

Produire et percevoir la prosodie en voix chuchotée : étude préliminaire sur la modalité interrogative et déclarative en français

Géraldine Vercherand & Hi-Yon Yoo

UPRES EA 33 « Atelier de Recherches sur la Parole », UFR Linguistique
Université Paris 7 – Denis Diderot, 2 place Jussieu, 75251 Paris Cedex 05 France
vercherand@linguist.jussieu.fr ; yoo@linguist.jussieu.fr
http://phon.linguist.jussieu.fr

ABSTRACT

In this paper, we propose to analyze the prosodic realization of two phrasal modalities (declarative and interrogative) through a production and perception study of whispered voice in French. We present the analysis of the acoustic parameters that are at stake when producing whispered voice and the results of an identification test. Results show that the most important acoustic cue in differentiating the two modalities in whisper is intensity, which has a higher value at the end of interrogative sentences. Moreover, the results of the identification test show that the two modalities are globally well perceived in whisper, while the good identification rate is higher in longer sentences.

1. INTRODUCTION

Le chuchotement est un mode de parole qui, même s'il n'a pas de fonction linguistique distinctive, est utilisé par pratiquement toutes les langues pour communiquer dans une situation discrète. La voix chuchotée implique la non vibration des cordes vocales et par conséquent l'absence de la fréquence fondamentale. Or si l'utilisation de ce mode de parole est courante, c'est parce que finalement l'intonation continue à être reconnue, ce qui est assez étonnant si l'on considère que la fréquence fondamentale est le paramètre acoustique le plus important dans la production de l'effet intonatif. Si l'intonation continue à être perçue en voix chuchotée, quels sont les moyens autres produisant l'intonation ? Nous avons cherché de donner des éléments de réponses à travers une étude de production et de perception de la voix chuchotée en français.

La plupart des travaux sur le chuchotement traitent de ses aspects acoustiques et physiologiques (comme par exemple, les caractéristiques glottales [Tartter [10]], les mouvements laryngaux [Bonnot [1]], la durée de la tenue des consonnes [Higashikawa [5] Schwartz [9]] ou encore la hauteur intrinsèque des segments en voix chuchotée [Harbold [4] et Meyer-Eppler [7]]. Le matériel examiné pour ces analyses consiste en des mots ou non mots isolés (voir par exemple les travaux de Thomas [11] sur la perception de la hauteur des voyelles), et l'on ne trouve que très peu d'études sur la réalisation prosodique de la voix chuchotée (voir notamment les quelques travaux sur

la perception des tons en chinois [Gao [3], Jensen [6], et en vietnamien [Miller [8]]). Enfin, en ce qui concerne les études sur la voix chuchotée en français, elles semblent être quasi inexistantes (Fonagy [2] étant l'un des rares chercheurs à avoir travaillé sur ce sujet).

Dans ce travail, nous nous proposons d'analyser la voix chuchotée produite dans un contexte de lecture, en se focalisant notamment sur la réalisation prosodique de deux modalités en français, la déclarative et l'interrogative.

Nous partons des deux hypothèses suivantes : a) les paramètres de durée et d'intensité sont suffisants pour permettre de différencier la modalité phrastique ; b) l'unité rythmique doit être suffisante pour que ces paramètres se réalisent pleinement, ce qui est reflété par un meilleur taux de reconnaissance dans les énoncés plus longs. Nous présenterons d'abord les résultats de l'analyse de production de la voix chuchotée, puis les résultats du test d'identification.

2. PRODUIRE LA VOIX CHUCHOTÉE

2.1. Protocole expérimental de l'étude de production de la voix chuchotée

Le corpus comporte 13 phrases sémantiquement neutres ayant une structure syntaxique simple. Pour les phrases de 6 et 7 syllabes, nous avons rajouté une phrase supplémentaire afin de voir si la constitution syllabique avait un effet sur la production.

Table 1 : Corpus de l'étude de production. Il faut combiner un élément par colonne, le chiffre entre parenthèses indiquant les phrases dans lesquelles cet élément apparaît

		Ø (1) bien (2,3) une noix (4,5b) un melon (5a,6b) des ananas (6a,7)une noix du Ghana (8) un melon du Ghana (9,10) des ananas du Ghana (11)
Elle (1,2) Anna (3,4,5,6) Annabelle (7,8,9) Annabella (10,11)	mange (ensemble du corpus sauf 5b et 6b) mangea (5b, 6b)	

Nous avons fait varier le paramètre du nombre de syllabes, allant de 2 à 12 syllabes ainsi que la modalité de phrase, le

même énoncé pouvant être prononcé comme une déclarative ou interrogative (un total donc de 26 phrases). Nous avons essayé de maintenir dans la mesure du possible une structure CV pour l'ensemble des mots, des voyelles ouvertes et le maximum de consonnes nasales (et bien sûr aucun segment non voisé) afin de minimiser l'influence d'autres facteurs comme la valeur intrinsèque des phones. L'ensemble du corpus a été mélangé de manière aléatoire. Le recueil des réponses a été présenté sous forme de petites fiches et avec une écriture orthographique.

Pour cette étude de production, nous avons demandé à trois locutrices francophones d'énoncer aussi naturellement que possible l'ensemble du corpus, en voix normale et en voix chuchotée. Sur un total de 5 occurrences obtenues par phrase et par locutrice, la première et la dernière occurrence pour chacune des locutrices n'ont pas été prises en compte dans l'analyse acoustique : notre étude porte donc sur l'analyse de 3 occurrences par phrase.

Les enregistrements se sont déroulés le même jour dans une salle calme des locaux de l'UFR Linguistique de l'Université Paris 7. Ils ont été effectués directement sur un ordinateur PC, en utilisant un microphone unidirectionnel. Pour chaque occurrence nous avons d'abord vérifié l'absence de fréquence fondamentale. Ensuite nous avons mesuré la durée des phones et calculé la durée des syllabes ainsi que l'intensité (une valeur par segment).

2.2. Résultats de l'analyse acoustique

Comparaison de la voix chuchotée à la voix normale

Nous avons dans un premier temps comparé les deux modes de production, voix chuchotée et voix normale. Sur l'ensemble des trois locutrices, on retrouve des résultats similaires. Mis à part la présence de la fréquence fondamentale en voix normale (avec un contour descendant pour la modalité déclarative et un contour montant pour la modalité interrogative) ainsi que des valeurs d'intensité inférieures pour la voix chuchotée (avec un écart de 22dB en moyenne), les deux modes de production présentent une même répartition d'intensité et de durée relatives.

En effet, les segments phoniques occupent une proportion équivalente dans les deux modes de production. On constate un allongement de la voyelle de la dernière syllabe porteuse de l'accent, cet allongement étant beaucoup plus marqué dans les phrases longues que dans les phrases courtes.

Pour mesurer l'intensité relative, nous avons soustrait la valeur d'intensité de la consonne à la valeur d'intensité de la voyelle qui suit. L'on obtient en moyenne une intensité relative de 4dB pour la voix chuchotée et de 7dB pour la voix normale, ce qui montre que la proportion d'intensité est préservée dans les deux modes de production.

L'examen rapide de la voix chuchotée par rapport à la voix normale permet de conclure que les paramètres acoustiques de durée et d'intensité interviennent de manière égale dans les deux modes de production.

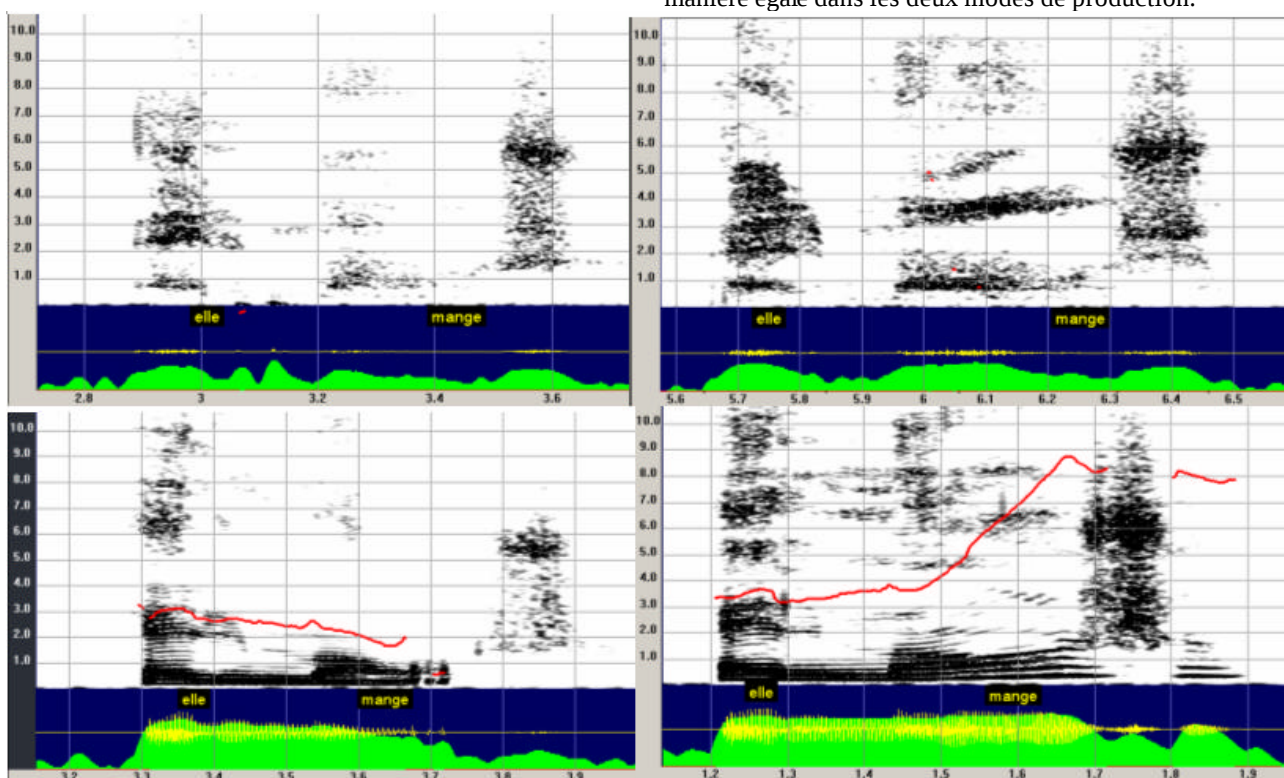


Figure 1 : Exemple de production de l'énoncé contenant 2 syllabes : en haut les productions chuchotées (déclarative à gauche et interrogative à droite) et en bas les productions à voix normale (déclarative à gauche et interrogative à droite)

Comparaison de la voix chuchotée à la voix normale

Passons maintenant aux caractéristiques acoustiques de chaque modalité considérée. De manière générale, Le débit est un plus rapide dans la modalité interrogative (la modalité interrogative contient en moyenne 1 syllabe de plus par seconde par rapport à la modalité déclarative). En

effet, si nous observons les analyses de durée, les segment des énoncés interrogatifs ont tendance à être plus courts que leur correspondant déclaratif. Par contre, la proportion des segments à l'intérieur de la phrase est égale dans les deux modalités, avec notamment un allongement de la dernière syllabe, allongement qui semble être plus marqué pour la modalité déclarative.

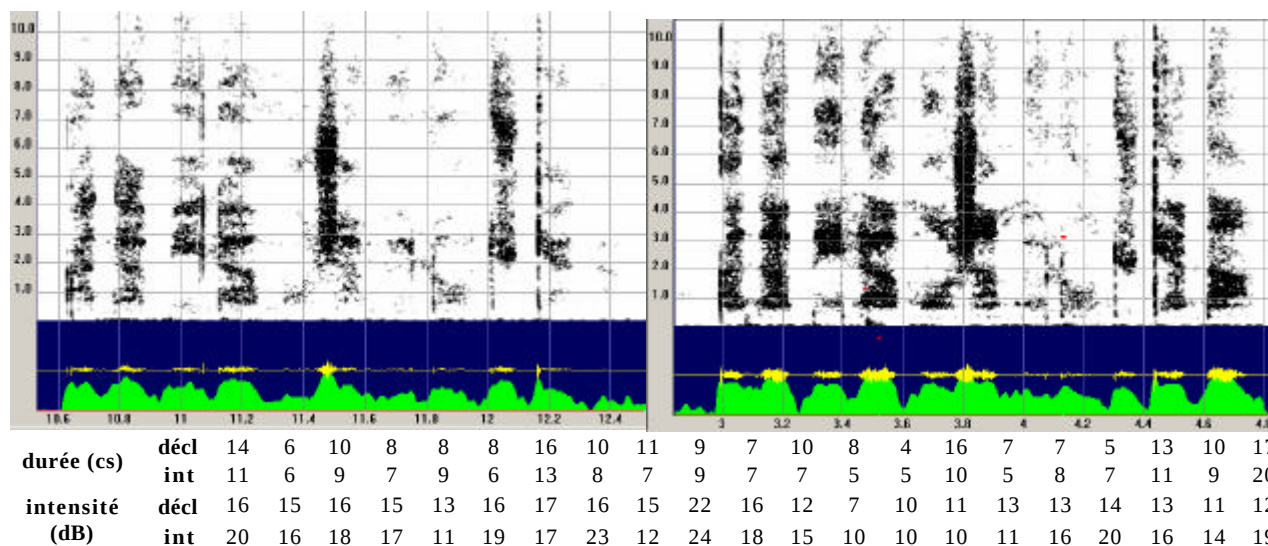


Figure 2 : Exemple de phrase longue (11 syllabes) en voix chuchotée : « Annabella mange un melon du Ghana » : Phrase déclarative à gauche et phrase interrogative à droite ; le tableau donne les valeurs brutes de durée et d'intensité correspondant à chaque segment phonique

En fait, il semble que l'intensité est le seul paramètre qui permette de marquer nettement la différence entre les deux modalités : les valeurs d'intensité sont plus élevées pour la modalité interrogative, l'écart étant important sur les dernières syllabes où l'intensité est maintenue contrairement à la modalité déclarative où la fin de la phrase est marquée par une déclinaison d'intensité.

En somme, nous retiendrons le débit et une intensité forte maintenue jusqu'à la fin de la phrase comme caractéristiques acoustiques distinctives définissant la modalité interrogative en voix chuchotée. La proportion occupée par chaque segment à l'intérieur de la phrase semble être un paramètre non significatif.

3. PERCEVOIR LA VOIX CHUCHOTÉE

3.1. Protocole expérimental de l'étude de perception de la voix chuchotée

Nous avons procédé à un test d'identification afin de tester la reconnaissance des deux modalités d'un point de vue perceptif. Ce test perceptif permet également de voir si la longueur des phrases (nombre de syllabes) a un effet sur le taux de reconnaissance de la bonne modalité. Les stimuli ont été choisis parmi les enregistrements effectués pour l'analyse acoustique. Le test comporte 134 phrases

(une phrase par nombre de syllabes et par modalité, répétée 5 fois de manière aléatoire, les 4 premières phrases n'étant pas prises en compte dans les résultats). Pour ce test de perception pilote, dix francophones normo-entendants, y ont participé. La tâche des sujets était de décider et de noter sur une feuille libre si la phrase entendue était une déclarative ou une interrogative.

3.2. Résultats du test d'identification

Nous pouvons dire que la modalité est globalement bien reconnue en parole chuchotée puisque nous obtenons un taux de reconnaissance de la modalité, tout auditeur confondu, de 76% (73 pour les phrases interrogatives, 78 pour les phrases déclaratives).

Nous avons effectué une ANOVA pour voir si le nombre de syllabes influençait la reconnaissance de la modalité. L'analyse statistique révèle un effet significatif de la longueur sur le taux de reconnaissance ($F=12$ et $p<0.001$ pour les phrases déclaratives et $F=9$ et $p<0.001$ pour les phrases interrogatives), même si cette analyse ne suffit pas à montrer si la reconnaissance croît parallèlement à l'augmentation du nombre de syllabes.

Par contre, l'observation du détail des résultats de bonne reconnaissance va dans le sens de notre hypothèse. Nous avons regroupé les résultats en trois groupes (table 2).

Table 2 : Résultats groupés des taux de reconnaissance de la modalité interrogative et déclarative pour l'ensemble des auditeurs

Modalité	Taux de reconnaissance		
	2 à 5 syllabes	6 à 8 syllabes	9 à 12 syllabes
Interrogative	67,5%	64%	91%
Déclarative	68,5%	80%	86%
Moyenne	68%	72%	88,5%

Les résultats donnés dans la table 2 montrent que plus la phrase est longue, plus le taux de reconnaissance devient meilleur, atteignant presque 90% pour les phrases de 9 à 12 syllabes. En dessous de 9 syllabes, les résultats sont plus disparates.

Ce résultat est conforté par l'observation du coefficient de variance. Pour les phrases interrogatives, nous observons une bonne homogénéité de la distribution des données à partir de la phrase 8 (coefficient inférieur à 15%). Cela montre qu'à partir de neuf syllabes, la modalité est reconnue unanimement.

En ce qui concerne la modalité déclarative, les résultats sont moins homogènes. Cette plus grande variance est due aux résultats d'un auditeur : si on ne prend pas en compte les résultats obtenus pour ce dernier, les résultats sont similaires à ceux de la modalité interrogative.

Ce premier test de perception pilote nous permet de faire les deux remarques suivantes :

- la modalité est globalement bien reconnue en voix chuchotée
- le taux de bonne reconnaissance est corrélé au nombre de syllabes de la phrase.

Toutefois, les résultats pour certaines phrases sont relativement hétérogènes et peuvent aller à l'encontre de nos hypothèses. D'autres tests de perception sont en cours pour déterminer le rôle de la longueur, et pour déterminer s'il existe un seuil plus précis de la reconnaissance.

4. DISCUSSIONS ET CONCLUSION

A travers cette étude sur la voix chuchotée, nous avons montré que malgré l'absence de fréquence fondamentale, la durée et l'intensité sont des paramètres suffisants pour percevoir la réalisation des deux modalités phrastiques, la déclarative et l'interrogative. Par ailleurs, la prise en compte du paramètre de la longueur des phrases montre que plus les phrases sont longues, plus le taux de bonne reconnaissance est élevé. Il semblerait que le facteur rythmique joue un rôle important, puisqu'un certain nombre de syllabes est nécessaire pour permettre aux paramètres d'intensité et de durée de se différencier suffisamment pour permettre la perception des différentes modalités. D'autres études sont en cours pour intégrer

d'autres modalités phrastiques, mais aussi d'autres tests de perception afin d'affiner nos résultats sur le rôle joué par la longueur de la phrase. Par ailleurs, l'analyse acoustique de la production des différentes modalités a permis de dégager que les paramètres de durée et d'intensité n'interviennent pas de manière égale, l'intensité étant un paramètre plus important que la durée, du moins dans la distinction des deux modalités de phrase étudiées.

Cette dernière remarque s'avère intéressante pour une application de l'analyse de la voix chuchotée dans le domaine des implants cochléaires. Il a été remarqué que les sujets implantés ont plus de facilités à percevoir et comprendre la voix chuchotée que la voix normale. Dans la suite de notre travail, nous comptons monter des tests de perception destinés à des implantés en utilisant des stimuli de synthèse où les paramètres d'intensité et de durée pourront être réglés, ce qui nous permettrait de déterminer les meilleurs réglages pour les implants cochléaires.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Bonnot J.F.P., « Electromyographical Correlates of Shouted and Whispered Voice », *Actes du 11ème Congrès International des Sciences Phonétiques*, volume 4, 20-23, Tallinn, Estonie, 1987.
- [2] Fonagy, I, Accent et intonation dans la parole chuchotée, *Phonetica* 20. 177-192, 1969.
- [3] Gao M., *Tones in whispered Chinese : Articulatory features and perceptual cues*, BA, Zhongshan University, 1999
- [4] Harbold. Pitch ratings of voiced and whispered vowels. *JASA* 30 (7), 600-601, 1958
- [5] Higashikawa, M., K. Nakai, A. Sakakura & H. Takahashi, Perceived Pitch of whispered voice—Relationship with formant frequencies: a preliminary study. *Journal of Voice* 10.2: 155~158, 1996.
- [6] Jensen M.C. Recognition of word tones in whispered speech. *Word* 14, 186-196, 1958
- [7] Meyer-Eppler, W., Realization of prosodic features in whispered speech. *JASA* 29: 104~106, 1957.
- [8] Miller, Word tone recognition in Vietnamese whispered speech, *Word* 17, 11-15, 1961.
- [9] Schwartz. M.F, Bilabial closure durations for /p/, /b/, and /m/ in voiced and whispered vowel environments. *JASA* 51 (6), 2025-2029, 1972.
- [10] Tartter V.C What's in whisper? *JASA* 86 (5), 1678-1683, 1989
- [11] Thomas, I.B., Perceived pitch of whispered vowels. *JASA* 46 468~470, 1969.