



HAL
open science

Variabilité intra-individuelle en parole lue

Cecile Fougeron, Nicolas Audibert

► **To cite this version:**

Cecile Fougeron, Nicolas Audibert. Variabilité intra-individuelle en parole lue. XXXIVe Journées d'Études sur la Parole – JEP 2022, Jun 2022, Noirmoutier, France. pp.145-153, 10.21437/JEP.2022-16 . hal-03928095

HAL Id: hal-03928095

<https://hal.science/hal-03928095>

Submitted on 8 Jan 2024

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.



Variabilité intra-individuelle en parole lue

Cécile Fougeron & Nicolas Audibert

(1) Laboratoire de Phonétique et Phonologie (UMR7018 CNRS/Univ. Sorbonne-Nouvelle), 19 rue des Bernardins, 75005 Paris, France
cecile.fougeron, nicolas.audibert@sorbonne-nouvelle.fr

RÉSUMÉ

Nos connaissances sur la parole sont historiquement construites sur des données comparant différents locuteurs ou moyennées entre locuteurs. On sait donc peu de choses sur la variabilité de la parole chez un même individu : dans quelle mesure varie-t-on d'une répétition à l'autre et sur quelles dimensions ? Dans cette étude, nous documentons la stabilité des caractéristiques de parole chez 9 individus, sur la lecture de deux textes enregistrés sur une dizaine de sessions des jours différents sur une période de deux mois. 21 descripteurs liés à des aspects temporels, spectraux, de f_0 , d'harmonicité ainsi que leur modulation au cours de la lecture sont étudiés. La stabilité de ces descripteurs entre sessions est évaluée en comparaison avec leur variabilité entre locuteurs et entre textes. Les caractéristiques de parole variant le plus d'une session à l'autre sont aussi celles qui séparent le mieux les différents locuteurs dans cette cohorte.

ABSTRACT

Intra-individual variability in read speech.

Our knowledge of speech is historically built on data comparing different speakers or averaged across speakers. We therefore know little about the variability of speech in the same individual: to what extent does it vary from one repetition to another and on what dimensions? In this study, we document the stability of speech characteristics in 9 individuals, on the reading of two texts recorded over ten sessions on different days over a two-month period. 21 features related to temporal, spectral, f_0 , harmonicity aspects as well as their modulation during reading are studied. The stability of these features between sessions is evaluated in comparison with their variability between speakers and between texts. The speech features that vary the most between sessions are also those that best separate the different speakers in this cohort.

MOTS-CLÉS : variabilité intra-locuteur, variabilité inter-locuteur, français

KEYWORDS: intra-speaker and inter-speaker variability, French.

1 Introduction

Nos connaissances sur la parole sont historiquement construites sur des données comparant différents locuteurs ou sur des données moyennées entre locuteurs. Par conséquent, on sait peu de choses sur la variabilité de la parole chez un même individu. Pourtant on sait que les locuteurs s'adaptent aux contextes linguistiques et de parole, et modifient leur discours en fonction de leur état émotionnel ou biologique, par exemple. Cependant, il n'est pas clair dans quelle mesure les locuteurs varient

d'une répétition à l'autre, sur quelles caractéristiques de parole et dans quelle mesure les enregistrements recueillis à quelques jours, mois ou années d'intervalle sont « naturellement » variables.

Si la variabilité de la parole est omniprésente, elle est aussi, du moins en partie, structurée et régie par des règles. Par conséquent, documenter la variation dans le discours et expliquer ses origines est au cœur de la recherche en phonétique. Deux types de variabilité peuvent être distingués en fonction de leurs origines et de leurs caractéristiques : la variabilité inter-locuteur et la variabilité intra-locuteur (Wright 2006). D'un côté, la *variabilité inter-locuteur* est enracinée dans des traits physiologiques ou anatomiques spécifiques au locuteur, ainsi que dans des facteurs démographiques, régionaux ou sociaux (Foulkes & Docherty 2006). Cette variabilité entre locuteurs est porteuse d'informations indexicales sur le locuteur : cette indexicalité est disséminée sur un faisceau de caractéristiques de parole et c'est généralement ce faisceau qui permet de discriminer un locuteur (ou un groupe de locuteurs) d'un autre. En outre, la variabilité inter-locuteur est censée rester constante au travers des conditions de parole, une stabilité qu'il reste pourtant à préciser. D'un autre côté, la *variabilité intra-locuteur* provient de diverses sources (voir par exemple Bürki 2018 pour une revue). Parmi celles-ci, les changements dans la réalisation phonétique des phonèmes, dus aux contextes phonétiques ou prosodiques, ont été les plus documentés et cette variabilité intra-locuteur est plus facile à modéliser. Ce n'est pas le cas des autres sources de variabilité intra-locuteur qui sont dues au fait que les individus adaptent leur discours au contexte de parole, à sa formalité, à la relation sociale qu'ils ont avec leurs interlocuteurs, aux besoins spécifiques des auditeurs (dans un contexte bruyant par exemple, ou s'ils partagent des connaissances communes sur le sujet ou non), etc. Ces différents facteurs, souvent regroupés sous la notion « d'effets de style », influencent principalement le degré de précision articulatoire (ou hyper/hypo articulation) et le débit de parole adopté par le locuteur. Un troisième type de facteur affectant la façon dont un locuteur parle est lié à l'état affectif et cognitif de celui-ci (par exemple l'émotion, l'attention, la fatigue..., entre autres Gelfer 1991, Barret & Paus 2002, Johnstone & Scherer 2000).

Bien qu'ils soient reconnus et documentés d'une manière ou d'une autre, les changements dans le discours résultant du contexte de parole ou de l'état du locuteur sont difficiles à délimiter et à manipuler expérimentalement (voir par exemple le travail sur le « clear speech » de Scarborough & Zellou, 2013). De plus, notre connaissance de la variation interne au locuteur est encore compliquée par le fait que tous ces effets ne sont pas statiques mais changent dynamiquement au cours d'un discours (Bates 2003). En effet, les conditions environnementales, le contenu du discours, les propriétés prosodiques affectant la prononciation ainsi que la fatigue, les émotions, l'excitation ou l'attention varient d'un moment à l'autre. Par conséquent, une croyance commune et répétée en sciences du langage est qu'une unité de parole ne sera jamais prononcée deux fois de la même manière par un même individu. Cependant, on sait peu de choses sur l'ampleur des divergences de la parole d'une itération à l'autre, sur quels aspects, sur le degré de granularité nécessaire pour mesurer ces variations, ou sur les facteurs qui contribuent à la variabilité d'une itération à l'autre. Heald & Nusbaum (2015), par exemple, ont exploré la variabilité de la production de voyelles isolées pour huit locuteurs sur 9 répétitions recueillies soit le même jour à trois moments différents, soit des jours différents. Ils ont constaté une plus grande variabilité de l'acoustique des voyelles entre les différentes sessions d'un même jour qu'entre les jours, et proposent un large ensemble de causes possibles pour ce phénomène. En criminalistique, où les enregistrements d'un individu doivent être comparés, ces questions semblent cruciales. En effet, à notre connaissance, nous ne disposons d'aucun argument scientifique concernant la validité d'une comparaison entre des enregistrements espacés de plusieurs semaines, mois ou années et sur les conditions permettant de telles comparaisons.

Dans le cadre d'un projet de recherche plus vaste sur la variabilité intra et inter-individuelle sur des bases de données collectées au LPP sur les mêmes locuteurs avec différents délais entre les sessions (bases PATATRA et PATAFreq), l'étude présentée ici se penche sur la stabilité des caractéristiques de parole lue sur des sessions rapprochées dans le temps. Afin de donner un sens à cette variabilité inter-sessions, observée sur 21 indicateurs acoustiques, nous la comparons à celle liée aux différents contenus des textes lus et aux différents locuteurs.

2 Méthode

2.1 Corpus et annotation

Les données de parole utilisées pour cette étude viennent de la base d'enregistrement PATAFreq qui a été collecté lors du premier confinement en 2020 par 9 locuteurs, dont 6 femmes, tous natifs du français, proches du milieu de la recherche sur la parole, et participant à la base d'enregistrements longitudinaux PATATRA coordonnée par le LPP. Ces 9 locuteurs se sont enregistrés régulièrement, avec un objectif d'environ 10 enregistrements sur environ un mois et demi. Cinq locuteurs sont parvenus à 10 sessions, 3 locuteurs à 8 sessions, et une locutrice qui a rencontré des problèmes techniques n'a pu effectuer que 6 sessions. Les locuteurs étaient invités à s'enregistrer régulièrement, avec un minimum de 24h entre les sessions et sur une période maximale de deux mois, ils étaient encouragés à changer le moment de la journée mais les contraintes familiales pendant le confinement ne l'ont pas toujours permis. Tous les intervenants étaient équipés de cartes son et microphones professionnels devaient s'enregistrer dans un environnement calme, toujours dans la même pièce si possible, et avec les mêmes réglages. Tout changement, ou bruit environnemental inattendu, devait être signalé dans un questionnaire à remplir après chaque enregistrement, avec une auto-évaluation sur une échelle à 4 niveaux de son état de fatigue, de son état émotionnel, de sa fatigue vocale et de la fréquence d'utilisation de sa voix le jour de l'enregistrement.

#	Texte1	Nsyl	Texte 2	Nsyl
1	o nɔʁ dy pei ð tʁuv yn espes də	11	ma sœʁ ɛ vœny ʃe mwa ijœʁ puʁ pʁɑ̃dœʁ lə te	13
2	ʃa dɔ̃ la kœ ɛ tʁɛ kuʁt il	8	ɛl mə pœʁlɛ də se vakɑ̃s ɑ̃ mœʁ dy nœʁ	12
3	sɔ̃ nwaʁ avek dœ taf blɑ̃ʃ syʁ lə do	10	lœʁskə dɑ̃ notœ do tɔ̃ba ẽ pœti twazo	13
4	lœʁ pwal ɛ bo e du zyst a kote	10	se dœ zel ete blese e il	9
5	vi yn koloni dwazo dɔ̃ le ni	10	ave kœsy ẽ ku vjɔlɑ̃ syʁ la kœ	11
6	sɔ̃t akœʁʃe o bœʁ də la falez	10	sɔ̃ kœʁ bate tʁɛ vit mɛ il ete tɑ̃ vi	12
7	il dwav fœʁ atɑ̃sjɔ̃ a nə pa fœʁ tɔ̃be lœʁ zœ	12	sɔ̃ plymaz ete bo e du zœ mapœʁʃe	13
8	dɑ̃ la mœʁ ma sœʁ na ka tʁavœʁse	12	dy bœʁ də la fœnetœʁ puʁ kœgavde dɑ̃ la ky	13
9	la ky puʁ vɑ̃kɔ̃tœ se dœ zespes	10	ẽ ʃa selwœʁ dẽ ni pœʁʃe syʁ ẽ nœʁbœʁ	12
10	vivɑ̃ ɑ̃ nœʁmoni o kœʁ dẽ pœʁk natyʁel	13	il ave dy fœʁ fɥiʁ lwazo apœʁ lavwœʁ atake	15
11	kœgyljœʁmɑ̃ syʁ lə ku də midi	10		
12	apœʁ zavwœʁ pœi ẽ bɔ̃ te nu sœʁtɔ̃ də	12		
13	ʃe zel puʁ ale œbœʁve se zanimo	12		

TABLE 1 : Transcription API et découpage en chunks des deux textes sur lesquels les descripteurs sont calculés (Nsyl : nombre de syllabes attendues sur lequel le débit est calculé)

Sept des neuf locuteurs ont piloté leurs enregistrements à l'aide d'une application dédiée développée en Python (une version modifiée de MonPaGe, Trouville et al. 2020) qui gérait la présentation des instructions, de l'enregistrement et de l'organisation des fichiers sonores. Deux locuteurs n'ont pas pu utiliser cette application et ont réalisé leurs enregistrements avec Praat, en s'appuyant sur une

version Powerpoint des instructions. Parmi les 7 tâches du protocole PATAFreq, nous n'étudions ici que la tâche de lecture de deux textes (texte1 et 2) qui ont été manuellement segmentés en chunks selon le découpage illustré dans la table 1, de façon à obtenir environ dix syllabes par chunks. En complément de la segmentation en chunks, l'ensemble des pauses silencieuses perçues ont été délimitées, les pauses se situant entre deux chunks étant par convention rattachées au premier.

2.2 Extraction des descripteurs acoustiques

Les 11 mesures acoustiques récapitulées dans la table 2 ont été extraites au moyen d'un script Praat sur chacun des chunks préalablement segmentés. A l'exception de la mesure de variabilité intra-chunk du rapport harmoniques/bruit (HNR_stdev), ces mesures ont également été dérivées pour obtenir une mesure de modulation entre chunks consécutifs, désignée par le préfixe « d_ » (par exemple, la variable d_debit_arti correspond à la modulation chunk-à-chunk du débit articulatoire), et normalisée comme suit : $|chunk_i - chunk_{i-1}| / ((chunk_i + chunk_{i-1}) / 2)$. Nous obtenons donc un total de 21 descripteurs sur chacun des chunks, à l'exception du dernier chunk de chaque texte sur lequel les mesures de modulation ne sont pas définies.

Dimension	Nom	Description
Temporelle	debit_arti	Débit sans pauses en phones/seconde
	tps_pause	Temps cumulé des pauses silencieuses
Hauteur mélodique	F0moy	Niveau moyen de F0 en Hz
	F0plage	Étendue de F0 en demi-tons recentrée sur les quantiles à 10% et 90% de l'ensemble des valeurs mesurées sur le chunk
Harmonicité	HNR_mean	Valeur moyenne du rapport harmoniques/bruit en dB
	HNR_stdev	Variation intra-chunk du rapport harmoniques/bruit en dB
Spectre / Voix	LTASslope	Pente du spectre moyen à long terme calculé sur les parties voisées du chunk, estimée comme la différence d'intensité en dB entre les bandes 0-1kHz et 1-4kHz
Spectre / Articulation	Int_stdev0_1kHz Int_stdev1_2.5kHz Int_stdev2.5_4kHz Int_stdev4_8kHz	Fluctuation locale de l'intensité trame à trame (au sein d'un chunk) sur 4 bandes de fréquences : 0-1kHz, 1-2.5kHz, et 2,5-4kHz ciblant des caractéristiques liées à l'articulation des segments ; et sur 4-8kHz : où les caractéristiques segmentales sont moins prégnantes.

TABLE 2 : Définition des 11 mesures sur chaque chunk, avant calcul de la modulation chunk-à-chunk (d_Nom) sur 10 d'entre-elles pour obtenir les 21 descripteurs retenus.

2.3 Analyses

L'importance relative des 21 mesures sélectionnées pour la distinction entre sessions produites par un même locuteur a été évaluée au moyen d'une procédure de sélection de variables fondée sur une classification par *random forest* et implémentée dans le package R Boruta (Kursa & Rudnicki, 2010). Dans cette procédure appliquée séparément aux données de chacun des locuteurs, l'évaluation de l'importance de chaque variable repose sur l'estimation de la perte de précision dans la tâche de classification induite par la permutation aléatoire de valeurs entre éléments à classifier (ici les sessions), chaque variable obtenant un score d'importance permettant de les hiérarchiser et éventuellement d'éliminer les variables ne résultant pas en une classification meilleure que la référence aléatoire. Afin de permettre une comparaison directe entre tâches de classification portant sur un

nombre de classes et d'observations non-homogène, nous retenons comme métrique le rang de chaque variable après les avoir classées par score d'importance. La même procédure a été appliquée à l'ensemble des données pour une classification en locuteurs et en texte.

3 Résultats

La table 3 présente les descripteurs ordonnés par rang en fonction de leur importance pour la séparation en sessions pour chaque locuteur, et pour la séparation en locuteurs et en textes. Les distributions des descripteurs discutés plus bas sont illustrées sur la figure 1, en fonction des sessions (courbes de densité juxtaposées) pour chaque locuteur et en fonction des textes (courbes bleue et rouge). Parmi les descripteurs qui ressortent parmi les 5 variables estimées les plus importantes par la classification Boruta pour la séparation en sessions (en haut de la table 3 et de la figure 1), on trouve : F0moy pour tous les locuteurs sauf F05 (pour qui elle apparaît au rang 7), Int_stdev_0-1KHz pour tous sauf F02 (au rang 7), LTASslope pour tous sauf F04 et H02 (au rang 6 et 7 respectivement) et HNR_mean pour tous sauf F05 (au rang 7) et H01 (au rang 12). Tous ces descripteurs qui caractérisent des aspects en basse fréquence, supposés liés à la voix, sont donc assez instables d'une session à l'autre et/ou au sein d'une même session (i.e. entre les chunks). La comparaison du poids de ces descripteurs avec ceux estimés les plus importants pour la classification en locuteurs et en textes (colonnes de droites sur la Table 3) est intéressante : les variables qui ont été relevées parmi les plus importantes pour la séparation en session sont aussi celles qui différencient le plus nos 9 locuteurs. En effet, dans le top-5 des descripteurs utiles pour la séparation des locuteurs on retrouve F0moy, LTASslope et Int_stdev0-1Khz. Ces descripteurs distinguent donc les sessions d'un même locuteur mais aussi les locuteurs entre eux. En effet, si on prend l'exemple de la locutrice F02 sur la figure 1, on voit qu'en fonction de sa session, sa f0 moyenne se rapproche plus de celle de la locutrice F05 (sessions 3-6-9-10), alors que sur les autres sessions elle est plus proche de celle de F03. Sur la sélection de descripteurs présentés dans la figure 1, il apparaît également clairement que certains locuteurs sont assez variables d'une session à l'autre alors que d'autres sont remarquablement stables, comme F03 et H01.

Comme cela est également illustré sur la figure 1, la stabilité/variabilité du débit articulatoire entre les sessions dépend du locuteur : ce descripteur se classe dans le top-10 des variables pour la séparation en session pour F01, F03, F04, F05 et H01, alors qu'il se situe à un rang bien inférieur pour F02, F07, H02 et H03. Parmi ces 4 locuteurs, les patterns sont différents : pour F02 et H02 le débit est particulièrement stable d'une session à l'autre et montre également une plage de variation restreinte au sein de la même session (entre chunks). F07 et H03, au contraire, varient beaucoup leur débit au sein d'une même session, entre les sessions et entre les textes. Ces caractéristiques individuelles de débit articulatoire expliquent que ce descripteur atteigne le top 5 de la séparation en locuteurs sur la cohorte. Le débit articulatoire apparaît aussi varier en fonction des textes. Les fluctuations fréquentielles dans la bande 4-8KHz (Int_stdev4-8KHz), également illustrés dans la figure 1, sont comme on l'attendait parmi les 5 descripteurs les plus importants pour la séparation en locuteurs. Toutefois, on remarque également que celles-ci sont aussi variables entre sessions et présentent une plage de variation importante au sein des sessions : pour tous les locuteurs sauf F03, ce descripteur apparaît donc dans le top-10 de la séparation en sessions. Sa variation en fonction du texte, nous laisse penser que dans cette bande de fréquence, on capture plus de variations liées au contenu segmental qu'attendu. Au niveau des mesures qui n'ont que peu de poids dans la séparation entre sessions d'un même locuteur (i.e. en bas de la liste Table3 et en de la figure 1), on retrouve les mesures de modulation des descripteurs chunk-à-chunk au long du texte (d_X). Parmi celles-ci la modulation chunk-à-chunk de la fluctuation d'énergie dans les moyennes fréquences (d_Int_stdev_1-2.5KHz ou 2.5-4KHz en fonction des locuteurs) se retrouvent parmi les 5 mesures

les plus stables entre les sessions pour tous les locuteurs. Si on considère que sur ces bandes de fréquences, on capture des variations spectrales liées à la dynamique de l'articulation des sons, ce résultat reflète la stabilité de cette dynamique articulatoire chunk-à-chunk. La modulation chunk-à-chunk en hautes fréquences (4-8KHz) apparaît aussi parmi les descripteurs les plus stables pour 7 des 9 locuteurs. La modulation du HNR (d_HNR_mean , non illustrée ici) est aussi stable chez tous les locuteurs sauf F05 et F07, et pourrait refléter chez ces locuteurs une variabilité plus importante de qualité de voix au long de la lecture, qui reste à vérifier avec des mesures plus fines prenant en compte l'évolution au cours du temps dans une même session.

4 Conclusion

Alors qu'elle est attestée depuis fort longtemps, la variabilité de la parole chez un même individu est rarement étudiée. Contrairement à d'autres études menées soit sur un large nombre de locuteurs mais sur un faible nombre de répétition (par ex. Keating et al. 2019 : 200 locuteurs, 12 tâches de parole, mais seulement 3 sessions d'enregistrement, analysé dans Lee et al. 2019), soit sur un faible nombre de locuteurs sur multiples répétitions de séquences isolées (par ex. Whalen & Chen 2019 = 1 locuteur, 300 répétitions de mots monosyllabique ; Tielsen 2017 : 6 locuteurs, 400 répétitions de la séquence /ip^ha/), nous avons pris le parti d'observer la variabilité intra-locuteur dans une cohorte restreinte de 9 locuteurs sur un nombre de sessions relativement importantes, et sur un contenu de parole lue permettant la caractérisation des productions sur plusieurs dimensions à l'aide de 21 descripteurs acoustiques. Pour faire sens, cette variabilité a été comparé à celle qu'on peut observer entre deux lectures de matériaux linguistiques différents ou entre les locuteurs de la cohorte. La similarité observée entre les descripteurs montrant le plus de variation entre les sessions et ceux qui varient le plus d'un locuteur à l'autre interroge sérieusement sur l'influence de la session d'enregistrement choisie dans des tâches de séparation en locuteurs. D'autre part, le fait que les mesures de modulation qui capturent la variabilité interne à une même session entre les différentes parties du texte (la modulation chunk-à-chunk) apparaissent comme les plus stables entre les sessions pour la plupart des locuteurs soulève des pistes intéressantes pour l'introduction de mesures définies sur des empan de parole assez long. De telles mesures pourraient venir compléter les paramétrisations traditionnellement considérées dans de nombreuses tâches de classification automatique de la parole ainsi qu'en clinique, plus locales car définies au niveau de trames de quelques dizaines de millisecondes au maximum et dans lesquelles la prise en compte de la dynamique se limite aux trames voisines à travers les dérivées première et seconde.

Enfin, une autre piste pour nos recherches futures est liée au comportement des descripteurs capturant la dynamique articulatoire dans des bandes de fréquence (mesures Int_stdev_bande et leur modulation). Leur stabilité d'une session à l'autre et au sein d'une même lecture suggère l'intérêt d'approfondir ces caractéristiques de la « façon » d'articuler et co-articuler qui semble propre au locuteur. Nos prochaines études sur ce corpus investigueront donc la variabilité de mesures de coordination temporelle entre sons produits successivement au travers de mesures de coarticulation et d'undershoot des cibles acoustiques, en parole lue mais aussi sur les enregistrements semi-spontanés que nous avons également enregistrés pour ces mêmes locuteurs.

Remerciements

Ce travail a été soutenu en partie par le programme "Investissements d'Avenir" ANR-10-LABX-0083 (Labex EFL) et le projet VoxCrim (ANR-17-CE39-0016).

rangSession										rangLoc	rangTexte
F01	F02	F03	F04	F05	F07	H01	H02	H03			
1	LTASslope	F0moy	Int_stdev0_1KHz	Int_stdev0_1KHz	Int_stdev0_1KHz	Int_stdev0_1KHz	LTASslope	Int_stdev0_1KHz	LTASslope	F0moy	d_HNR_mean
2	F0moy	HNR_mean	F0moy	F0moy	LTASslope	F0moy	F0moy	HNR_mean	Int_stdev0_1KHz	LTASslope	d_LTASslope
3	F0plage	LTASslope	F0plage	Int_stdev1_2_5KHz	Int_stdev1_2_5KHz	HNR_stdev	Int_stdev0_1KHz	F0moy	HNR_mean	Int_stdev4_8KHz	d_tps_pause
4	HNR_mean	Int_stdev1_2_5KHz	HNR_mean	Int_stdev2_5_4KHz	Int_stdev4_8KHz	HNR_mean	HNR_stdev	Int_stdev1_2_5KHz	F0moy	Int_stdev0_1KHz	d_Int_stdev2_5_4KHz
5	Int_stdev0_1KHz	HNR_stdev	LTASslope	HNR_stdev	Int_stdev2_5_4KHz	LTASslope	debit_arti	HNR_stdev	Int_stdev1_2_5KHz	debit_arti	d_Int_stdev4_8KHz
6	Int_stdev1_2_5KHz	Int_stdev4_8KHz	debit_arti	HNR_mean	HNR_stdev	Int_stdev1_2_5KHz	Int_stdev1_2_5KHz	LTASslope	Int_stdev2_5_4KHz	HNR_mean	Int_stdev2_5_4KHz
7	Int_stdev2_5_4KHz	Int_stdev0_1KHz	d_debit_arti	LTASslope	F0moy	F0plage	F0plage	d_Int_stdev0_1KHz	HNR_stdev	F0plage	tps_pause
8	debit_arti	Int_stdev2_5_4KHz	HNR_stdev	Int_stdev4_8KHz	debit_arti	Int_stdev4_8KHz	Int_stdev2_5_4KHz	d_F0moy	d_Int_stdev0_1KHz	HNR_stdev	Int_stdev4_8KHz
9	Int_stdev4_8KHz	F0plage	tps_pause	debit_arti	d_F0plage	d_F0plage	d_Int_stdev2_5_4KHz	Int_stdev4_8KHz	Int_stdev4_8KHz	Int_stdev1_2_5KHz	d_debit_arti
10	HNR_stdev	d_F0moy	d_LTASslope	F0plage	d_LTASslope	Int_stdev2_5_4KHz	Int_stdev4_8KHz	d_LTASslope	F0plage	Int_stdev2_5_4KHz	debit_arti
11	d_debit_arti	d_Int_stdev0_1KHz	Int_stdev1_2_5KHz	d_Int_stdev0_1KHz	d_Int_stdev2_5_4KHz	d_debit_arti	d_debit_arti	F0plage	d_LTASslope	tps_pause	Int_stdev0_1KHz
12	d_F0plage	d_tps_pause	d_Int_stdev2_5_4KHz	d_Int_stdev1_2_5KHz	d_HNR_mean	debit_arti	HNR_mean	d_Int_stdev1_2_5KHz	d_F0moy	d_Int_stdev4_8KHz	d_Int_stdev0_1KHz
13	d_F0moy	d_LTASslope	d_F0plage	d_debit_arti	d_F0moy	d_F0moy	d_Int_stdev0_1KHz	tps_pause	d_debit_arti	d_Int_stdev0_1KHz	d_Int_stdev1_2_5KHz
14	d_Int_stdev0_1KHz	tps_pause	d_F0moy	tps_pause	d_tps_pause	d_Int_stdev0_1KHz	d_F0plage	d_F0plage	d_tps_pause	d_LTASslope	d_F0plage
15	d_tps_pause	debit_arti	d_Int_stdev4_8KHz	d_Int_stdev4_8KHz	HNR_mean	d_LTASslope	d_F0moy	d_tps_pause	debit_arti	d_Int_stdev1_2_5KHz	HNR_mean
16	tps_pause	d_Int_stdev2_5_4KHz	d_Int_stdev0_1KHz	d_LTASslope	F0plage	d_HNR_mean	d_LTASslope	Int_stdev2_5_4KHz	d_Int_stdev2_5_4KHz	d_tps_pause	F0plage
17	d_Int_stdev2_5_4KHz	d_HNR_mean	d_HNR_mean	d_Int_stdev2_5_4KHz	d_Int_stdev1_2_5KHz	d_Int_stdev2_5_4KHz	d_Int_stdev1_2_5KHz	d_Int_stdev2_5_4KHz	d_Int_stdev1_2_5KHz	d_F0plage	Int_stdev1_2_5KHz
18	d_HNR_mean	d_F0plage	d_Int_stdev1_2_5KHz	d_F0plage	d_debit_arti	d_tps_pause	d_HNR_mean	d_HNR_mean	tps_pause	d_F0moy	LTASslope
19	d_LTASslope	d_Int_stdev4_8KHz	Int_stdev4_8KHz	d_tps_pause	d_Int_stdev4_8KHz	d_Int_stdev4_8KHz	d_Int_stdev4_8KHz	debit_arti	d_F0plage	d_Int_stdev2_5_4KHz	d_F0moy
20	d_Int_stdev4_8KHz	d_debit_arti	d_tps_pause	d_F0moy	tps_pause	d_Int_stdev1_2_5KHz	tps_pause	d_Int_stdev4_8KHz	d_HNR_mean	d_debit_arti	F0moy
21	d_Int_stdev1_2_5KHz	d_Int_stdev1_2_5KHz	Int_stdev2_5_4KHz	d_HNR_mean	d_Int_stdev0_1KHz	tps_pause	d_tps_pause	d_debit_arti	d_Int_stdev4_8KHz	d_HNR_mean	HNR_stdev

TABLE 3 : Classement des descripteurs pour la séparation en sessions par locuteurs (rangSession), et pour la séparation en locuteurs (rangLoc) et en textes (rangTexte). En gras noir (meilleure séparation) et rouge (moins bonne séparation), les descripteurs discutés dans le texte. Les couleurs des cases du tableau correspondent au rang moyen sur l'ensemble des 9 locuteurs pour la séparation entre sessions : bleu = meilleure séparation entre sessions ; jaune = rang intermédiaire ; rouge = moins bonne séparation.

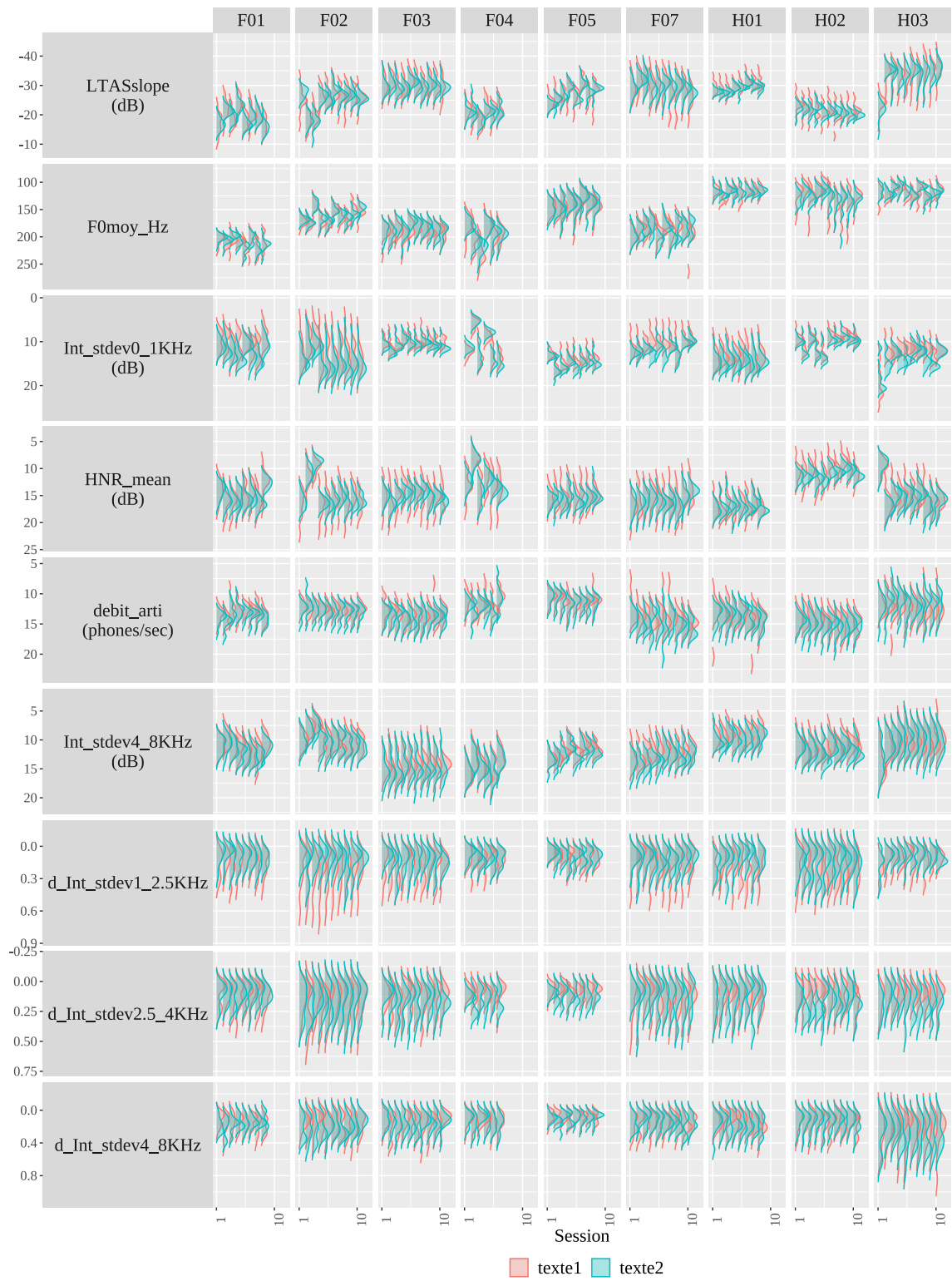


FIGURE 1 : Distribution de 9 des descripteurs pour chaque session (1 à 10), chacun des deux textes (rouge et vert) et les 9 locuteurs.

Références

- BARRETT J., & PAUS T. (2002). Affect-induced changes in speech production. *Experimental Brain Research*, 146(4), 531–537. PMID: 12355282
- BÜRKI A. (2018) Variation in the speech signal as a window into the cognitive architecture of language production. *Psychonomic Bulletin & Review* 25:1973–2004 <https://doi.org/10.3758/s13423-017-1423-4>
- FOUGERON, C., DELVAUX, V., MENARD, L., & LAGANARO, M. (2018). The MonPaGe_HA database for the documentation of spoken French throughout adulthood. *Proceedings of LREC 2018*, Miyazaki, Japan, pp. 4301-4306.
- GELFER M.P. (1991). Effects of prolonged loud reading on selected measures of vocal function in trained and untrained singers. *Journal of Voice*, 5, 15–167.
- HEALD SLM, NUSBAUM HC (2015) Variability in Vowel Production within and between Days. *PLoS ONE* 10(9): e0136791. doi:10.1371/journal.pone.0136791
- JOHNSTONE, T., & SCHERER, K. R. (2000). Vocal communication of emotion. *Handbook of emotions*, 2, 220-235.
- KEATING, P. A., KREIMAN, J., and ALWAN, A. (2019). “A new speech database for within- and between-speaker variability,” in *Proceedings of the ICPHS XIX*.
- LEE Y., KEATING, P., KREIMAN J. (2019). Acoustic voice variation within and between speakers. *Journal of the Acoustical Society of America*. 146, 1568, <https://doi.org/10.1121/1.5125134>
- SCARBOROUGH, R., & ZELLOU, G. (2013). Clarity in communication: “Clear” speech authenticity and lexical neighborhood density effects in speech production and perception. *JASA*, 134(5), 3793-3807
- TIELSEN, S. (2017). Exertive modulation of speech and articulatory phasing. *Journal of Phonetics*, <http://dx.doi.org/10.1016/j.wocn.2017.03.001>
- TROUVILLE, R., DELVAUX, V., FOUGERON, C., LAGANARO, M., (2021). Logiciel d'évaluation de la parole (version screening) MonPaGe-2.0.s [Computer program] Retrieved on 05/01/2022 from <https://lpp.in2p3.fr/monpage/>
- WOISARD, V., BODIN, S., & PUECH, M. (2004). The Voice Handicap Index: impact of the translation in French on the validation. *Revue de laryngologie-otologie-rhinologie*, 125(5), 307-312.
- WHALEN, D. H., CHEN, W. R., TIEDE, M. K., and NAM, H. (2018). Variability of articulator positions and formants across nine English vowels. *J. Phonetics* 68,1–14. doi: 10.1016/j.wocn.2018.01.003