Contraste de voisement en parole chuchotée

Yohann Meynadier, Yulia Gaydina

▶ **To cite this version:**

Yohann Meynadier, Yulia Gaydina. Contraste de voisement en parole chuchotée. XIXè Journées d'Étude sur la Parole, 2012, Grenoble, France. pp.361-368. hal-01211217

HAL Id: hal-01211217
https://hal.science/hal-01211217

Submitted on 4 Oct 2015

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.

Contraste de voisement en parole chuchotée

Yohann Meynadier & Yulia Gaydina
Laboratoire Parole et Langage, CNRS URM7309 & Université d'Aix-Marseille

yohann.meynadier@lpl-aix.fr, yulia.gaydina@lpl-aix.fr

RÉSUMÉ

Ce travail porte sur le contraste phonologique de voisement en parole chuchotée qui se caractérise par une configuration semi-ouverte des cordes vocales empêchant leur vibration. En parole modale, outre la vibration des cordes vocales, le contraste entre consonnes voisées et sourdes est supporté par d'autres corrélats phonétiques : durées des consonnes et des voyelles, pression intraorale, entre autres. Les analyses acoustiques et aérodynamiques des consonnes voisées *vs* sourdes montrent que ces corrélats secondaires du voisement sont préservés en parole chuchotée, pouvant donner une assise à la persistance de la perception de ce contraste malgré l'absence de vibration des cordes.

ABSTRACT

Voicing contrast in whispered speech

This paper presents analyses on the phonological voicing contrast in whispered speech, which is characterized by a semi-open configuration of the vocal folds preventing them from vibrating. In modal speech, in addition to vocal fold vibration, the contrast between voiced and unvoiced consonants is realized by other phonetic correlates: e.g. consonant and pre-consonantal vowel durations, intraoral pressure differences. Acoustic and aerodynamic analyzes show that these voicing correlates are preserved in whispered speech. These findings seem consistent with those showing that voiced contrast is maintained in perception despite the absence of vocal fold vibration.

MOTS-CLÉS : phonétique, voisement, voix chuchotée, aérodynamique, durée segmentale
KEYWORDS: phonetics, voicing, whisper, aerodynamics, segmental duration

1 Introduction

Le chuchotement est un mode de phonation naturellement utilisé dans le but de réduire la perceptibilité de la parole. En phonation modale (voix normale), l'adduction complète des cordes vocales permet leur mise en vibration pour les segments voisés. En voix chuchotée, les cordes vocales sont accolées uniquement dans leur partie antérieure, laissant une ouverture étroite inter-aryténoïdienne pour l'échappement de l'air. Cette configuration glottique permet l'établissement de turbulences aérodynamiques à l'origine de la source acoustique bruitée. Outre l'absence de voisement, la voix chuchotée induit des modifications spectrales importantes comme la perte d'énergie, notamment dans les basses fréquences (Ito et al. 2004, Jovicic & Saric 2008), l'aplatissement du spectre, particulièrement des voyelles et des consonnes voisées, l'élévation des formants vocaliques (Sharifzadeh et al. 2009)… Concernant plus spécifiquement l'absence de voisement en parole chuchotée, un certain nombre d'études rapporte que tant les informations tonales et intonatives même en absence de la f0 (Faraco 1984, Nicholson &

Teig 2003, Vercherand 2010), indexiales (sexe du locuteur), segmentales (timbre vocalique) (Eklund & Traunmüller 1996) que de voisement (Mills 2003, 2009, Vercherand 2010) sont en bonne partie préservées, même si leur perception est plus réduite qu'en parole modale. Notre étude s'intéresse aux traits phonétiques susceptibles de participer au maintien de la perception du trait phonologique de voisement en parole chuchotée, à savoir sans voisement physiologique et acoustique.

En parole modale, outre la vibration périodique des cordes vocales, d'autres indices phonétiques du voisement sont communément observés (Catford 1977, Eklund & Traunmüller 1997, Silbert & de Jong 2008). Par exemple, les voyelles sont plus longues avant une consonne voisée qu'une sourde (Lehiste 1970). La durée des consonnes obstruantes est plus importante pour les sourdes que les sonores. Cet écart de durée consonantique relèverait d'une contrainte aérodynamique liée au différentiel transglottique de pression ($\Delta P = Ps - Po$, où Po = pression intraorale et Ps = pression sous-glottique) nécessaire au maintien de la vibration des cordes vocales durant la production des consonnes voisées (Ohala 1997). En effet, en-deçà de 1 à 2 hPa la vibration s'arrête faute de flux d'air assez puissant. Cette contrainte joue en défaveur d'une durée longue des obstruantes voisées, au contraire des sourdes pour lesquelles le voisement est absent, empêché par une abduction des cordes. D'autre part, cette contrainte aérodynamique se traduit aussi par une Po plus importante pour les obstruantes sourdes par rapport aux voisées. En effet, l'adduction glottique réalisée pour les obstruantes voisées impose une résistance au flux d'air transglottique, réduisant d'autant la quantité d'air s'accumulant dans la cavité supraglottique en amont de la constriction (Malécot 1955). Lors du chuchotement, cette contrainte aérodynamique ne devrait plus jouer du fait de la fuite glottique constante caractéristique de la configuration phonatoire chuchotée. On s'attendrait donc à ce que les écarts de durée et de Po (Wiesmer & Longstreth 1980) soient absents.

Les travaux, rapportés par Weismer & Longstreth (1980), sur les gestes laryngés et leurs conséquences aérodynamiques sont contradictoires s'agissant du contraste voisé-non voisé en parole chuchotée. Par ailleurs, Weismer & Longstreth (1980) ne mettent pas en évidence une différence homogène de Po entre /p/ et /b/ chuchotés. Or, les études de Mills (2003, 2009) sur l'anglais et Vercherand (2010) sur le français montrent qu'en parole chuchotée la différence de durée entre consonnes voisées et non voisées est maintenue. De plus dans une étude récente, Mills (2009) mesure précisément par fibroscopie que l'ouverture glottique est dynamiquement ajustée en fonction de la propriété de voisement des consonnes en parole chuchotée : les obstruantes sourdes montrent une ouverture plus importante que les voisées. Ces résultats posent la question de la raison de l'existence de ces marques corrélatives du voisement en l'absence de vibration glottique. Dans cette étude, nous avons reproduit et complété ces résultats et nous discutons de la production et de la perception du trait phonologique [voisé]. Nous avons menés deux expériences de production en parole chuchotée : l'une sur les indices temporels du voisement et l'autre sur son corrélat aérodynamique.

2 Expérience 1 : durées segmentales

Cette expérience porte sur l'analyse acoustique de la durée des consonnes voisées *vs* sourdes et des voyelles pré-consonantiques en parole modale *vs* chuchotée.

2.1 Corpus

Enregistrés acoustiquement en chambre sourde, 4 locuteurs français (2 hommes, 2 femmes), non linguistes ou étudiants en linguistique, ont lu à haute voix et à débit normal, après une brève session d'entraînement, deux listes de mots présentés isolément. La première liste était composée de 12 logatomes cibles noyés parmi 36 distracteurs. Chaque locuteur a produit cette liste randomisée 5 fois en 10 sessions de 48 items alternant la phonation modale et la phonation chuchotée. Les logatomes cibles étaient de forme VC_1VC_2V, où V est toujours /e/ et C_1 et C_2 une consonne obstruante cible toujours différente parmi les couples voisé-non voisée : /p b/, /t d/, /k g/, /f v/, /s z/, /ʃ ʒ/, par exemple /epeʒe/ (écrit « épéjé ») ou /edese/ (écrit « édécé »). Les distracteurs répondaient au même patron mais comportaient au moins une consonne non cible, par exemple /ekene/ (écrit « équéné ») ou /eleme/ (écrit « élémé »). La liste était équilibrée en fréquence d'occurrence des consonnes. La seconde liste était composée de 24 verbes cibles conjugués et mélangés au hasard avec 72 distracteurs de même gabarit phonologique. Chaque locuteur a répété cette liste en phonation chuchotée et modale de la même manière que la première. Les mêmes consonnes cibles étaient initiales de syllabe médiane ou finale de mot. La voyelle pré-consonantique était /e/ dans tous les cas sauf 4 (/a/). Pour contrôler un minimum la voyelle suivante, les paires de consonnes voisée-non voisée correspondaient à une paire minimale, par exemple « (il) é**c**outait » *vs* « (il) é**g**outtait » ou « il é**c**oute » *vs* « il é**g**outte ». Le sujet ne devait lire que les mots hors parenthèses, afin de produire des séquences toujours trisyllabiques. Les distracteurs variaient de la même façon et ont été équilibrés en fonction du nombre de syllabe du verbe. Le corpus comporte donc 8640 items enregistrés, dont la durée acoustique des voyelles et des consonnes des 1440 items cibles ont été manuellement mesurées.

2.2 Analyses

L'étiquetage du signal acoustique a été réalisé sous Praat (www.fon.hum.uva.nl/praat). Afin de garantir une segmentation identique dans les deux modes de phonation, elle a essentiellement été effectuée à partir du spectrogramme. Le début et la fin des voyelles sont respectivement localisés au début et à la fin de l'énergie des F2-F3. Deux phases consonantiques ont été mesurées : tenue et relâchement (Figure 1). Pour les plosives, la tenue a été étiquetée de la fin de la voyelle précédente jusqu'au début de l'explosion, le relâchement du début de l'explosion au début de la voyelle suivante, incluant le bruit de friction et les traces formantiques de la transition CV. Pour les fricatives, la tenue, présentant un bruit de friction intense de 2000 à 8000 Hz (selon la consonne), a été séparée de la phase de transition CV (ou relâchement) identifiée par l'apparition de traces formantiques de la voyelle suivante et une diminution marquée de l'énergie du bruit de friction. La tenue va de la fin de la voyelle précédente au début des traces formantiques de la voyelle suivante. Son relâchement court de la fin du bruit intense au début de F2-F3 de la voyelle qui suit.

Seules les consonnes en syllabe inaccentuée (position non finale de mot) et les voyelles précédant celles-ci ont été analysées ici. Les analyses statistiques effectuées sont des ANOVA à mesures répétées : les locuteurs et la lexicalité (mot *vs* logatome) sont en facteurs aléatoires ; le mode de phonation (modale *vs* chuchotée), le voisement (voisé *vs* sourd) et le mode d'articulation (plosive *vs* fricative) en facteurs indépendants.

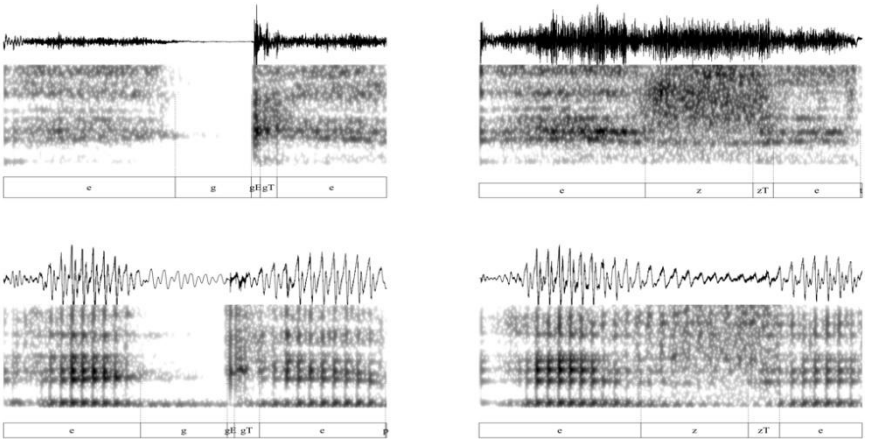

Figure 1 – Étiquetage acoustique des voyelles et des consonnes des mots [**ege**pe] (à gauche) et [**eze**te] (à droite) en voix modale (en bas) et chuchotée (en haut)

2.3 Résultats et discussion

Nous présentons ici les résultats relatifs à l'effet du voisement sur les durées consonantiques et vocaliques en phonation modale et chuchotée pour les fricatives et les occlusives, les données relatives aux mots et aux logatomes sont confondues. La phonation modale constitue la situation de référence à laquelle sont comparées les observations en parole chuchotée. Comme attendu, en phonation modale les consonnes sourdes (118 ms) sont en moyenne plus longues que les voisées (78 ms) [$F(1,3) = 161{,}79$; $p = 0{,}00105$]. Les fricatives et occlusives sourdes sont plus longues que les voisées : respectivement 133 *vs* 81 ms et 102 *vs* 74 ms en moyenne. De même, en phonation chuchotée, les consonnes sourdes sont en moyenne plus longues de 31 ms que les voisées [$F(1,3) = 56{,}014$; $p = 0{,}00494$] : respectivement, 136 ms *vs* 99 ms pour les fricatives et 108 ms *vs* 82 ms pour les occlusives.

Ainsi, on peut observer le maintien d'une différence importante des durées consonantiques associée au contraste de voisement en voix chuchotée, malgré une réduction notable de cet écart pour les fricatives chuchotées (37 *vs* 52 ms en modal). Pour les occlusives, cet écart est constant entre les deux modes phonatoires : 28 ms en modal et 26 ms en chuchoté. On constate par ailleurs que la durée des consonnes augmente significativement en mode de phonation chuchotée (107 ms) par rapport à modale (98 ms) [$F(1,3) = 13{,}881$; $p = 0{,}03368$], ce qui rejoint les autres études et l'observation générale d'une parole chuchotée souvent plus lente qu'en modale (Schwartz 1967, Jovicic & Saric 2008, Mills 2003, 2009, Vercherand 2010).

L'analyse plus détaillée de la durée des phases de tenue et de relâchement des consonnes révèle une différence remarquable. En parole modale, ces deux phases consonantiques participent significativement à la distinction temporelle entre consonnes sourdes et voisées. En parole chuchotée, la tenue des occlusives est significativement plus longue pour les sourdes (84 ms) que les voisées (59 ms) [$F(1,3) = 55{,}134$; $p = 0{,}00505$]. Cette différence de 25 ms est renforcée par rapport à la voix modale (18 ms, soit 72 ms pour les sourdes et 54 ms pour les voisées). La phase de relâchement ne participe pas

significativement au contraste de voisement en chuchoté (24 ms pour les sourdes *vs* 23 ms pour les sonores), au contraire des occlusives produites en phonation modale [$F(1,3) = 13{,}670$; $p = 0{,}03434$]. S'agissant des fricatives, la durée de la tenue est significativement différente en fonction du voisement en voix chuchotée [$F(1,3) = 49{,}841$; $p = 0{,}00584$]. La différence observée est de 36 ms en faveur des fricatives sourdes (126 *vs* 90 ms pour les voisées). On remarque cependant une légère réduction par rapport à la phonation modale, où cet écart compte 41 ms. Comme pour les occlusives, la phase de relâchement des fricatives ne montrent pas en phonation chuchotée de différence temporelle significative selon le voisement (10 ms pour les sourdes *vs* 9 ms pour les sonores), au contraire de leur réalisation en phonation modale. Ainsi, on peut supposer que seule la phase de tenue porte des informations temporelles liées au caractère voisé des consonnes obstruantes en parole chuchotée (Figure 2).

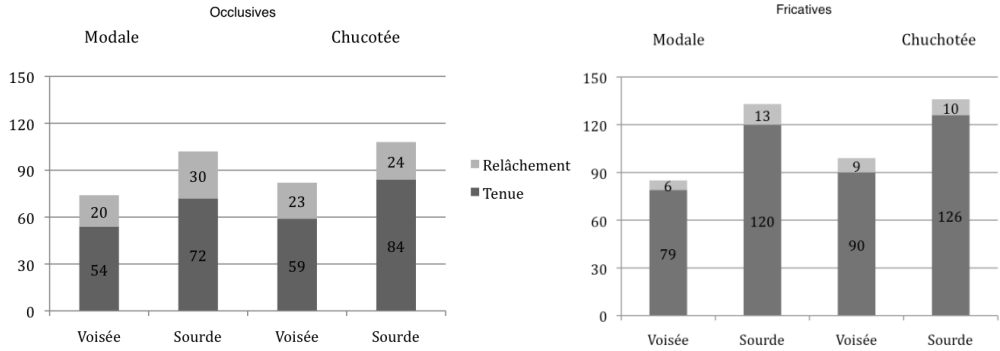

Figure 2 – Durée moyenne (en ms) de la tenue (en bleu) et du relâchement (en rouge) des consonnes selon le voisement et la phonation

Ces résultats sur la durée des consonnes voisées *vs* sourdes, d'une part, sont en accord avec ceux obtenus dans les études précédentes (Mills 2003, 2009, Jovicic & Saric 2008, Vercherand 2010), et d'autre part, complètent ces connaissances par une observation plus détaillée des patrons temporels en jeu dans la production du contraste de voisement en parole chuchotée.

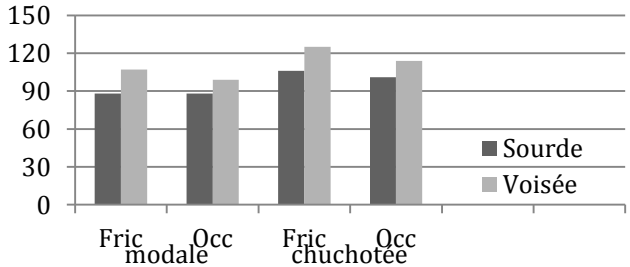

Figure 3 – Durée moyenne (en ms) des voyelles pré-consonantiques en fonction de la nature de la consonne suivante en phonation modale et chuchotée

Outre la durée consonantique, en parole modale le voisement est également marqué par une durée plus importante des voyelles pré-consonantiques devant consonne voisée. Nos données confirment ce résultat attendu : une voyelle avant une consonne sourde est en moyenne plus courte (88 ms) que celle précédant une consonne voisée (103 ms)

[F(1,3) = 93,103 ; p = 0,00236]. Cet écart de 15 ms est similaire à celui observé en parole chuchotée dans notre corpus [F(1,3) = 245,76 ; p = 0,00056], soit 16 ms entre les voyelles avant une consonne sourde et une consonne voisée. Quel que soit le mode phonatoire ou d'articulation, cet écart reste assez faible (Figure 3), tenant potentiellement au fait que la voyelle pré-consonantique observée ici n'est pas tautosyllabique avec la consonne qui suit. L'observation de voyelles suivies d'une consonne codaïque devrait montrer des différences plus nettes selon le voisement, comme dans Mills (2003). Reste que ces différences de durée vocalique pré-consonantique sont préservées en parole chuchotée et pourraient ainsi comme en parole modale participer au marquage phonétique du contraste de voisement.

3 Expérience 2 : pression intraorale

Cette expérience porte sur l'analyse du pic de pression intraorale (Po) atteint lors la constriction des obstruantes sourdes et voisées en fonction du mode de phonation.

3.1 Corpus et analyses

8 locuteurs français (6 femmes, 2 hommes) ont lu à haute voix et à débit normal en chambre sourde une liste randomisée et équilibrée de 72 mots présentés isolément, une fois en voix modale puis une fois en voix chuchotée. Les mots comportaient les obstruantes labiales cibles /p b/ ou /f v/ en initiale ou finale de mots monosyllabiques, composés des voyelles /a ɛ ɔ/ (par exemple, « **v**ache, **b**eige, **p**omme » ou « che**f**, ra**p**, ro**b**e » ou en médiane de mots bisyllabiques, en contexte vocalique /a_a ɛ ɔ / (par exemple, « sa**v**ate, a**ff**aire, ra**pp**ort »). Chaque contexte a été itéré dans 2 mots différents, représentant au total 18 occurrences différentes par consonne cible, soit 1152 items cibles analysés. Le signal acoustique, le débit d'air oral et la Po ont été enregistrés synchroniquement avec l'aérophonomètre d'EVA (Ghio & Teston 2004). La Po a été acquise par voie orale via un cathéter dont l'extrémité dépassait juste l'arrière des incisives. Lors de l'articulation de l'obstruante, ce cathéter pointait dans la cavité orale fermée par l'occlusion ou la constriction labiale ou labiodentale. La prise de mesure manuelle de la Po a été réalisée sous Phonédit (www.lpl-aix.fr/~lpldev/phonedit). Mesuré en hecto Pascal (hPa), le pic de Po a été localisé sur le maximum de la courbe de Po (son point d'inflexion) atteint lors de la constriction consonantique.

Des ANOVA à mesures répétées portent sur le voisement au regard des modes de phonation et d'articulation (facteurs indépendants). Les locuteurs, la position syllabique de la consonne et la nature des voyelles sont en facteur aléatoire.

3.2 Résultats et discussion

Le contraste de voisement est significativement marqué par une Po maximale plus haute en parole modale. Mais surtout, cela est vrai également en parole chuchotée. Les tests montrent que seuls le voisement [F(1,7) = 138,162 ; p = 0,00007] et l'interaction voisement*phonation [F(1,7) = 16,631 ; p = 0,0047] sont significatifs. Dans les deux modes, les obstruantes sourdes (5,42 hPa en modal et 4,91 hPa en chuchoté) montrent une pression plus importante que les voisées (3,56 en modal et 4,02 en chuchoté).

La différence de Po entre sourdes et voisées réduite de moitié en phonation chuchotée (0,89 hPa) par rapport à la phonation modale (1,86 hPa) peut s'expliquer. D'une part, l'intensité sonore, largement soutenue par la pression sous-glottique, est plus faible d'environ 20 dB en parole chuchotée, ce qui peut entraîner une Po plus basse et/ou écraser sa dynamique par un effet plafond et/ou planché. D'autre part, la configuration glottique spécifique au chuchotement propose, par rapport à la configuration modale, une résistance au flux d'air égressif moindre pour les consonnes voisées chuchotées (plus ouvertes du fait de la fuite glottique) et peut-être plus grande pour les sourdes (plus fermées du fait de l'accolement partiel des cordes). Cela aurait pour effet d'abaisser la Po des obstruantes sourdes et d'augmenter celle des voisées en mode chuchoté. En outre, l'absence quasi totale de significativité du facteur phonation sur les valeurs de Po (seules les occlusives voisées montrent un écart de Po significatif [$p = 0{,}029$] selon la phonation : 4,28 hPa en chuchoté *vs* 3,38 en modal) pourrait aller dans le sens d'une Po relativement variable dans les deux modes, mais malgré tout fortement sous-tendue par une nécessité de préserver une différence de configuration glottique et/ou de pression entre consonnes voisées et sourdes, assez robuste dans les deux modes de phonation.

Ainsi, +1 hPa environ distingue les obstruantes sourdes des voisées en parole chuchotée. Ce résultat est contradictoire avec ceux de Weismer & Longstreth (1980) portant sur le contraste /p b/ en anglais. Il supporte l'idée que la configuration glottique adoptée lors du chuchotement n'est pas une position statique et constante, mais que l'ouverture glottique peut être spécifiquement contrôlée pour véhiculer une information relative au voisement de la consonne. Ainsi, nos résultats aérodynamiques sur le français confirment indépendamment ceux obtenus en fibroscopie par Mills (2009) sur l'anglais, qui constitue l'une des rares mesures (et pas seulement une observation) empiriques directes de ce phénomène physiologique.

4 Conclusion

Notre étude soutient que la production du contraste de voisement en parole chuchotée est supportée par des traits phonétiques secondaires produits en parole modale.

La durée intrinsèque des consonnes et la différence de pression intraorale sont en bonne partie la conséquence phonétique de la contrainte aérodynamique liée au différentiel de pression transglottique nécessaire à la vibration des cordes vocales (Malécot 1955). Or, en l'absence de cette contrainte en voix chuchotée, ces différences phonétiques semblent se maintenir, bien qu'affaiblies. Concernant la constriction glottique distincte entre obstruantes voisées et sourdes en voix chuchotée, une analyse acoustique de la qualité du bruit émis à la source glottique et de sa résonance supraglottique est encore à mener pour déterminer si des informations spectrales du voisement existent en voix chuchotée. Dès lors, la question de l'implication et de la hiérarchie de ces indices phonétiques dans la perception du trait [voisé] en phonation chuchotée se pose également. Au regard des études antérieures effectuées en perception, on peut raisonnablement le supposer. Par contre à notre connaissance, aucune étude complète n'a encore tenté de déterminer à quel niveau du traitement perceptif ces indices interviennent.

Enfin, il est à noter que notre étude constitue la première analyse physiologique indirecte sur le français qui confirme empiriquement l'existence de gestes glottiques

contrastifs associés à l'opposition de voisement des consonnes en parole chuchotée.

Références

CATFORD, J.C. (1964). Phonation types: the classification of some laryngeal components of speech production. IN D. Abercrombie, D.B. Fry, P.A.D. MacCarthy, N.C. Scott & J.L.M. Trim (eds.), *In honour of Daniel Jones*, p. 26-37. London: Longmans.

CATFORD, J.C. (1977). *Fundamental problems in phonetics*. Edinburgh University Press.

EKLUND, I. & TRAUNMÜLLER, H. (1997). Comparative study of male and female whispered and phonated versions of long vowels of Swedish. *Phonetica* 54(1): 1-21.

FARACO, M. (1984). *Comparaison des intonations affirmative et interrogative en voix normale et chuchotée.* Thèse de doctorat, Université de Provence, Aix-en-Provence.

GHIO, A. & TESTON, B. (2004). Evaluation of the acoustic and aerodynamic constraints of a pneumotachograph for speech and voice studies. *Proceedings of International Conference on Voice Physiology and Biomechanics*, p. 55-58. Marseille.

ITO, T., TAKEDA, K. & ITAKURA, F. (2005). Analysis and recognition of whispered speech. *Speech Comunication* 45(2): 139-152.

JOVICIC, S.T. & SARIC, Z. (2008). Acoustic analysis of consonants in whispered speech. *Journal of Voice* 22(3): 263-74.

LEHISTE, I. (1970). *Suprasegmentals*. Cambridge: MIT Press.

MALÉCOT, A. (1955). An experimental study of force of articulation. *Studia Ling.* 9: 35-44.

MILLS, T.I.P. (2003). *Cues to voicing contrasts in whispered Scottish obstruents*. Master of Science, University of Edinburgh.

MILLS, T.I.P. (2009). *Speech motor control variables in the production of voicing contrasts and emphatic accent.* Phd dissertation, University of Edinburgh.

NICHOLSON, H. & TEIG, A.H. (2003). How to tell beans from farmers: cues to the perception of pitch accent in whispered Norwegian. *Nordlyd* 31(2): 315-325.

OHALA, J. J. (1997). Aerodynamics of phonology. *Proceedings of the 4th Seoul International Conference on Linguistics*, p. 92-97.

SHARIFZADEH, H.R., MCLOUGHLIN, I.V. & AHAMDI, F. (2009). Voiced speech from whispers for post-laryngectomised patients. *IAENG International Journal of Computer Science* 36.

SCHWARTZ, M.F. (1967). Syllabe duration in oral and whispered reading. *JASA* 41:1367-9.

SILBERT, N. & DE JONG, K. (2008). Focus, prosodic context, and phonological feature specification: Pattern of variation in fricative production. *JASA* 123: 2769-79.

VERCHERAND, G. (2010). *Production et perception de la parole chuchotée en français: analyse segmentale et prosodique*. Thèse de doctorat. Université de Paris 7.

WEISMER, G. & LONGSTRETH, D. (1980). Segmental gestures at laryngeal level in whispered speech: evidence from an aerodynamic study. *Journal of Speech Hear. Res.* 23: 383-92.