31	0.4	0.349	0.904	Thibaud_Le_Floch
	0.31	0.021	0.333	0.31	0.321	0.897	Claude>Weill
	0.42	0.022	0.4	0.42	0.41	0.901	Roland_Cayrol
	0.31	0.025	0.298	0.31	0.304	0.921	Serge_Hefez
	0.292	0.029	0.284	0.292	0.288	0.905	Thomas_Soulie
	0.27	0.024	0.284	0.27	0.277	0.844	Xavier_Leherpeur
	0.569	0.056	0.377	0.569	0.453	0.9	Guillaume_Erner
	0.41	0.025	0.363	0.41	0.385	0.926	Thomas_Legrand
Weighted Avg.	0.321	0.025	0.304	0.321	0.308	0.874	

Figure 15 : détails de validation croisée en locuteur sous LMT



Figure 16: Taux de vrai positif--Classification en locuteur par LMT (validation croisée)

L'ensemble de résultats de la classification sous LMT est en l'annexe. (annexe 3).

Nous voyons ici en bas, la matrice de confusion sortie à partir d'une validation croisée de classification par LMT. Correspondante au graphe de la figure 15 au-dessus, on observe ainsi les proximités rythmiques entre locuteurs.

Pour être précis, l'observation fine à quelques locuteurs est nécessaire.

```

=== Confusion Matrix ===
 a b c d e f g h i j k l m n o p q r s t u v w x y z aa ab ac <-- classified as
50 5 2 1 1 0 5 0 1 0 0 2 3 0 0 5 0 3 1 0 1 2 0 0 4 0 4 10 0 | a = Arnaud_Ardoin
10 38 9 1 0 1 2 0 0 0 0 2 0 0 0 7 0 0 1 2 0 6 1 0 4 4 6 6 0 | b = Benjamin_Sportouch
9 4 36 5 3 4 1 0 0 0 0 1 0 0 3 5 0 1 0 1 3 0 1 3 0 5 3 10 2 | c = Bruno_Jeudy
4 0 3 32 8 0 1 1 4 9 0 3 0 0 2 0 0 4 0 1 2 6 3 5 1 1 2 2 6 | d = Carl_Meeus
7 2 7 13 23 0 3 0 4 5 0 6 3 0 2 1 0 2 0 4 0 4 3 5 0 1 0 3 2 | e = Christophe_Conte
0 2 1 0 0 34 1 0 0 0 4 3 0 0 0 0 2 2 0 3 1 0 5 0 26 0 0 13 3 | f = Daniel_Psenny
8 4 4 3 2 1 13 0 1 2 0 15 2 0 0 2 1 3 1 1 3 4 8 0 5 2 12 3 0 | g = Fernand_Tavares
7 7 3 2 1 1 1 1 6 1 1 3 1 3 4 7 5 3 3 0 1 4 0 2 2 6 1 21 3 | h = Francois_de_Rugy
3 0 0 3 5 0 1 0 24 1 1 0 16 1 9 2 1 5 4 0 3 3 0 5 0 7 1 5 0 | i = Frederic_Haziza
0 0 0 2 4 0 1 0 1 58 2 3 2 2 0 0 0 0 0 0 1 6 6 4 0 8 0 0 0 | j = Frederic_Pommier
4 14 1 1 1 9 2 0 1 1 9 2 1 0 0 5 6 5 6 6 0 4 2 2 3 2 7 6 0 | k = Germain_Andrieux
1 0 0 3 1 3 5 0 1 1 1 42 0 0 3 0 3 1 0 10 5 1 6 1 2 0 1 3 6 | l = Guillaume_Tabard
0 1 1 2 0 1 1 1 21 0 0 0 25 2 8 3 3 1 3 1 2 6 0 2 0 9 3 4 0 | m = Jean-Baptiste_Daoulas
0 0 0 0 0 0 0 0 0 3 3 0 4 64 1 1 16 0 0 1 0 1 1 4 0 0 0 0 1 | n = Jean-Marc_Ayrault
1 0 1 1 3 0 0 2 9 1 0 0 8 1 34 3 5 0 1 4 7 0 0 10 1 1 0 3 17 | o = Jerome_Garcin
2 7 12 1 2 0 0 1 6 0 3 1 5 0 0 14 0 8 14 0 1 4 0 3 0 7 2 7 0 | p = Laurent_Neumann
0 0 0 0 1 0 0 0 1 1 5 4 3 24 1 0 39 0 1 3 0 0 1 4 1 0 0 8 3 | q = Manuel_Valls
5 1 3 2 1 2 9 0 4 0 6 0 2 1 0 8 0 19 4 2 1 9 1 1 2 0 8 9 0 | r = Michel_Ciment
0 1 2 2 0 1 1 0 7 1 0 1 6 0 1 8 0 1 36 0 0 10 2 0 0 14 4 2 0 | s = Olivier_Truchot
4 4 0 1 3 2 2 1 0 3 2 19 0 1 3 0 8 1 0 10 3 0 4 3 10 0 2 2 12 | t = Pierre_Murat
8 0 2 3 6 2 5 0 4 3 2 4 2 0 9 0 1 0 0 2 19 2 7 1 2 0 1 2 13 | u = Pierre_Vavasseur
4 4 0 0 3 0 0 0 1 5 4 1 3 1 0 3 1 8 6 0 0 40 1 1 0 8 4 2 0 | v = Thibaud_Le_Floch
1 0 0 5 3 6 10 0 0 10 1 8 0 0 0 0 1 1 0 4 6 5 31 0 2 1 4 1 0 | w = Claude_Weill
1 0 0 4 3 0 0 0 8 9 0 1 0 2 13 0 3 3 0 0 2 1 0 42 0 1 0 4 3 | x = Roland_Cayrol
3 0 4 0 2 21 0 0 0 0 3 4 0 0 0 0 0 2 0 7 0 1 2 0 31 0 0 19 1 | y = Serge_Hefez
3 13 8 0 0 0 0 0 5 7 0 0 8 3 1 7 0 1 8 0 2 6 1 2 0 33 0 5 0 | z = Thomas_Soulie
10 4 5 3 0 2 9 0 1 0 4 2 0 0 0 1 0 10 4 0 0 4 4 1 0 4 27 5 0 | aa = Xavier_Leherpeur
4 12 4 1 0 6 2 2 2 0 6 2 1 0 3 6 4 1 0 2 0 0 0 1 8 2 3 95 0 | ab = Guillaume_Erner
1 0 1 1 0 2 0 1 1 3 1 11 2 1 6 0 1 0 0 6 13 0 3 3 0 0 0 2 41 | ac = Thomas_Legrand

```

Figure 17 : Matrice de confusion---Validation Croisée sous LMT

3.2.2 Confusion entre locuteurs

Limité par la méthode d'apprentissage, nous avons des difficultés pour comparer le niveau de confusion entre les locuteurs, c'est-à-dire, tel locuteur est très confondu avec un autre locuteur, par rapport aux autres.

Nous présentons le niveau de confusion par le « Ratio de confusion » pour que ce soit plus facile à comprendre et interpréter, parce que la matrice de confusion n'est pas une forme normalisée.

Ce ratio est calculé à partir de la matrice de confusion, avec la formule dessous :

$$\text{Ratio de confusion} = \frac{L_j(1...29)}{M_j}, (j=1,2,3,4...29);$$

Lj représente les valeurs de chaque ligne de matrice, soit les extraits passés en test d'un même locuteur. Mj représente la valeur qui se trouvent sur la diagonale de la matrice, c'est-à-dire les extraits correctement classifiés d'un locuteur.

Considérons le locuteur Jean-Marc Ayrault. Il porte le rôle « homme politique », un rôle bien distingué que ceux des autres si nous considérons son statut social, sa fonctionnalité devant le publique et entre ses collègues, ses missions professionnelles, etc.

Voici en dessus le graphe de niveau de confusion :

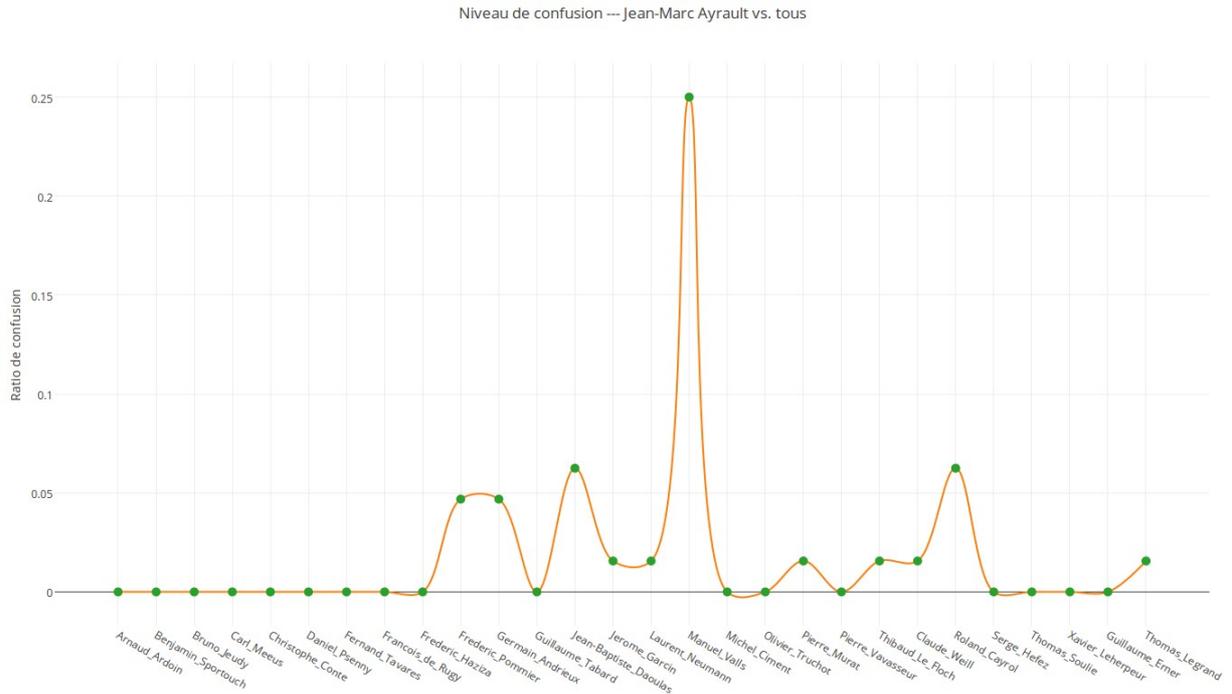


Figure 18 : Ratio de confusion---Jean-Marc Ayrault

Jean-Marc Ayrault est très confondu avec Manuel Valls, qui est également un homme politique comme nous l'avons noté au chapitre précédent et qui était également présenté un certain nombre des fois dans la même émission « topquestions », disons la même situation communicationnelle dans cette études.

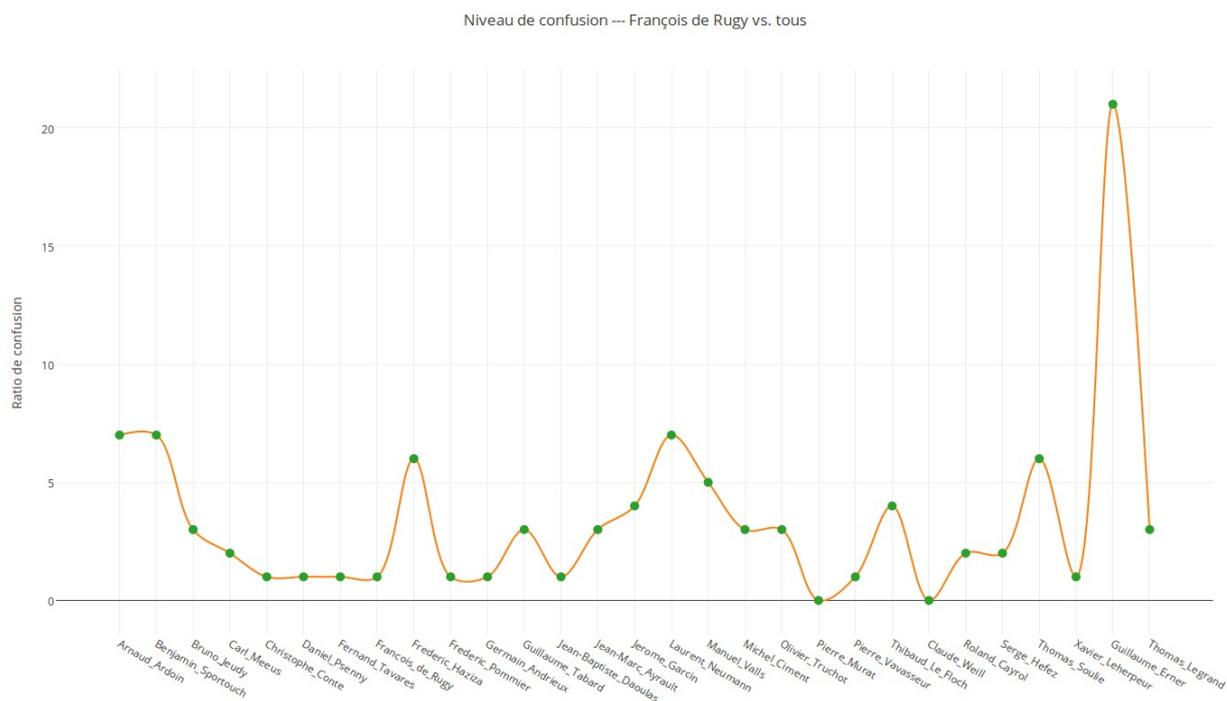


Figure 19 : Ratio de confusion---François de Rugy

Si on considère le ratio de confusion correspondant au François de Rugy. Verticalement, l'axe Y indique les valeurs de Ratio de confusion, varient normalement entre 0 et infini. Horizontalement, l'axe X, correspond à chaque classe, soit les locuteurs. Par contre, nous constatons ici que François de Rugy est très souvent classifié en tant que les autres locuteurs, tel que Guillaume Erner, Laurent Neumann, au lieu de lui-même.

Voici en plus, le troisième graphes de niveau de confusion entre les locuteurs, sous LMT. Ceil sélectionné d'après la précision et le taux de vrai positif qui sont relativement élevés. On peut remarquer qu'Arnaud Ardoin, qui est lui-même un journaliste, est plus confondu avec Christophe Conte, un chroniqueur.

Différent de François de Rugy, tous les ratios de confusion sont inférieurs à un. Cela indique que : le locuteur peut être bien distingué aux autres locuteurs qu'avec les traits rythmiques. De plus, nous pouvons observer la puissance du rythme sur la classification des signaux ou des locuteurs.

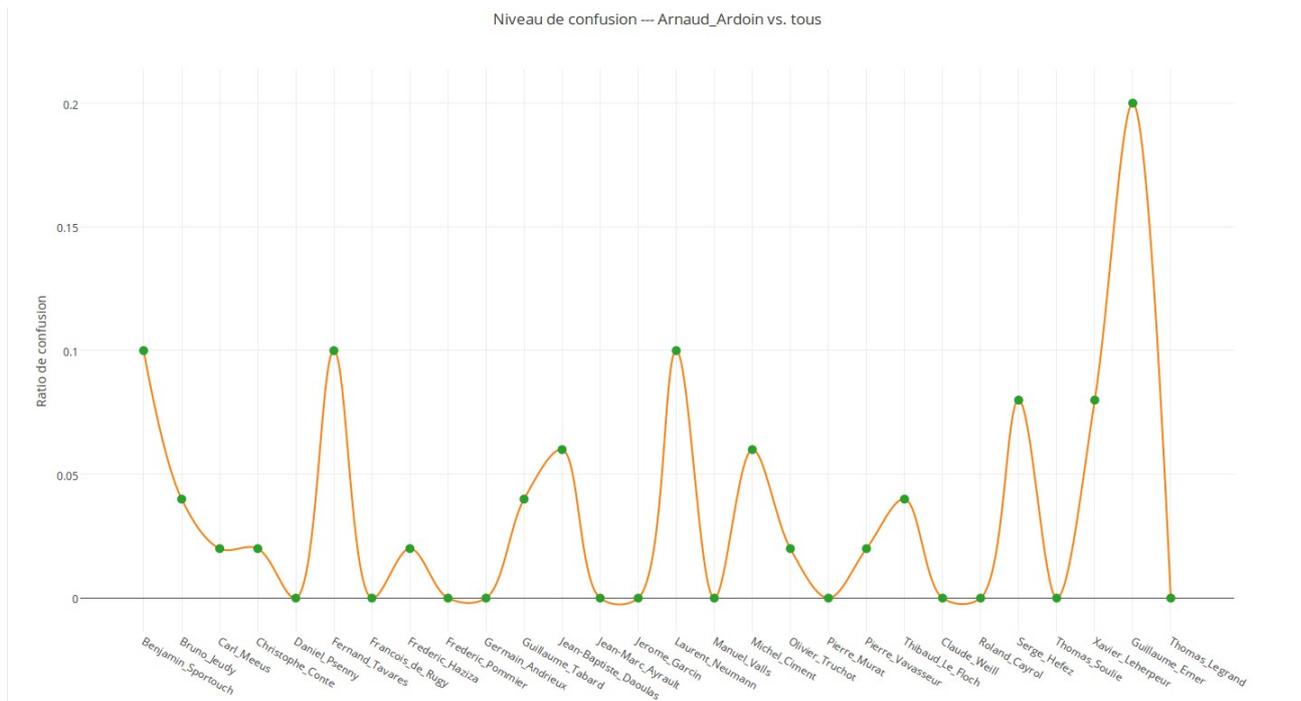


Figure 20 : Ratio de confusion---*Arnaud Ardoin*

3.2.3 Un test post-hoc sur la classification par LMT

Comme la robustesse des réseaux de neurones est très connue dans le domaine, nous avons aussi, effectué un test post-hoc pour montrer le niveau d'exactitude de notre classification faite sous LMT.

Voici en bas le tableau de statistique de la fonction *MultilayerPerception*, qui est une implémentation de réseaux de neurone sous Weka.

=== Evaluation on test split ===		
=== Summary ===		
Correctly Classified Instances	316	31.0413 %
Incorrectly Classified Instances	702	68.9587 %
Kappa statistic	0.2848	
Mean absolute error	0.0585	
Root mean squared error	0.1689	
Relative absolute error	87.8622 %	
Root relative squared error	92.5459 %	
Total Number of Instances	1018	

Tableau 8 : statistique de test post-hoc

Notons ici un taux de bonne classification égale à 31.0413 %, légèrement faible que celui de LMT, 32.1218%. Cela montre que les réseaux neurones sont aussi adéquates pour la

tâche et en plus, également très importantes pour nos études pour montrer que les traits rythmiques aient un impact significatif et stable.

3.3 Les macro-classes rythmiques et son impact sur le système

A partir de la structure de l'arbre LMT, nous pouvons trouver une cohérence entre la structure de l'arbre et la taille de l'effet des mesures (valeurs marquées entre parenthèses).

Structure de l'arbre LMT: -M 500

```

VarcoUV <= 81.77739 ( $\eta^2$ : 0.45485)
|
|   %VO <= 59.14389: LM_1:88/264 (560) ( $\eta^2$ : 0.42056)
|   %VO > 59.14389: LM_2:88/264 (1121)
|
VarcoUV > 81.77739: LM_3:88/176 (1312)
  
```

Nous avons fait un lien entre la macro-classe rythmique et le score du système RAL.

Tableau de croisement entre les macro-classes : Moyennes des scores TAR et les macro-classes

	Lm1	Lm2	Lm3
Lm1	11,4967068221935	5,15064021998401	7,19020598375138
Lm2	4,04238601433278	7,8182056637643	6,93539824313755
Lm3	7,61129804033532	5,7555017822539	8,16178475378811

Tableau 9 : interaction entre macro-classes

Nous constatons ici en dessus que les valeurs moyennes de scores sorties par le système lors des tests *Target*, se situent sur la diagonale de tableau de croisement. D'après ce tableau, nous pouvons noter que les moyennes de scores de tests *Target* sont plus élevées pour les paires d'extraits sonores qui sont dans la même macro-classe rythmique; les moyennes de scores inter classes sont relativement faibles par rapport à ceux d'intra classe. Pour être explicite, nous présentons aussi un tableau détaillé d'interaction de macro-classes rythmiques par sujet.

Selon le tableau, qui a présenté séparément et précisément l'impact des traits rythmiques sur chaque locuteur, nous pouvons noter que le rythme de la parole a une influence assez importante sur la tâche reconnaissance automatique du locuteur. La valeur VRAI indique que l'hypothèse d'interaction entre les classes est éprouvée, FAUX sinon. Nous avons une

interaction significative 68.97%, soit 20 sujets ont l'interaction significative, 9 non significative.

Par contre, le résultat total est significatif d'après notre hypothèse.

Comparaison des scores Inter et Intra macro-classe rythmique par sujet:

Locuteur	Inter	Intra	Total	S(Intra)>S(Inter)
Arnaud_Ardoin	21,3275362058	28,7872538374	26,2323495364	VRAI
Benjamin_Sportouch	-6,7175690091	-3,1118276607	-4,838941345	VRAI
Bruno_Jeudy	-0,8702928864	4,4370640988	2,6186242105	VRAI
Carl_Meeus	0,588033104	1,3552505282	1,0857169725	VRAI
Christophe_Conte	27,4261096261	27,2315607074	27,3263197868	FAUX
Claude_Weill	19,0062125343	19,5826771939	19,3539546501	VRAI
Daniel_Psenny	21,3504485324	21,7711612137	21,6157949953	VRAI
Fernand_Tavares	11,2917759786	11,5698254306	11,3832795255	VRAI
Francois_de_Rugy	-13,476970618	-7,1720390294	-10,1016032019	VRAI
Frederic_Haziza	24,7768972047	26,0791119265	25,4216907973	VRAI
Frederic_Pommier	---	-0,8050730806	-0,8050730806	FAUX
Germain_Andrieux	-8,1991790098	-0,6529760278	-5,7157922103	VRAI
Guillaume_Erner	5,5852674122	10,0350180865	7,9701745444	VRAI
Guillaume_Tabard	-2,6277129345	-2,5637720476	-2,5931331862	VRAI
Jean-Baptiste_Daoulas	-9,6402974364	-7,2427980974	-8,5219487549	VRAI
Jean-Marc_Ayrault	-0,3329544802	-0,9749162583	-0,8125453197	FAUX
Jerome_Garcin	9,0068322784	13,8379748268	11,6987742231	VRAI
Laurent_Neumann	-4,3145524788	-3,6229877276	-3,803073983	VRAI
Manuel_Valls	14,9406346056	14,5632738784	14,7179536593	FAUX
Michel_Ciment	7,6105615557	5,3761216463	6,6395958496	FAUX
Olivier_Truchot	6,6456855631	10,0749173249	8,3076506129	VRAI
Pierre_Murat	3,3974650806	2,8730222625	3,174762292	FAUX
Pierre_Vavasseur	9,0611124648	9,4101057368	9,2273601689	VRAI
Roland_Cayrol	16,5018291331	15,9163084175	16,0515104737	FAUX
Serge_Hefez	10,8222154158	14,0674860322	13,1148843745	VRAI
Thibaud_Le_Floch	0,0129884754	1,3784752326	0,7001455486	VRAI
Thomas_Legrand	9,080872765	7,9155684277	8,3098885822	FAUX
Thomas_Soulie	7,9750778873	16,2991430284	12,3054859853	VRAI
Xavier_Leherpeur	2,5799024091	2,0180307225	2,3529970552	FAUX
TotalRésultat	6,0943787496	8,4802807288	7,4447031057	VRAI

Tableau 10 : comparaison de score par sujet

Conclusion

Selon cette étude, nous pouvons conclure que le rythme est spécifique au locuteur. Le style de parole, plus particulièrement, les paramètres rythmiques ont un impact sur le système d'identification du locuteur.

D'après nos analyses, le rythme de la parole est influencé par le facteur locuteur, le facteur situation communicationnelle et le facteur rôle porté par le locuteur lors de la production de la parole. Cependant, ces trois facteurs n'ont pas l'influence au même niveau. Le facteur locuteur est à la place prédominante. La classification en locuteur d'après les paramètres rythmiques montrent encore une fois l'effet de ce facteur sur le rythme.

De plus, l'interaction entre les macro-classes rythmiques et le système a été trouvée. Ceci est un premier essai. L'étude ultérieure pourrait aussi être intéressante.

Bibliographie

1. Anne Catherine Simon, Antoine Auchlin, Mathieu Avanzi, Jean-Philippe Goldman. 2011. *Les phonostyles: une description prosodique des styles de parole en français*.
2. Ajili, Moez, Jean-François Bonastre, Juliette Kahn, Solange Rossato, et Guillaume Bernard. 2016. *FABIOLE, a Speech Database For Forensic Speaker Comparison*.
3. Ajili, Moez, Solange Rossato, Juliette Kahn, et al. « Inter-speaker variability in forensic voice comparison: a preliminary evaluation ». In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2114–2118. IEEE, 2016.
4. Baptiste Hupin, Anne-Catherine Simon. 2007. *Analyse phonostylistique du discours radiophonique*.
5. Bonastre, Jean-François, et Joseph P. Campbell. 2009. *Forensic Speaker Recognition*. IEEE SIGNAL PROCESSING MAGAZINE.
6. Bonastre, Jean-François, Frédéric Wils, et Sylvain Meignier. 2005. *ALIZE, a free toolkit for speaker recognition*. In *ICASSP (1)*, 737–740.
7. Bjursäter, Ulla. Department of Linguistics, Stockholm university. 2004. *Speaking styles and Phonetic variation*.
8. Caroline Émond et Lucie Ménard. 2010. *Les marques prosodiques des styles de parole dans les téléjournaux québécois*. Communication, Vol. 27/2.
9. Colonna Romain. 2014. *Des paroles, des langues, et des pouvoirs*. L'Harmattan.
10. Danielle Duez. 1999. *La fonction symbolique des pauses dans la parole de l'homme politique*. In: *Faits de langues*, n°13, Mars 1999. Oral-Ecrit : Formes et théories. pp. 91-97.
11. Dehak, N., Kenny, P., Dehak, R., Dumouchel, P., and Ouellet, P. *Front-end factor analysis for speaker verification*. Audio, Speech, and Language Processing, IEEE Transactions on, pp. 788-798 (2011).
12. Dellwo, Volker, Adrian Fourcin, et Evelyn Abberton. 2007. *Rhythmical classification of languages based on voice parameters*. Proceedings of ICPhS XVI, 1129–1132.
13. Dellwo, Volker, et Adrian Fourcin. 2013. *Rhythmic characteristics of voice between and within languages*. *Revue Tranel (Travaux neuchâtelois de linguistique)* 59: 87–107.

14. Dellwo, Volker; Leemann, Adrian; Kolly, Marie-José (2012). *Speaker idiosyncratic rhythmic features in the speech signal*. In: Interspeech 2012, Portland (OR), USA, 9 September 2012 - 13 September 2012, 1-4.
15. Eskenazi, M. (1993). *Trends in Speaking Style Research*. Keynote speech, Proceedings of Eurospeech'93.
16. Fant, GUNNAR, Anita Kruckenberg, et Lennart Nord. 1989. *Stress patterns, pauses, and timing in prose reading*. Speech Transmission Laboratory Quarterly Progress and Status Report, 1-7.
17. Feldstein, Stanley, et Ronald N. Bond. « Perception of speech rate as a function of vocal intensity and frequency ». *Language and Speech* 24, n° 4 (1981): 387-394
18. Furui, Sadaoki. 2005. *50 years of progress in speech and speaker recognition* . SPECOM 2005, Patras.
19. G. R. Doddington. 2001. *Speaker recognition based on idiolectal differences between speakers*. Proc. Eurospeech, pp. 2521-2524.
20. I. Pollack, J. M. Pickett, and W.H. Sumbly. 1954. *On the identification of speakers by voice*. Journal of the Acoustical Society of America, 26(3):403-406.
21. Fónagy, I. 1983. *La vive voix*. Essais de psycho-phonétique.
22. Fónagy, I. (1977) Le statut de la phonostylistique, *Phonetica*, 1-18.
23. Jean-Philippe Goldman, Tea Pršir, George Christodoulides, Antoine Auchlin. 2014. *Speaking style prosodic variation: an 8-hour 9-style corpus study*. Campbell, Gibbon, and Hirst (eds.) *Speech Prosody* 7.
24. Joaquim Llisterri . 1992. *Speaking styles in speech research*. p19.
25. Kerbrat-Orecchioni, C. 2003. *Les genres de l'oral : Types d'interactions et types d'activités*. Journée d'étude Les genres de l'oral. Lyon.
26. Kolly, Marie-José, Adrian Leemann, Philippe Boula de Mareüil, et Volker Dellwo. « SPEAKER-IDIOSYNCRASY IN PAUSING BEHAVIOR: EVIDENCE FROM A CROSS-LINGUISTIC STUDY ». Scotland, UK, 2015.
27. Landwehr, Niels, Mark Hall, et Eibe Frank. 2003. *Logistic model trees*. In European Conference on Machine Learning, 241-252. Springer.

28. LEON, P. 1993, Précis de phonostylistique, Parole et expressivité, Nathan Université, Paris.
29. Philippe Boula de Mareüil.. 2014. *Qu'est-ce qu'un (phono) style?* Cahiers de linguistique française, no 31 (2014): 9-19.
30. Sandrine Brognaux, Thomas Drugman. 2014. Cental, ICTEAM (Université catholique de Louvain), Belgium TCTS Lab (Université de Mons), Belgium. *Phonetic variations: Impact of the communicative situation.*
31. Sven Grawunder, Hans Engert, Helen Kaiser, Stephanie Unger, Ines Bose. 2008. Max Planck Institute for Evolutionary Anthropology, Leipzig. Martin Luther University, Halle, Germany. University of Trier, University of Cologne, Germany. *Pitch, speech rate and rhythm as speaking style characterizing parameters in German news presentation.*
32. Vaissiere, Jacqueline. 2008. « De la voix et du sens-autour de l'oeuvre de Iván Fónagy ». *Temps, Espaces, Langages: La Hongrie à la croisée des disciplines*, 56–70.
33. William labov. *Sociolinguistique*. 1972, université de Pennsylvania Presse. Collection Le sens commun. 1976, LES EDITIONS DE MINUIT.
34. Nick Campbell. 2007. *Differences in the Speaking Styles of a Japanese Male According to Interlocutor; Showing the Effects of Affect in Conversational Speech.*
35. Xu, Yi. 2013. *ProsodyPro—A tool for large-scale systematic prosody analysis*. In TRASP. Aix-en-Provence.

ANNEXE 1

```
files = dir('data/*_f0/*.f0');
n = 1;
for file = files'
    d = importdata(files(n).name);
    data = [d.data];
    data_length = length(data);
    t = data(1:data_length);
    time = t.'; % convertir le tab en vertical avec transpose .'

x = time;

% mesures des durees a partir des instants x;
duree=diff(x); % attention taille de x -1
% d'apres Dellwo, Fourcin 2013 Rhythmic ...
% duree < 20 ms = voisement
% detection des durees < 0.02
% I est un booleen, I = 1 quand la condition est vrai, 0 sinon.
I = duree < 0.02;

% Detection des SEGMENT VOISES pour lesquels , duree < 0.02 donc I vaut 1
% instants de debut des segments voises =
V = (I==1);
Vx = x(V);
% duree des segments voisee =
VOduree = duree(V);

UV = (I==0);
% instants de debut des segments non voises =
UVx = x(UV);
% duree des segments non voisee =
UVduree = duree(UV);
```

```

% creer 2 tableau pour stocker les valeurs de calcul
% TabV: Tableau des durees de segments voisies, la taille du tableau
% n'affecte pas le plot de graphe
TabVO = zeros(1000,1);

% TabNV: tableau de segments non voisies
TabUV = zeros(1000,1);

i = 1;
k = 1;
j = 1;

% utiliser la longueur de tableau duree comme indice de calcul
d = length(duree);

% quand !eof(duree), faire le calcul de duree et les stocker dans deux tabs
while (i<= d)
    % initialiser la valeur de s
    s = 0;

    %les segs voisies sont mesure consecutivement
    while ((i< d) && I(i)~=0) % quand c'est un seg voisie
        s = s+duree(i);% concatener les segs voisies pour avoir une seule duree
        i = i+1;
    end

    if s ~= 0
        TabVO(j) = s;
        j = j+1;
    end

    %les segs non voisies sont mesures discretement
    if I(i) == 0
        TabUV(k) = duree(i);

```

```

        k = k+1;
    end
    i = i+1;
end
n = n+1; % deplacer le handle au prochain fichier

% Matrice de UV, structure: la duree d'un UV en miliseconde, temps debut de ce seg, fin
de ce seg;
duree_en_ms=UVduree*1000;
End_UV=UVx+UVduree;
M_UV=[duree_en_ms UVx End_UV];

% Cette boucle trace les pauses, d'apres Zellner 1992, le UV >=200 est
% rarement trouve au milieu d'un mot, donc on le considere comme une pause
% (silencieuse)

h=1;
m=1;
pause=zeros(30,3);

while h<=length(M_UV)

if M_UV(h,1)>=300;
    pause(m,:)=M_UV(h,:);
    m=m+1;
end

h=h+1;
end
disp(pause(1:(m-1),:));
% Trace les debuts et les fins des seg VO
Debut_VO=UVx-TabVO(1:(k-1));
End_VO=UVx;
Dernier_VO_debut=Vx(end)-TabVO(k);

```

```

Dernier_VO_fin=Vx(end);
% dans le matrice M_VO, le dernier elt de VO est raté. Mais il est
% important.
M_VO=[TabVO(1:(k-1)) Debut_VO End_VO];

name = files(n-1).name;
string = strsplit(name, '.'); % elimine l'extension .f0
Moy_VO = mean(TabVO(1:k));
VO_proportion = sum(TabVO(1:k))/(sum(TabVO(1:k))+sum(TabUV(1:j)))*100; %VO
VarcoVO=(std(TabVO(1:k)))/Moy_VO*100;

%D'apres Dellwo, une correlation entre VarcoUV et %VO
VarcoUV=(std(TabUV(1:k)))/mean(TabUV(1:k))*100;
ProportionUV = sum(TabUV(1:k)>TabVO(1:k))/k*100;

%moyenne de la duree de "interstress interval"
moy_VOUV=mean(TabVO(1:k)+TabUV(1:k));
VarcoPair = (std(TabVO(1:k)+TabUV(1:k)))/moy_VOUV*100;

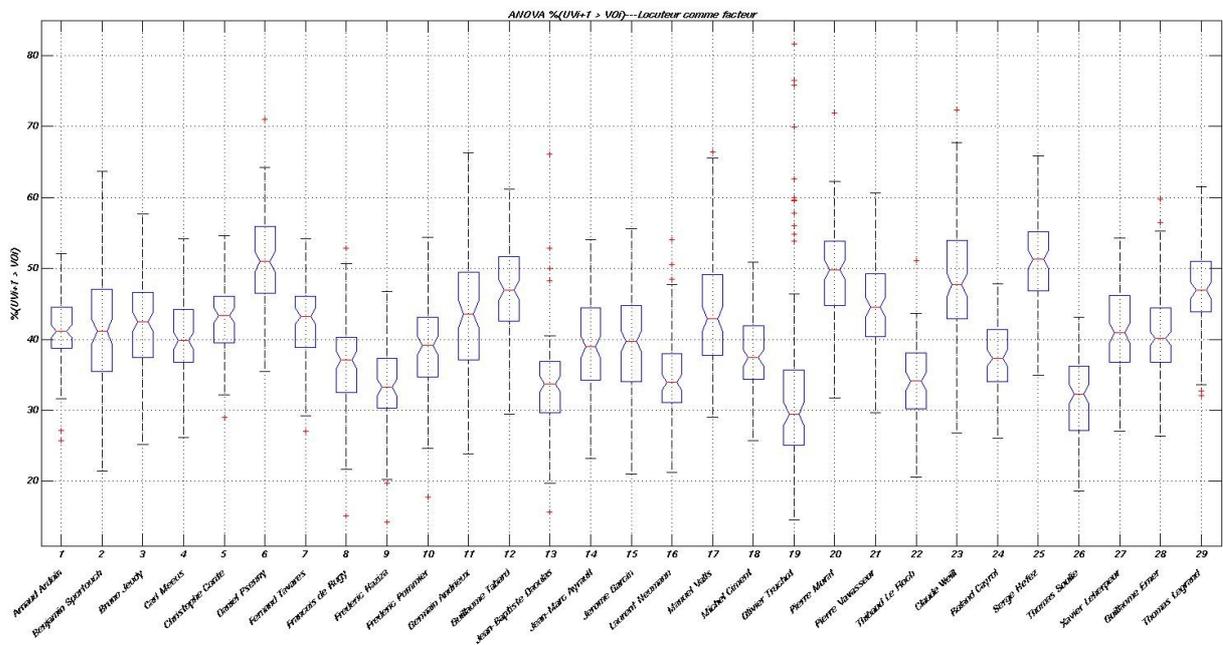
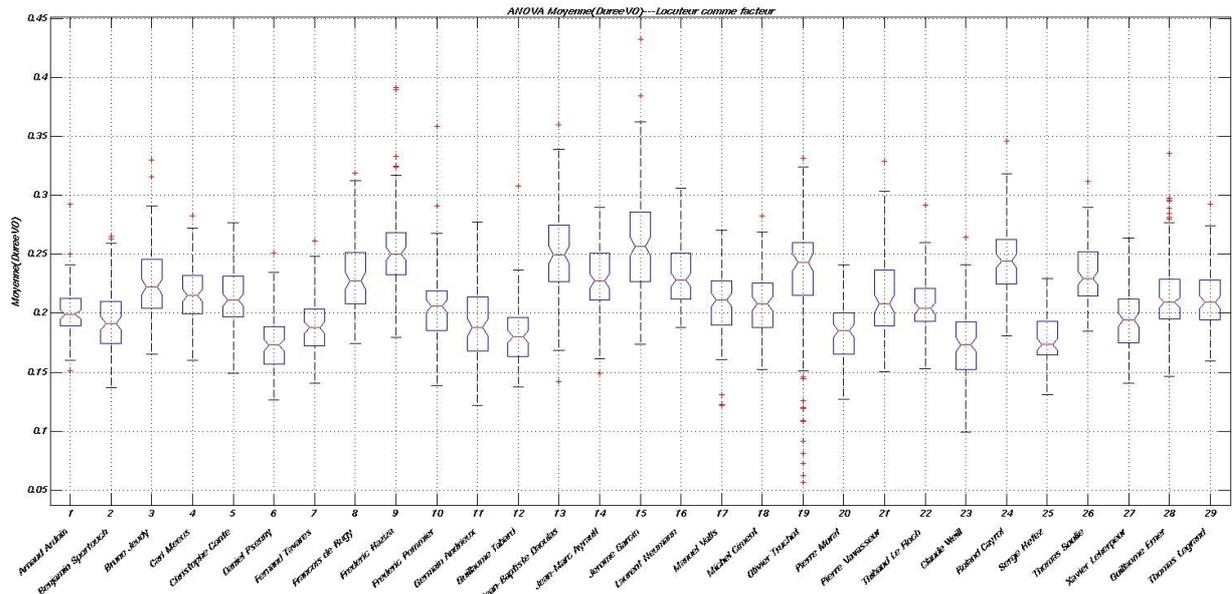
%proportion
%ecrire les donnees dans un fichiers
fileID = fopen('Nom_Loc_Rythme_mesure.csv','a+');
fprintf(fileID, '\n');
fprintf(fileID, '%6s\t', string{1}); % {} permet de acceder le contenu d'un cell au lieu de ()
fprintf(fileID, '%6.5f\t', VO_proportion);
fprintf(fileID, '%6.5f\t', Moy_VO);
fprintf(fileID, '%6.5f\t', VarcoVO);
fprintf(fileID, '%6.5f\t', VarcoUV);
fprintf(fileID, '%6.5f\t', ProportionUV);
fprintf(fileID, '%6.5f\t', moy_VOUV);
fprintf(fileID, '%6.5f\t', VarcoPair);
fclose(fileID);
end

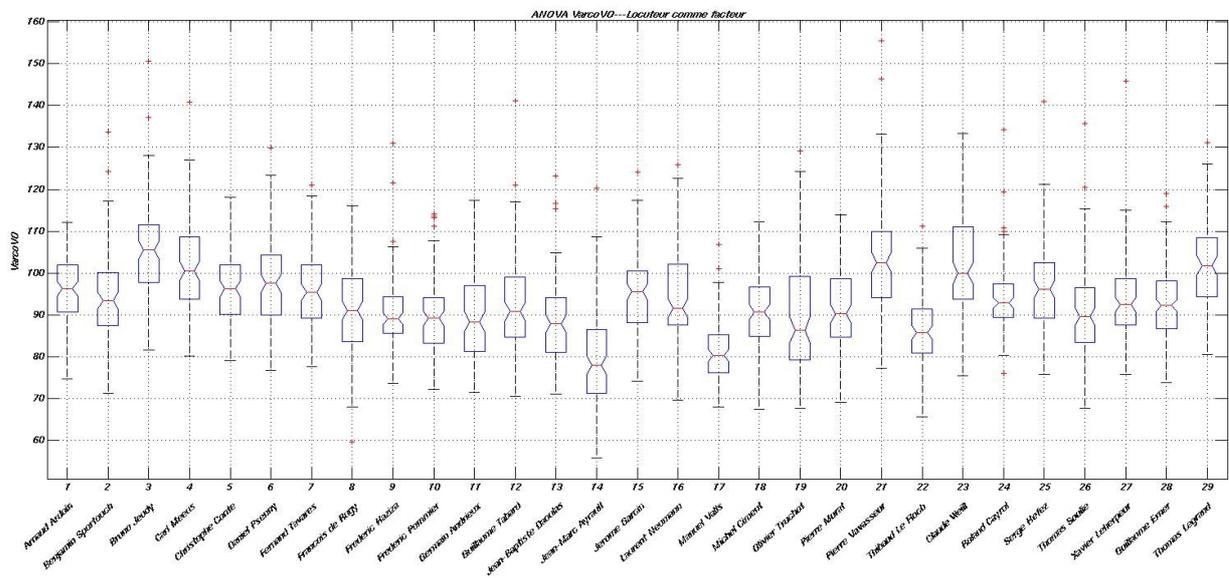
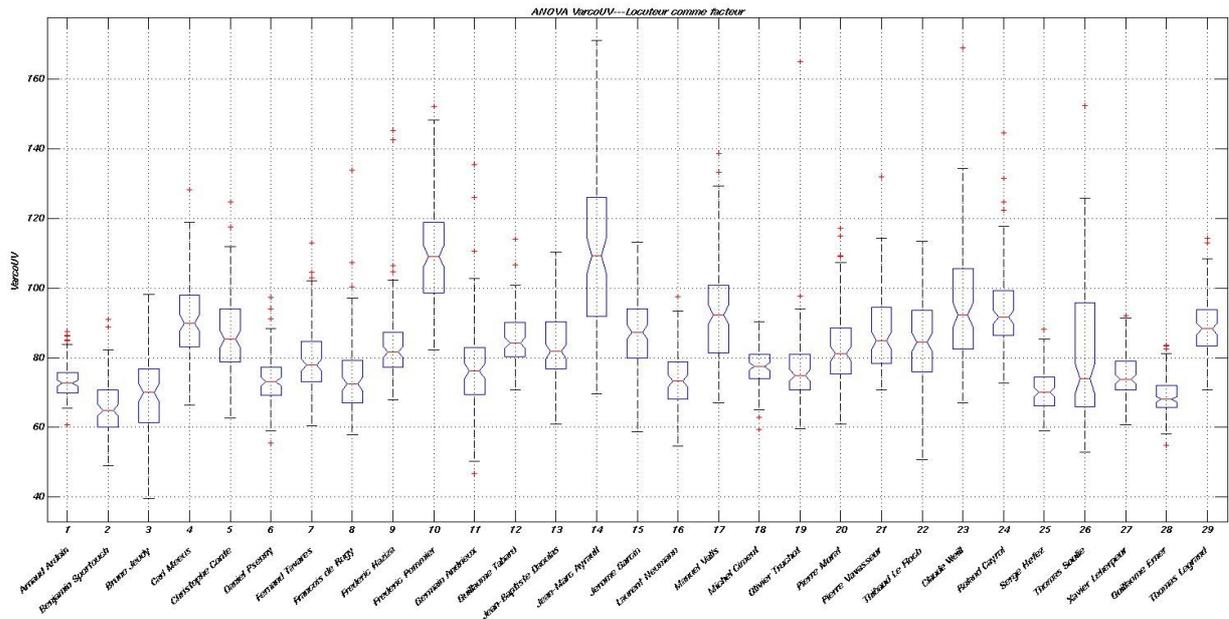
```

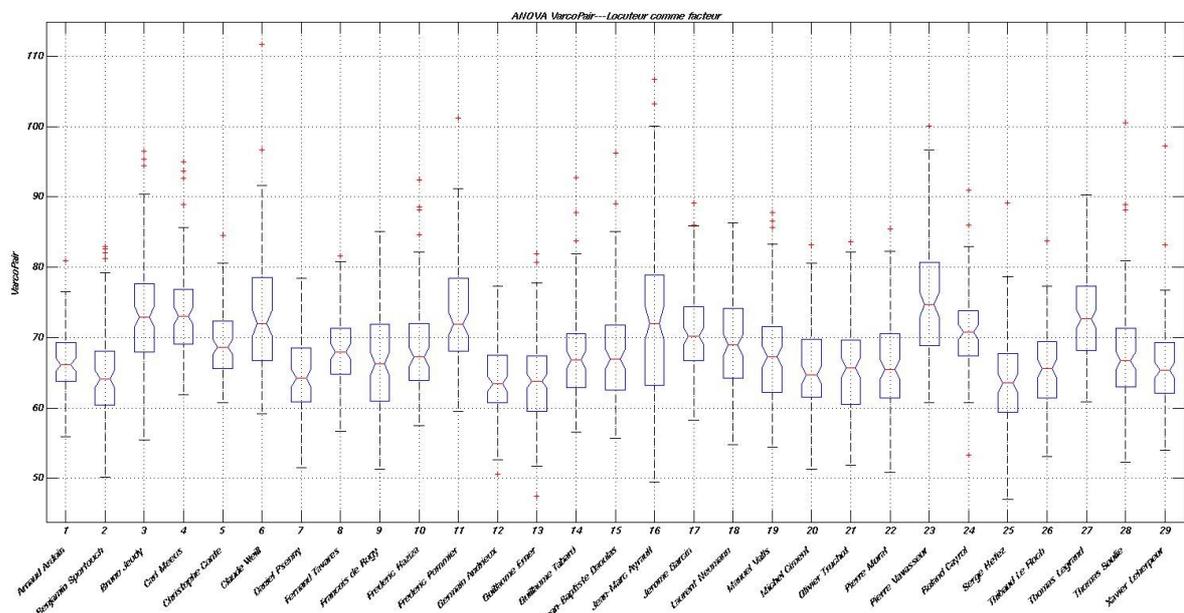
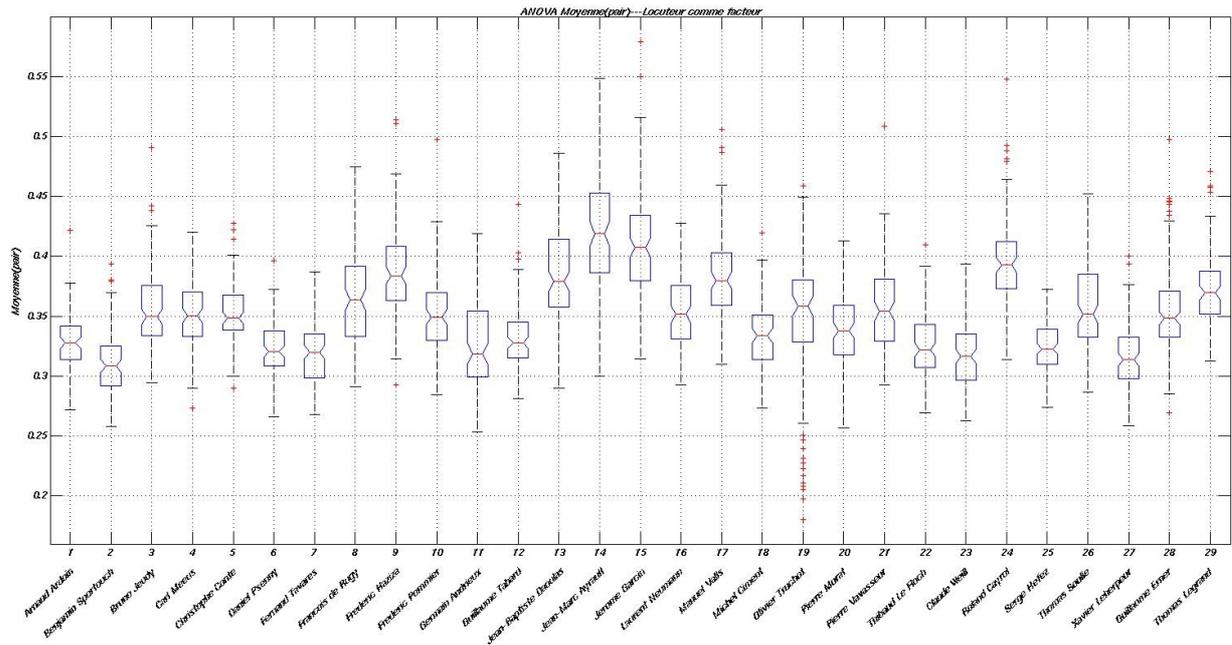
ANNEXE 2

Référence de locuteurs d'annexe 2 (de gauche à droite) :

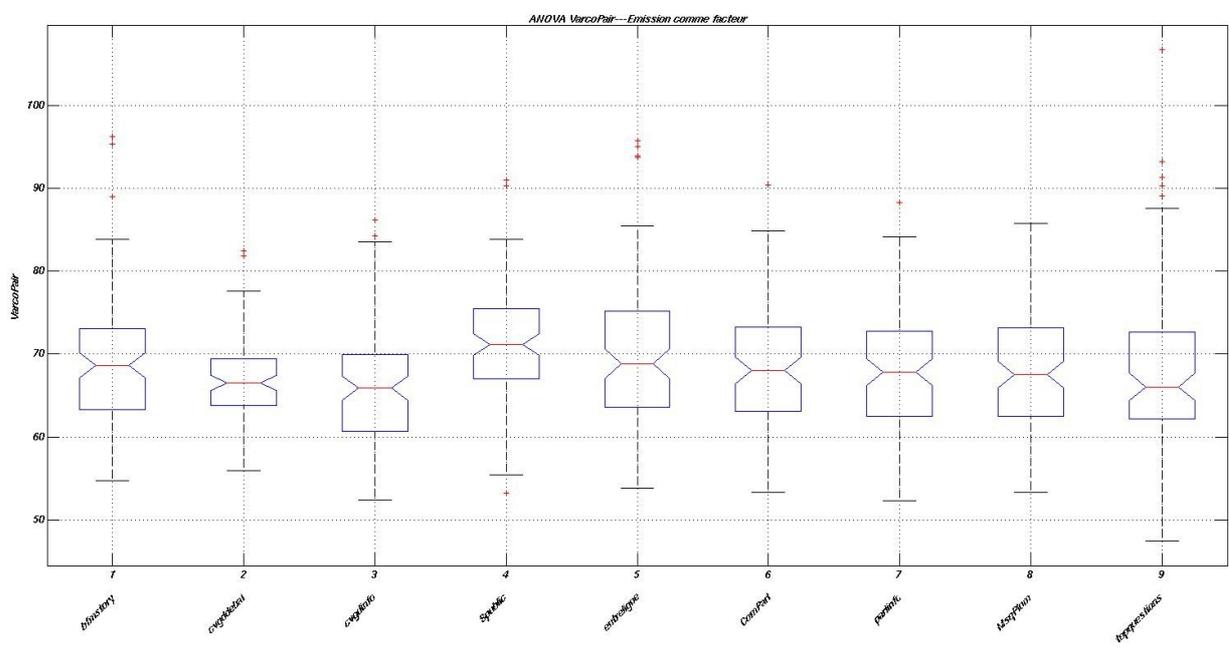
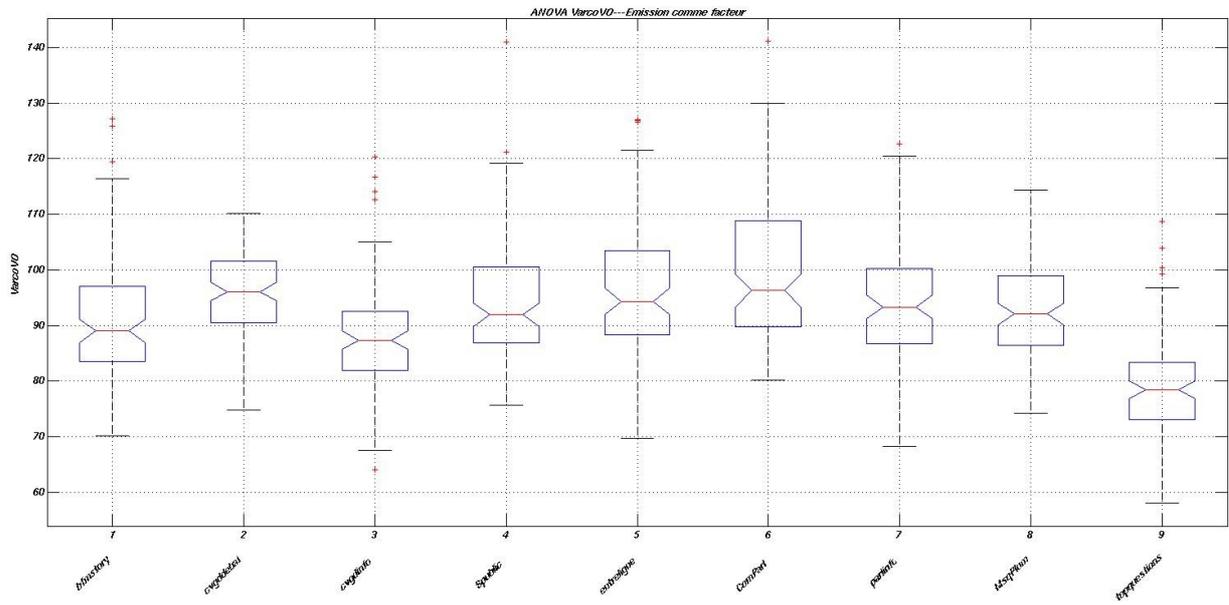
1. Arnaud Ardoin
2. Benjamin Sportouch
3. Bruno Jeudy
4. Carl Meeus
5. Christophe Conté
6. Daniel Psenny
7. Fernand Tavares
8. François de Rugy
9. Frédéric Haziza
10. Frédéric Pommier
11. Germain Andrieux
12. Guillaume Tabard
13. Jean-Baptiste Daoulas
14. Jean-Marc Ayrault
15. Jérôme Garcin
16. Laurent Neumann
17. Manuel Valls
18. Michel Ciment
19. Olivier Truchot
20. Pierre Murat
21. Pierre Vavasseur
22. Thibaud Le Floch
23. Claude Weill
24. Roland Cayrol
25. Serge Hefez
26. Thomas Soulié
27. Xavier Leherpeur
28. Guillaume Erner
29. Thomas Legrand

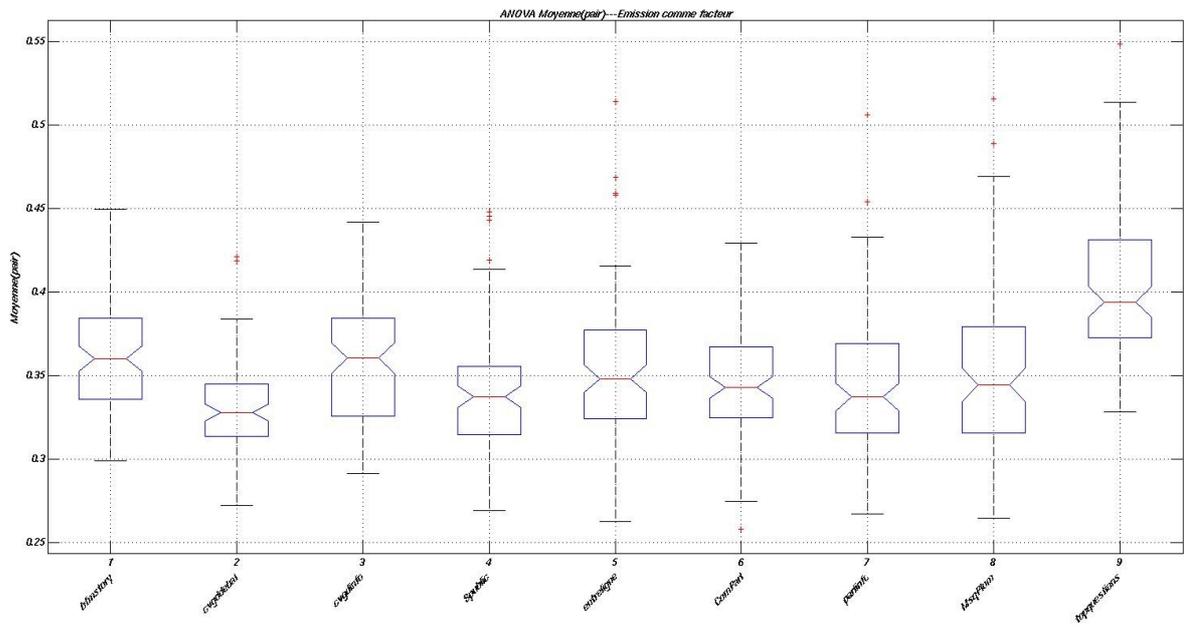
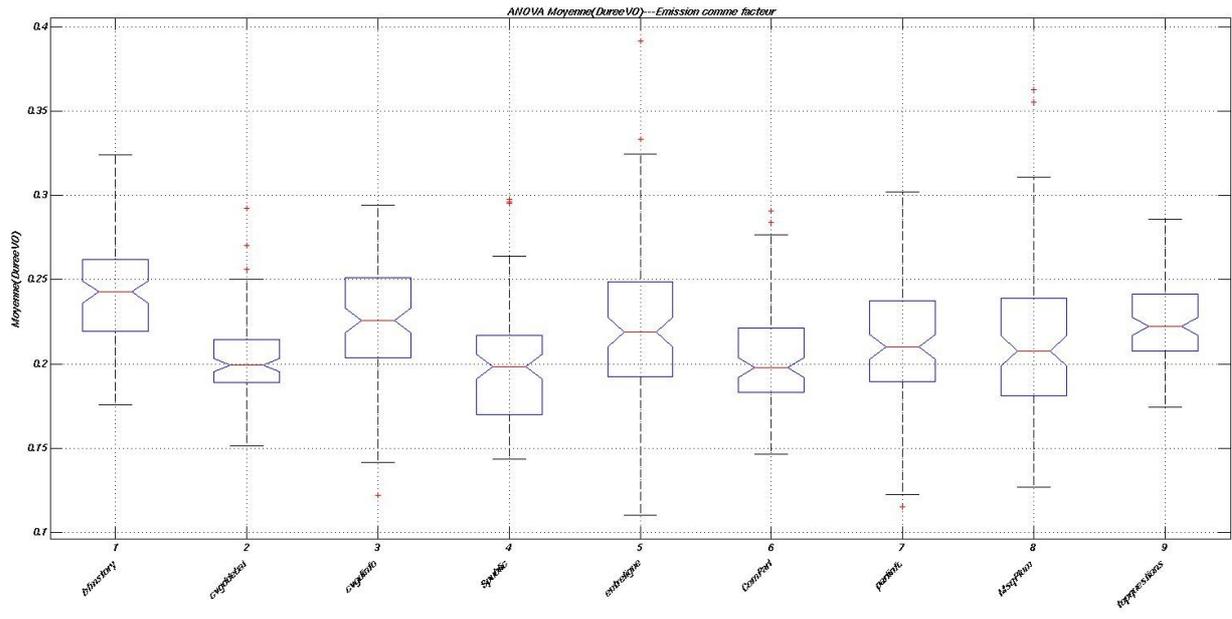


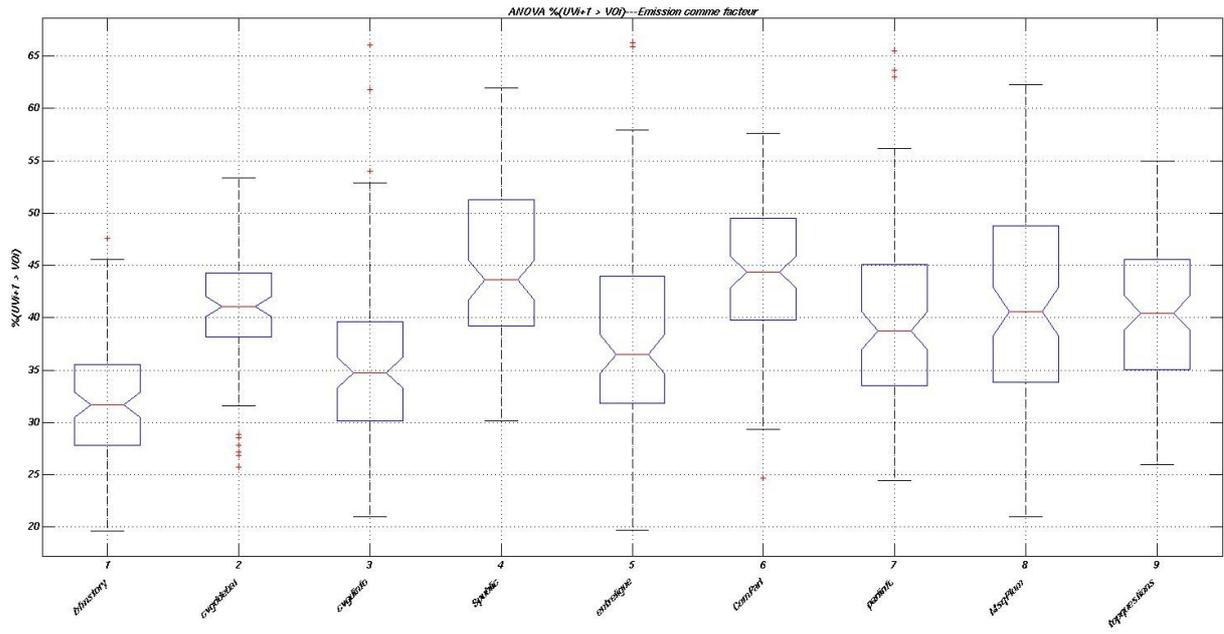




ANNEXE 3







ANNEXE 4

Détails de la classification sous LMT

==== Run information ====

Scheme:weka.classifiers.trees.LMT -I -1 -M 500 -W 0.0

Relation: 29_loc_toutes_mesures-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-R1

Instances: 2993

Attributes: 8

Nom

VO

moydVO

VarcoVO

VarcoUV

dUVdVO

moydVodUV

VarcoPair

Test mode:split 66.0% train, remainder test

Time taken to build model: 42.14 seconds

==== Evaluation on test split ====

==== Summary ====

Correctly Classified Instances	327	32.1218 %
Incorrectly Classified Instances	691	67.8782 %
Kappa statistic	0.2966	
Mean absolute error	0.055	
Root mean squared error	0.1675	
Relative absolute error	82.6007 %	
Root relative squared error	91.794 %	
Total Number of Instances	1018	

Table des matières

Remerciements.....	4
Introduction.....	6
Chapitre 1 : État de l'art et le contexte de recherche.....	7
1.1 Styles de parole.....	7
1.1.1 Les définitions du style de parole.....	7
1.1.2 Caractériser les styles de parole.....	9
1.1.3 Les paramètres prosodiques utilisés.....	13
1.2 Le rythme.....	17
1.2.1 Définition du rythme de la parole.....	17
1.2.2 Les caractéristiques rythmiques de la parole.....	18
1.3 Les systèmes de RAL.....	20
Conclusion Chapitre 1.....	22
Problématique.....	23
Chapitre 2 : Méthodologies et expérimentations.....	24
2.1 Introduction base de données.....	24
2.1.1 Base de données : Le projet et corpus FABIOLÉ.....	24
2.1.2 Émissions et rôle du locuteur.....	25
2.2 Mesures des paramètres rythmiques.....	27
2.2.1 Mesures des paramètres rythmiques de la parole.....	27
2.2.3 Une méthode semi-automatique pour les mesures des paramètres rythmiques.....	30
2.3 Méthodologies d'analyse d'influence des facteurs.....	32
2.3.1 Les facteurs étudiés.....	32
2.3.2 ANOVA à un facteur.....	33
2.4 Classification en locuteur à partir des données rythmiques.....	33
2.5 Influence du rythme sur l'identification du locuteur.....	36
2.5.1 Définition de macro-classes rythmiques basée sur LMT.....	36
2.5.2 Association de la macro-classe avec un extrait sonore.....	37
Chapitre 3 : Résultats.....	39
3.1 Influences des facteurs sur le rythme.....	39
3.1.1 Le facteur locuteur.....	39
3.1.2 Influence de la situation communicationnelle : le facteur émission.....	42
3.1.3 L'effet du rôle du locuteur sur le rythme de la parole.....	43

3.2 Validation croisée en locuteur et le test post-hoc.....	45
3.2.1 La classification en locuteur par LMT.....	45
3.2.2 Confusion entre locuteurs.....	48
3.3 Les macro-classes rythmiques et son impact sur le système.....	53
Conclusion.....	55
Bibliographie.....	56
ANNEXE 1.....	59
ANNEXE 2 Référence de locuteurs d'annexe 2 (de gauche à droite) :.....	63
ANNEXE 3.....	67
ANNEXE 4.....	70

Index des tableaux

Tableau 1 : résumé des paramètres pour la caractérisation des styles de parole	11
Tableau 2 : tableau de résumé des styles de parole d'après différents critères.....	13
Tableau 3 : comparaison d'algorithmes pour la classification.....	34
Tableau 4 : seuils de LMT.....	36
Tableau 5 : tableau d'ANOVA, locuteurs comme facteur.....	39
Tableau 6 : liste d'êta-carré des 7 paramètres en fonction de locuteurs.....	41
Tableau 7 : liste d'êta-carré des 7 paramètres en fonction d'émission	43
Tableau 8 : statistique de test post-hoc.....	51
Tableau 9 : interaction entre macro-classes.....	53
Tableau 10 : comparaison de score par sujet.....	54

Index des illustrations

Modèle de regroupement prosodique.....	14
Diagramme de la dispersion des métiers des locuteurs.....	25
Distribution des extraits des 29 locuteurs étudiés sous chaque émission.....	26
Un segment voisé et un segment non voisé d'un signal.....	27
Capture d'écran console ProsodyPro.....	31
Exemple fichier NomWav.f0.....	32
Extrait de la structure de données pour la classification.....	36
Classification en fonction de locuteurs par LMT structure de l'arbre.....	37
Boîte à moustache---%VO locuteur comme facteur.....	40
Radar-chart d'éta-carré des paramètres---locuteur comme facteur.....	40
Boîte à moustache---%VO Émission comme facteur.....	42
Boîte à moustache---VarcoUV émission comme facteur.....	43
Boîte à moustache---%VO Jean-Marc Ayrault vs. Manuel Valls.....	44
Statistique de validation croisée par LMT.....	45
Détails de validation croisée en locuteur sous LMT.....	47
Taux de vrai positif---Classification en locuteur par LMT.....	48
Matrice de confusion---Validation Croisée sous LMT.....	48
Ratio de confusion---Jean-Marc Ayrault.....	49
Ratio de confusion---François de Rugy.....	50
Ratio de confusion---Arnaud Ardoin.....	51

RÉSUMÉ

La variation de la parole et l'idiosyncrasie du locuteur sont deux facteurs importants qui influencent la performance du système RAL. La notion de styles de parole est un point d'unification des deux facteurs. Nous nous intéressons dans cette étude l'information du locuteur portée par les paramètres rythmiques, et en même temps, l'impact des styles de parole en particulier les traits rythmique de la parole sur l'identification du locuteur. Sept mesures sont appliquées afin d'obtenir les paramètres rythmiques. Les macro-classes rythmiques sont définies selon la classification. Par rapport à cette étude, le rythme a été prouvé spécifique au locuteur. De plus, il est aussi influencé par le facteur de situation communicationnelle et le rôle du locuteur. Les paramètres rythmiques montrent aussi sa puissance en classification en locuteur. Les macro-classes rythmiques sont démontrées d'avoir une interaction avec le système d'identification.

MOTS-CLÉS : rythme, styles de parole, variabilités intra-inter locuteur, idiosyncrasie, identification du locuteur.

Abstract

The variation of the speech and the idiosyncrasy of the speaker are two important factors which are supposed to have influence on the performance of the Speaker Recognition System. The notion « speaking styles » is the point who can join these two factors. In this research, we are interested in the rythme parameters of speech which carry the speakers' information. At the same time, the impact on the SR system of the speaking styles, especially the rythme parameters of the speech are also of high interest. Seven measures are implemented in this study in purpose of finding the rythme parameters in the speech. The rythme macro-classes are defined according the the classification results. The rythme of speech is demonstrated to be different between speakers. The factor « communication situation » and the « rôle of the speaker » also have influence on the rythme parameters, however, the speaker factor is proved to be the main one. The rythme parameters are relatively efficient for the classification task. The macro-classes are proved to have an interaction with the SR system performance.

KEY-WORDS : rythme, speaking styles, intra-inter speaker variabilities, idiosyncrasy, speaker recognition.