

Dis-moi comment tu varies ton débit, je te dirai qui tu es

Estelle Chardenon, Cécile Fougeron, Nicolas Audibert, Cédric Gendrot
Laboratoire de Phonétique et Phonologie - UMR 7018
Université Sorbonne Nouvelle - CNRS
19 Rue des Bernardins 75005 PARIS, France
(estelle.chardenon, cecile.fougeron, nicolas.audibert,
cedric.gendrot)@sorbonne-nouvelle.fr

RÉSUMÉ

Si l'étude de la variabilité entre locuteurs permet d'identifier des caractéristiques phonétiques potentiellement discriminantes, voire spécifiques, il est essentiel de comprendre, si et comment, ces caractéristiques varient chez un même locuteur. Ici, nous examinons la variabilité de caractéristiques liées à la gestion temporelle de la parole sur un nombre limité de locuteurs, enregistrés sur plusieurs répétitions dans une même session, et sur 6 à 7 sessions espacées d'une année. Sur cette vingtaine d'enregistrements par locuteur, nous observons comment le débit articulatoire, les modulations de ce débit, et la durée des pauses varient en fonction de la répétition et de la session et en interaction avec le locuteur. Les résultats montrent que c'est dans la variation de gestion temporelle de la parole que les locuteurs se distinguent les uns des autres, en termes de régularité ou non entre enregistrements et au sein d'un même enregistrement.

ABSTRACT

Tell me how you vary your speech flow, I'll tell you who you are

Studying inter-speaker variability makes it possible to identify discriminating or even identifying phonetic characteristics. It is however essential to understand if and how these characteristics vary within the same speaker. Here, we examine the variability of characteristics related to temporal organization of speech over a limited number of speakers, recorded over several repetitions in the same session, and over 6 to 7 sessions spaced one year apart. With this 20 recordings per speaker, we observe how the articulatory rate, the modulations of this rate, and the duration of pauses vary according to the repetitions and the sessions and interact with the speaker. The results show that it is in the variation of the temporal organization of speech that speakers differ from each other, in terms of regularity or not between recordings, and within the same recording.

MOTS-CLÉS: débit articulatoire, pause, variation intra et inter locuteur

KEYWORDS: articulatory rate, pause, inter-speaker variability, intra-speaker variability

1 Introduction

Le débit d'un locuteur véhicule des informations spécifiques au locuteur. Par exemple, on sait que les hommes ont tendance à parler plus vite que les femmes même si l'écart de débit entre sexes n'est en général pas très élevé (Byrd, 1994 ; Jacewicz, 2009 ; Schwab, 2012). Des différences de variété régionale peuvent aussi être liées à des différences de débit. Ceci a été montré entre des variétés de l'anglais américain (Jacewicz, 2009), du néerlandais (Quené, 2007), mais aussi sur le français. Cependant, la comparaison entre certaines variétés du français (Suisse, Belge, Français), ne montre pas cette même tendance puisque les études se contredisent voire présentent des variabilités nulles (Schwab, 2015). Le débit articulatoire peut aussi différencier un locuteur natif d'un locuteur L2, même si celui-ci est très fluent (Schwab, 2012, Bordal et al. 2012 ; Barquero, 2012).

Si l'étude de la variabilité entre locuteurs permet d'identifier des caractéristiques de parole potentiellement discriminantes, voire identifiantes, il est essentiel de savoir comment ces caractéristiques varient pour un même locuteur pour qu'elles prennent sens (Kahn et al., 2010, Ajili et al, 2018). Par exemple, Campbell et al. (2009) montrent que les performances d'un système de reconnaissance automatique du locuteur sont fortement impactées lorsqu'il est testé sur deux enregistrements d'un même locuteur à plus d'un mois d'écart. Une des explications possibles à cette variation dans les performances serait liée à l'évolution naturelle de la parole en fonction de l'âge du locuteur. En effet, on sait que le vieillissement a des effets sur la parole, et la diminution du débit et la gestion des pauses chez les locuteurs âgés en est un des effets les plus connus (Ramig, 1983, Jacewicz, 2009, Dellwo, 2015). Pour autant, on ne sait pas sur quel laps de temps on peut observer ces effets. De plus, d'autres facteurs peuvent faire varier la façon dont un locuteur va organiser sa parole entre deux enregistrements. Cette variation peut être contrôlée et/ou volontaire, par exemple pour s'adapter à la situation de communication, au style de parole ou à la personne à qui le locuteur s'adresse (ex. personnes malentendantes ou non natives). On peut ainsi observer des différences de débit entre parole lue et parole spontanée, bien que les différences n'aillent pas toujours dans le même sens (Tsao et Weismer, 2006, Schwab, 2012, Dellwo, 2015, Crystal et House, 1982 ; Avanzi, Schwab, Bubosson et Goldman, 2012, Grosjean et Deschamps, 1975 ; Lucci, 1983). Une variation dans la gestion temporelle de la parole et des pauses peut aussi refléter des changements non contrôlés liés à son état général, par exemple sa consommation de tabac, une prise de médicament, le stress ou la fatigue. En outre, les personnes utilisant quotidiennement leur voix (acteurs, professeurs, chanteurs) sont les premières touchées par le syndrome de fatigue vocale alors qu'il existerait pourtant une meilleure stabilité vocale et articulatoire chez les personnes pratiquant à haut niveau une de ces activités (Lortie, 2016). Enfin, une habitude à la tâche lors d'une expérience avec plusieurs répétitions, ou une fatigue lors d'une longue prise de parole peut aussi survenir au sein de la même session d'enregistrement et impacter la vitesse de production d'un locuteur.

Si ces multiples facteurs pouvant impacter la parole ont été examinés ou mentionnés comme des prédicteurs potentiels de la variation intra et inter-locuteurs, ceux-ci ont rarement été comparés directement et évalués. Dans cette étude, notre objectif est d'examiner la variabilité de paramètres liés à la gestion temporelle de la parole entre un nombre limité de locuteurs, enregistrés plusieurs fois dans une même session, et sur 6 à 7 sessions espacées d'une année.

2 Méthode

2.1 Enregistrements et description de la population

Les enregistrements de 5 locuteurs de la base de données PATATRA ('Parole Adulte A TRavers les Ages'), en cours d'acquisition au LPP, ont été utilisés. Cette base est constituée d'enregistrements d'une dizaine de locuteurs, collectés une fois par an depuis 2013, sur plusieurs tâches de parole. Les enregistrements audio (et EGG) sont effectués en chambre sourde avec le même matériel, et un questionnaire de qualité de vie, axé sur les potentiels facteurs de variation de la voix/parole, est recueilli à chaque session.

Cinq locutrices francophones natives ayant entre 6 et 7 sessions d'enregistrement ont été sélectionnées. Les informations concernant ces locutrices sont résumées dans la table 1. Lors du premier enregistrement, les locutrices étaient âgées de 39 à 68 ans, les locutrices F05 et F04 étant les plus âgées et les trois autres étant dans la quarantaine. Elles présentent toutes un niveau d'étude similaire et vivent en région parisienne.

	F01	F02	F03	F04	F05
Age <2013-2018>	<39-45>	<42-48>	<42-48>	<68-73>	<57-63>
Résidence	IDF	IDF	IDF	IDF	IDF
Nb sessions (manque)	7	6 (2015)	7	6 (2013)	7
Pratique vocale	Oui	Oui	Oui	Oui	Oui
Fatigue vocale	Rare	Modérée	Modérée	Modérée	Modérée

TABLE 1 : Description des 5 locutrices (F01-F05) et des sessions d'enregistrement

2.2 Traitement des données

Notre étude se base sur la tâche de lecture du texte « la bise et le soleil » que le locuteur doit lire à 3 reprises avec 5 minutes de pause entre deux lectures. Si une erreur de lecture est commise, le locuteur est invité à recommencer au début du paragraphe lu. Pour chaque locutrice, le matériel analysé inclut 3 répétitions sur 6 ou 7 sessions, donc 18 à 21 enregistrements. De façon à pouvoir aussi évaluer la variabilité au sein d'une même répétition, le texte a été divisé en 18 *chunks* comprenant chacun une dizaine de syllabe, les pauses internes éventuelles et la pause finale éventuelle (signalées par un dièse).

Chunk	Transcription API	Nombre de phonèmes	Nombres de syllabes
1	/la biz e lə solej sə dispytɛ #/	22	10
2	/ʃakɛ asyɔ̃ kil etɛ lə ply fɔ̃ #/	23	11
3	/kã ilz ɔ̃ vy ɛ vwajazɔ̃ ki/	19	9
4	/savãse # ãvəlope dã sɔ̃ mãto #/	21	10 ou 11
5	/il sɔ̃ tɔ̃be dakɔ̃ kə səlɥi/	20	9
6	/ki akɪvɔ̃kɛ lə prɔ̃mjɛ a/	18	9
7	/fɛɔ̃ ote sɔ̃ mãto o vwajazɔ̃ #/	21	10
8	/sɔ̃kɛ kɔ̃gɔ̃kɛde kɔ̃m lə ply fɔ̃ #/	22	9
9	/alɔ̃ la biz sɛ miz a sufle/	20	9
10	/dɔ̃ tut sɛ fɔ̃s me plyz ɛl sufle #/	24	9
11	/ply lə vwajazɔ̃ sɛkɛ sɔ̃ mãto/	23	10
12	/otɔ̃k dɔ̃ lɥi # e a la fɛ̃#/	15	8
13	/la biz a kɔ̃nɔ̃sɛ a lə lɥi fɛɔ̃ ote #/	24	12
14	/alɔ̃ lə solej a komãse a brije #/	24	12
15	/e o bu dɛ momã # lə vwajazɔ̃#/	20	10
16	/kɛʃofɛ # a ote sɔ̃ mãto #/	16	9
17	/ɛ̃si la biz a dy kɔ̃konɛtɔ̃/	19	9
18	/kə lə solej etɛ lə ply fɔ̃ de dɔ̃ #/	24	11

TABLE 2 : Découpage et transcription des chunks

Ce choix de découpage est justifié par le fait que nous voulions obtenir des extraits de parole contenant à peu près le même nombre d'unités de parole pour pouvoir les comparer entre eux et avoir un même découpage pour tous les locuteurs. Cette segmentation a été faite en respectant au mieux le découpage syntaxique, sans couper de mots. Les 18 chunks contiennent 20.8 phonèmes en moyenne ($\sigma=2.7$, $15 \diamond 24$) correspondant à 9.7 syllabes en moyenne ($\sigma =1.14$, $9 \diamond 12$).

La durée des 18 chunks ainsi que celles des pauses, a été mesurée de façon à calculer, pour chaque chunk, un *débit articulatoire* (nombre de phonème sur temps de parole sans pauses), *le nombre de pauses* éventuelles et *la durée de ces pauses*. Une mesure de *modulation du débit articulatoire* au long du texte a été estimée à partir des différences de débit entre chunks adjacents. Ainsi pour chaque paire de chunks adjacents (17 paires) nous avons calculé une différence relative de débit articulatoire entre chunks à l'aide de la formule suivante : $(d_N - d_{N+1})/((d_N + d_{N+1})/2)$ où d est le débit du $N^{\text{ième}}$ chunk.

Pour prédire les variations de débit articulatoire, de durée des pauses et de différence relative entre les chunks, nous avons utilisé des modèles mixtes testant les effets du locuteur, de la répétition, de la session, mais également l'interaction entre les facteurs : locuteurs et répétition, ainsi que locuteur et session en conservant l'identité du chunk comme facteur aléatoire (lmer(VD~ LOCUTEUR + RÉPÉTITION + SESSION + LOCUTEUR: RÉPÉTITION + LOCUTEUR: SESSION + (1| CHUNK))).

Afin de vérifier si les variations de débit observées entre les sessions et répétitions n'étaient pas dues à des différences intrinsèques de débit entre locuteurs, nous avons également fait des analyses sur des débits relatifs.

3 Résultats

	<i>débit articulatoire</i>	<i>durée des pauses</i>	<i>diff_inter_chunk</i>
LOCUTEUR	$\chi^2(4)= 425.29$ $p<0.0001$, $R^2=0.008$	$\chi^2(4)= 81.38$ $p<0.0001$, $R^2=0.007$	$(\chi^2(4)= 33.47$ $p<0.0001$, $R^2=0.006$
RÉPÉTITION	$\chi^2(2)= 8.87$, $p=0.01$, $R^2=0.002$	NS	NS
SESSION	$\chi^2(6)= 57.87$ $p<0.0001$, $R^2=0.01$	$\chi^2(6)= 16.17$ $p=0.01$, $R^2=0.005$	NS
LOCUTEUR: RÉPÉTITION	$\chi^2(8)=32,52$, $p<0.0001$, $R^2=0.05$	$\chi^2(8)=22.73$, $p=0.004$, $R^2=0.03$	NS
LOCUTEUR: SESSION	$\chi^2(22)=167.82$, $p<0.0001$, $R^2=0.04$	$\chi^2(22)=74.46$, $p<0.0001$, $R^2=0.02$	NS

TABLE 3 : Résumé des résultats statistiques

Les résultats sont résumés dans la Table 3 et les figures 1, 2 et 3. Globalement, un effet du LOCUTEUR est observé sur tous les paramètres étudiés mais en interaction avec les facteurs RÉPÉTITION et SESSION pour le débit et la durée des pauses. Pour ces trois paramètres, ces interactions peuvent être expliquées par des profils de variation différents en fonction des locuteurs. Notre étude ne présente que des statistiques descriptives puisque nous n'avons qu'une seule valeur pour chaque enregistrement et locuteur.

En ce qui concerne le *débit articulatoire*, illustré Figure 1a, toutes les locutrices sauf F02 varient leur débit de manière significative entre sessions. Les comparaisons par paires de sessions montrent toutefois que les locutrices F04, F03 et F01 varient plus d'une session à l'autre que la locutrice F05 qui varie essentiellement entre sa première session et les suivantes. D'autre part, les différences de débit entre sessions semblent essentiellement aléatoires et ne suivent pas de tendance particulière qui pourrait refléter un ralentissement d'une année à l'autre ou sur les dernières sessions. Seul le fait que la première session soit globalement plus rapide ressort pour 3 locutrices (F01, F03, F04). La variation entre répétitions (intra session) dépend aussi du locuteur. Seuls les locutrices F03 et F04 varient leur débit de façon significative entre répétitions, avec une première répétition plus lente que les suivantes.

En ce qui concerne la *durée des pauses*, illustrée Figure 1b, les différences entre sessions et entre répétitions dépendent aussi du locuteur. F05 et F03 gardent une durée de pause stable entre les sessions. Les locutrices F04, F01 et F02 varient d'une session à l'autre, F04 étant la moins régulière. Pour ces trois locutrices, la variation entre sessions semble aléatoire sans suivre une tendance particulière au cours du temps. Entre répétitions, il y a moins de variation dans la durée des pauses : seuls deux locuteurs, F01 et F05, montrent des différences significatives, avec des pauses plus courtes sur leurs premières lectures.

La *mesure de modulation du débit*, illustrée Figure 1c, qui capture la variation de débit d'un chunk à l'autre au sein d'une répétition, est particulièrement intéressante car c'est la seule qui distingue les locuteurs indépendamment de la répétition ou la session. Cette mesure permet de distinguer deux groupes : des locutrices qui modulent beaucoup leur débit (les locutrices F01 et F05) et des locutrices qui modulent moins (F02 et F04). La locutrice F03 présente un profil intermédiaire : similaire à F04 mais modulant plus que F02. N'ayant qu'une seule valeur ici, nous n'avons pas de barres d'erreur.

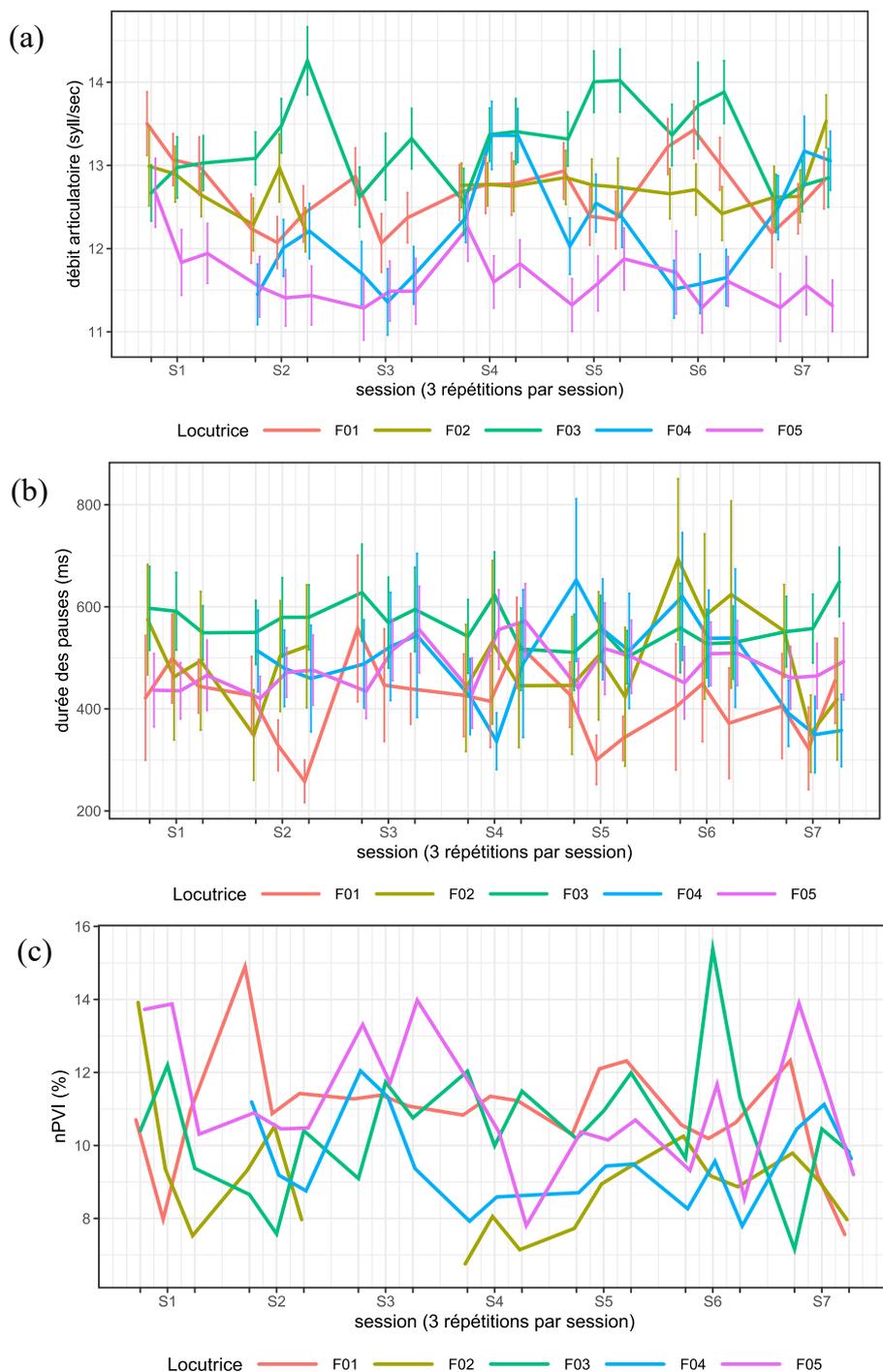


Figure 1 : Évolution (a) du débit articuloire moyen, (b) de la durée moyenne des pauses, (c) de la modulation du débit (différence relative de débit articuloire entre chunks consécutifs) au cours des 6/7 sessions successives et des 3 répétitions par session. Les barres d'erreur représentent la variation entre chunks au sein de chaque enregistrement.

4 Discussion et conclusion

Cette étude basée sur une vingtaine d'enregistrements par locuteur montre que c'est dans la variabilité de leur gestion temporelle (débit, pause, modulation de débit) que les locuteurs se distinguent les uns des autres.

Cette variabilité intra-locuteur s'observe premièrement dans la modulation de débit d'un chunk à l'autre au court d'une même lecture. Certaines locutrices ont un débit qu'on pourrait qualifier de monotone ou isochrone (F02, F04) alors que d'autres lisent le texte en variant beaucoup leur débit. Puisque le débit articulatoire et les pauses dépendent beaucoup du découpage prosodique, nous prévoyons dans les prolongements ultérieurs de cette étude de contrôler dans quelle mesure ces différents profils se retrouvent aussi dans d'autres aspects de l'organisation prosodique du texte produit par les locuteurs dans différents enregistrements.

Deuxièmement, cette variabilité intra-locuteur s'observe entre les différents enregistrements. Certaines locutrices sont très régulières d'une session à l'autre ou d'une répétition à l'autre. C'est le cas de la locutrice F02 qui ne change pas de débit articulatoire, mais aussi de la locutrice F05 qui change peu son débit et a aussi des pauses de durée stable. Au contraire, les locutrices F04, F03 et F01 adoptent des débits plus changeants, accompagnés de durées de pauses variées pour F01 et F04.

Le cas de la locutrice F04 est intéressant. Son débit est très régulier en termes de modulation au court d'une lecture, par contre elle varie beaucoup entre sessions et répétitions. Il est à noter qu'elle est aussi celle qui produit le plus de pauses. Un examen plus poussé de la structuration prosodique du texte lu sera nécessaire pour voir si celle-ci est également plus changeante d'une lecture à l'autre. Cette locutrice est la plus âgée de la cohorte, toutefois son débit n'est pas plus lent sur toutes les répétitions/sessions, ni ses pauses toujours plus longues que celles des autres. C'est donc bien en termes de variabilité que cette locutrice se distingue des autres. Ce résultat est en accord avec les études montrant que s'il y a une tendance à ralentir le débit avec l'âge (5% chez Verhoeven, 2004) ce ralentissement est très variable entre les individus (Decoster, 2000) mais c'est aussi la variabilité intra-sujet qui augmente avec l'âge (Morris et Brown 1994). Toutefois, il reste à démontrer si cette tendance est dû à l'âge ou à l'individu puisque la locutrice F05, également plus âgée que les autres locutrices, présente quant à elle plus de régularité.

Dans ces données, il ne ressort pas d'évolution particulière entre les premières années d'enregistrement et les dernières qui pourrait refléter une évolution longitudinale. Les enregistrements sont réalisés sur une période de 6 ou 7 ans, un laps de temps qui n'apparaît pas suffisant pour montrer un effet d'âge, comparé à d'autres études longitudinales qui observent un changement sur un laps de temps beaucoup plus long (30 ans pour Decoster, 2000 ; 35 pour Harrington, 2007). De plus, l'âge auquel on commence à observer de nombreux changements dans la parole se situe entre 60 et 70 ans, ce qui coïncide avec des changements physiologiques (Decoster, 2000).

Cette étude sera poursuivie sur une plus longue durée en exploitant les enregistrements à venir de la base PATATRA. Il sera également intéressant d'explorer la variabilité intra-locuteur sur d'autres caractéristiques de la voix et la parole. Si les recherches en phonétique ont eu pour tradition dans le passé de moyenniser les caractéristiques de parole entre répétitions, entre sessions, voire entre locuteurs, des résultats comme ceux présentés ici illustrent l'intérêt de mieux comprendre la variation dans la parole ainsi que la façon dont cette variation peut indexer des particularités individuelles.

Remerciements

Ce travail a été partiellement financé par l'ANR VoxCrim (ANR-17-CE39-0016), le programme Investissements d'Avenir (ANR-10-LABX-0083), et le subside FNS CRSII5_173711/1 Sinergia du FNRS.

Références

- AJILI, M., ROSSATO, S., ZHAND, D., & BONASTRE, J. F. (2018). Impact of rhythm on forensic voice comparison reability. HAL : [hal-01962531](#)
- AVANZI, M., SCHWAB, S., DUBOSSON, P., & GOLDMAN, J. P. (2012). La prosodie de quelques variétés de français parlées en Suisse romande. *La variation prosodique régionale en français*. Bruxelles: De Boeck/Duculot, 89-118.
- BARQUERO-ARMESTO, M. Á. (2012). A comparative study on accentual structure between Spanish learners of French interlanguage and French native speakers. *In Speech Prosody 2012*.
- BORDAL, G., AVANZI, M., OBIN, N., & BARDIAUX, A. (2012). Variations in the realization of the French Accentual Phrase in the light of language contact. HAL : [hal-01161304](#)
- BYRD, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, 15(1-2), 39-54.
- CAMPBELL, W., BONASTRE, J. F., SCHWARTZ, R., DRISS, M. (2009). Forensic Speaker Recognition. *Signal Processing Magazine*. IEEE 26.2 : 95-103. DOI: [<10.1109/MSP.2008.931100>](#)
- CRYSTAL, T. H., & HOUSE, A. S. (1982). Segmental durations in connected speech signals: Preliminary results. *The JASA*, 72(3), 705-716.
- DECOSTER, W., & DEBRUYNE, F. (2000). Longitudinal voice changes: facts and interpretation. *Journal of Voice*, 14(2), 184-193. DOI : [<doi.org/10.1016/S0892-1997\(00\)80026-0>](#)
- DELLWO, V., LEEMAN, A., KOLLY, M. J. (2015). Rythmic variability between speakers: Articulatory, prosodic, and linguistic factors. *JASA*, 137(3): 1513-1528. DOI: [<10.1121/1.4906837>](#)
- GROSJEAN, F., & DESCHAMPS, A. (1975). Analyse contrastive des variables temporelles de l'anglais et du français: vitesse de parole et variables composantes, phénomènes d'hésitation. *Phonetica*, 31(3-4), 144-184. DOI : [<doi.org/10.1159/000259667>](#)
- HARRINGTON, J., PALETHORPE, S., & WATSON, C. I. (2007). Age-related changes in fundamental frequency and formants: a longitudinal study of four speakers. In Eighth Annual Conference of the International Speech Communication Association.
- JACEWICZ, E., FOX, R. A., O'NEILL, C., & SALMONS, J. (2009). Articulation rate across dialect, age, and gender. *Language variation and change*, 21(2), 233-256. DOI : [<10.1017/S0954394509990093>](#)
- KAHN, J., AUDIBERT, N., ROSSATO, S., & BONASTRE, J. F. (2010). Intra- speaker variability effects on Speaker Verification performance. *In Odyssey* p. 21.
- LORTIE, C. L., RIVARD, J., THIBEAULT, M., & TREMBLAY, P. (2017). The moderating effect of frequent singing on voice aging. *Journal of Voice*, 31(1), 112-e1.
- LUCCI, V. (1983). Etude phonétique du français contemporain à travers la variation situationnelle (débit, rythme, accent, intonation, e muet, liaisons, phonèmes). *Publications de l'Université des Langues et Lettres de Grenoble*, 1-360.

- MORRIS, R. J., & BROWN Jr, W. S. (1994). Age-related differences in speech variability among women. *Journal of Communication Disorders*, 27(1), 49-64.
- QUENE, H. (2007). On the Just Noticeable Difference for tempo in speech. *Journal of Phonetics*. 35(3), 353– 362. DOI: <[10.1016/j.wocn.2006.09.001](https://doi.org/10.1016/j.wocn.2006.09.001)>
- RAMIG, L. A. (1983). Effects of physiological aging on speaking and reading rates. *Journal of communication disorders*, 16(3), 217-226. DOI : <[doi.org/10.1016/0021-9924\(83\)90035-7](https://doi.org/10.1016/0021-9924(83)90035-7)>
- SCHWAB, S., & AVANZI, M. (2015). Regional variation and articulation rate in French. *Journal of Phonetics*, 48, 96-105. DOI: <[10.1016/j.wocn.2014.10.009](https://doi.org/10.1016/j.wocn.2014.10.009)>
- SCHWAB, S., DUBOSSON, P., & AVANZI, M. (2012). Etude de l'influence de la variété dialectale sur la vitesse d'articulation en français (Dialectal Effect on Articulation Rate in French). *In Proceedings of the Joint Conference JEP-TALN-RECITAL 2012*,1: JEP, 521-528.
- TSAO, Y.C., WEISMER, G., IQBAL, K. (2006). Interspeaker Variation in Habitual Speaking Rate: Additional Evidence. *Journal of Speech, Language and Hearing Research*. 49: 1156-1164.
- VERHOEVEN, J., DE PAUW, G., KLOOTS, H. Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands. *Language and Speech* 2004;47:297–308.