

Sexe, mensonges et F_0

Estelle Campione, Jean Véronis

Equipe DELIC, Université de Provence
29, Av. Robert Schuman, 13621 Aix-en-Provence Cedex 1, France
Tél.: ++33 (0)4 42 95 31 37 - Fax: ++33 (0)4 42 95 34 95
Mél: Estelle.Campione@up.univ-aix.fr - <http://www.up.univ-mrs.fr/delic>

ABSTRACT

Many contradictory results have been published on male and female voice characteristics, and the debate was sometimes tinged by sexist stereotypes. In a detailed study, Tielen [Tie92] seemed to partly conclude the debate by showing that there is no difference in F_0 range among sexes. We show in this paper that her conclusion was misled by the measure she and many other researchers use (the 90% range), which precisely erases the differences to be observed. We show on a large multilingual corpus involving 60 different speakers, that there are indeed strong differences in the shape of F_0 distribution between sexes. Female voices show high values of skewness and kurtosis, characteristic of long tails in the distribution, whereas no such tendency can be observed for men.

1. INTRODUCTION

S'il n'est contesté par personne que les voix de femmes sont en moyenne plus aiguës que les voix d'hommes, tous les autres aspects de la F_0 , et en particulier l'étendue, ont fait l'objet de publications contradictoires, concluant tantôt à une influence du sexe des locuteurs, tantôt à une absence de différences. Le débat a été coloré d'emblée par divers clichés ou préjugés sexistes, qui remontent aux considérations (fondées sur aucune observation concrète) de Jespersen [Jes22], associant diverses caractéristiques langagières supposées à des traits de caractères également supposés (caractère fort des hommes, caractère impulsif des femmes, etc.). On a ainsi longtemps attribué aux voix de femmes une plus grande étendue et une plus grande variabilité (qui reflèteraient la féminité et l'émotivité), et aux voix d'hommes une étendue réduite et une plus grande monotonie (qui reflèteraient la force et la stabilité).

Les résultats publiés sont disparates et très difficiles à comparer, étant données les conditions expérimentales extrêmement différentes dans lesquelles ils ont été recueillis (différences de langues, de types de parole, d'âges, de techniques d'analyse, etc.). Le débat semblait avoir été clos par l'étude de Tielen [Tie92], qui fait une revue de tous les résultats publiés sur la question et, à travers plusieurs expériences sur le néerlandais, conclut à une absence de différence entre sexes en ce qui concerne l'étendue.

Nous montrons ici, à travers la première étude de grande ampleur, portant sur l'observation de près de 60000 mouvements mélodiques dans un corpus de 5 heures et demi de parole en cinq langues, faisant intervenir deux types de parole (lue et spontanée) et 60 locuteurs hommes et femmes, que les résultats de Tielen sont biaisés par un artefact méthodologique et qu'il existe bel et bien des différences de F_0 imputables au sexe — que rien n'autorise bien sûr à relier à des traits de caractères réels ou supposés.

2. CORPUS

Nous avons utilisé dans cette étude deux corpus faisant intervenir autant de locuteurs hommes que femmes et comportant environ 4 heures et demi de parole lue en cinq langues ainsi qu'environ une heure de parole spontanée en français. Il aurait sans doute été préférable de disposer de parole spontanée dans les cinq langues, mais l'effort correspondant était hors de notre portée. Le présent plan d'expérience, même incomplet, permet néanmoins une étude des facteurs sexe, langue (sur la parole lue) et type de parole (pour le français).

Le corpus lu est le « Few Talker Set » de la base de données EUROM1 produite dans le cadre du projet Esprit SAM [Cha95]. Ce corpus est constitué de passages de cinq phrases chacun, lus dans différentes langues, parmi lesquelles nous avons sélectionné l'allemand, l'anglais, l'espagnol, le français et l'italien. Les passages dans les différentes langues sont la traduction de la version originale anglaise. Ils sont lus par 5 hommes et 5 femmes dans chaque langue (50 locuteurs au total).

Le corpus spontané est extrait du *Corpus Français de Référence du Français Parlé* (ci-après FREF) récemment collecté notre équipe et qui comporte 150 enregistrements d'environ 15 minutes, réalisés dans 40 villes de France. Nous avons sélectionné 10 enregistrements (5 femmes et 5 hommes) et nous avons utilisé dans chacun des passages d'environ 5 à 6 minutes correspondant à des monologues.

3. MÉTHODOLOGIE

La courbe de F_0 brute contient une grande quantité de détails qui constituent du « bruit » pour notre étude. D'une part, ces tracés, quelle que soit la méthode adoptée, contiennent des erreurs de détection (sauts

d'octaves, etc.) ; d'autre part, ils contiennent à la fois une information fréquentielle suprasegmentale (*macroprosodie*) et une information liée aux phonèmes (*microprosodie*) : interruption de la courbe due aux consonnes non voisées, abaissements dus aux constrictives, etc. Nous ne souhaitons retenir dans cette étude que la première composante, macroprosodique (l'étude des différences microprosodiques liées au sexe n'est pas inintéressante, mais devrait faire l'objet d'études séparées).

La F_0 a donc été stylisée sur l'ensemble des deux corpus à l'aide l'algorithme MOMEL qui permet de réduire la F_0 à un ensemble de points-cibles représentant la succession des mouvements macrointonatifs des énoncés. Il a été proposé par [Hir93], publication à laquelle nous renvoyons le lecteur pour des détails techniques (voir aussi [Cam00, [Cam01]). Par interpolation à l'aide d'une courbe lisse (spline), on peut obtenir par resynthèse une F_0 qui doit normalement être perceptuellement indiscernable de l'original. L'algorithme produit toutefois environ 5% d'erreurs [Cam01] dues à des causes diverses, et la totalité des quelques 60000 points-cibles a été corrigée par des experts de chaque langue.

La fréquence des points-cibles a été convertie en demitons (ST), comme le font [Tie92], et de nombreux autres auteurs travaillant sur les différences liées au sexe. D'autres échelles pourraient être utilisées (Mels, Barks, ERB, etc.), mais elles reposent sur des hypothèses supplémentaires concernant la perception qui ne nous semblent pas justifiées au niveau d'analyse que nous proposons, et que nous situons volontairement au niveau physique. De plus, ces échelles ont été élaborées à partir de l'écoute de sons purs ou de voyelles isolées, et il reste à démontrer qu'elles sont appropriées dans le cas de la parole continue, particulièrement la parole spontanée. [Ter00] font remarquer qu'il y a peu d'études détaillées sur le sujet, et qu'elles sont contradictoires : [Her91] ont cru montrer la supériorité de l'échelle ERB sur l'échelle logarithmique pour l'étude de l'intonation, mais [Tra95] prouvent que ce résultat est dû à un biais expérimental.

4. FRÉQUENCES MOYENNES

En parole lue, nous observons pour les fréquences moyennes un décalage de 9,7 ST entre sexes toutes langues confondues, du même ordre que celui observé par [Tie92:83] (9,5 ST). Par contre, en parole spontanée, le décalage est beaucoup moins important, puis qu'il passe de 10,5 ST pour la partie française du corpus EUROM1 à seulement 4,9 ST, soit moins de la moitié. Il est dû à un tassement vers le grave des voix de femmes en parole spontanée, alors que la moyenne des voix d'hommes reste stable. Une analyse de variance montre que cet effet est hautement significatif [$F(1,16459)=1391,74$; $p<0,000$]. Ce résultat contredit celui de [Tie92] qui ne trouve pas de différence signifi-

cative dans le décalage hommes-femmes selon le type de parole.

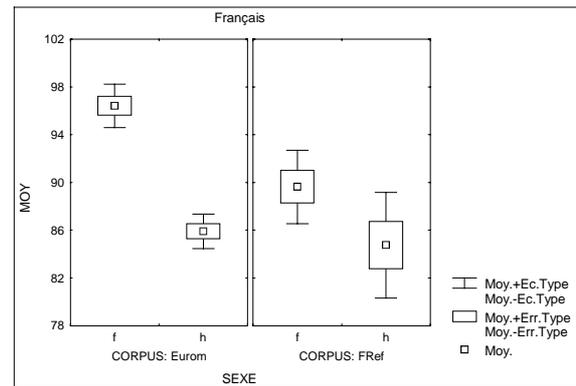


Figure 1. Fréquences moyennes des locuteurs français

5. ETENDUE

L'étendue (en anglais *range*) est définie comme la différence entre la fréquence maximale et la fréquence minimale d'un locuteur. Cette mesure est toutefois fragile, car elle est dépendante de l'échantillon considéré, et qu'il suffit d'un point aberrant d'un côté ou de l'autre de la série pour en changer totalement la valeur. La fragilité de l'étendue a amené de nombreux auteurs, dont [Tie92], à utiliser une valeur corrigée beaucoup plus stable, l'étendue à 90% (*90% range*). Cette étendue est celle mesurée entre les 5ème et 95ème centiles de la distribution. Elle élimine donc la quasi-totalité des valeurs aberrantes ou extrêmes.

En utilisant cette mesure sur nos corpus, on ne note de différence liée au sexe dans aucune des langues. Les valeurs observées de l'ordre de 11 ST pour chaque sexe, ce qui est du même ordre que les valeurs mentionnées dans la littérature.

L'absence de différence significative concernant l'étendue à 90% entre sexes semble donc confirmer l'étude de [Tie92], qui s'inscrivait en faux contre l'association d'une plus grande étendue et de la « féminité ». Toutefois, l'étendue à 90% « gomme » par définition les points extrêmes et les « queues » de la distribution, et ignore la forme de celle-ci. Nous allons donc ci-après examiner plus en détail la forme exacte des distributions de la F_0 pour chaque sexe.

6. DISTRIBUTION

6.1 Asymétrie

Le coefficient d'asymétrie permet de juger de l'égalité de répartition de la F_0 des deux côtés de la moyenne. Une asymétrie négative est révélatrice d'une longue queue à gauche de la distribution, une asymétrie positive d'une longue queue à droite. Il est intéressant de noter que, sur le corpus EUROM1, les voix de femmes sont beaucoup plus asymétriques que les voix

d'hommes (Figure 2). L'asymétrie moyenne est de -0,41 pour les voix de femmes et à peu près nulle pour les voix d'hommes. Les voix de femmes comportent de plus de nombreux cas très asymétriques (voir locuteurs notés « aberrants » et « extrêmes » sur la Figure 2). Il existe par ailleurs de fortes différences entre langues : les langues pour lesquelles l'asymétrie des voix de femmes est la plus prononcée sont l'allemand, l'espagnol et l'italien (voir détail dans [Cam01]).

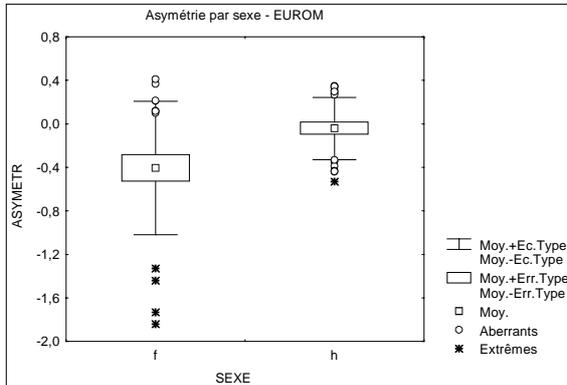


Figure 2. Asymétrie par sexe (EUOM1)

La comparaison parole lue/parole spontanée révèle une forte différence entre les deux corpus sur le français. On note tout d'abord une beaucoup plus grande variabilité inter-locuteurs en parole spontanée. On peut supposer que la parole lue étant fortement codifiée et faisant l'objet d'un apprentissage scolaire spécifique, les différences inter-locuteurs ont tendance à être moins fortes. Plus surprenant, la différence d'asymétrie, négligeable sur le français en parole lue, se révèle très importante en parole spontanée (Figure 3).

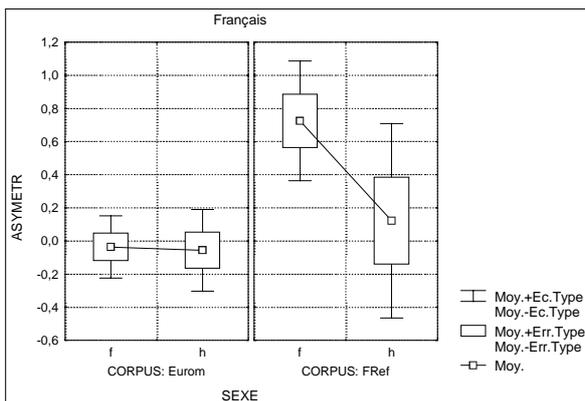


Figure 3. Asymétrie (fr., parole lue vs spontanée)

6.2 Aplatissement

L'aplatissement ou *kurtosis* permet de mesurer la « hauteur » d'une distribution. L'aplatissement de la distribution normale est de 0. Si l'aplatissement est positif, la distribution est plus « plate » que la distribution normale avec de longues « queues », s'il négatif, la distribution est plus « pointue » que celle-ci.

Les différences entre sexes sont importantes sur le corpus EUOM1, puisqu'on observe un aplatissement moyen important pour les femmes (1,54) et à peu près nul pour les hommes (Figure 4). On note à nouveau une variabilité inter-locuteurs importante, en particulier chez les femmes, avec divers cas d'aplatissement extrême. Les différences entre langues sont également très importantes pour les femmes, avec des valeurs importantes pour l'allemand, l'anglais et l'espagnol.

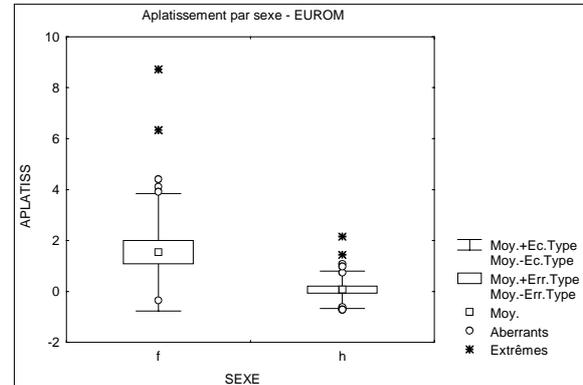


Figure 4. Aplatissement par sexe (EUOM1)

La différence parole lue/parole spontanée est également notable, puisque l'aplatissement du français est à peu près nul pour les deux sexes sur le corpus EUOM1 (avec une variabilité inter-locuteurs très faible) alors qu'il devient important pour les femmes (0,93) et augmente seulement légèrement pour les hommes dans le corpus FREF (0,22) (Figure 5).

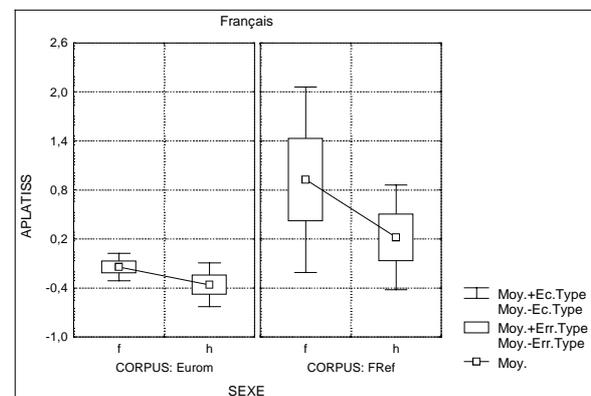


Figure 5. Aplatissement (fr., parole lue vs spontanée)

La représentation combinée de l'asymétrie et de l'aplatissement sous forme de nuage de points, toutes langues combinées, est extrêmement révélatrice (Figure 6). Dans la parole lue, les voix d'hommes sont toutes centrées autour de valeurs moyennes de l'asymétrie et de l'aplatissement. A l'inverse, de nombreuses voix de femmes s'en écartent et font preuve à la fois d'aplatissement et d'asymétrie négative, ce qui est révélateur d'une longue « queue » de points-cibles dans le grave.

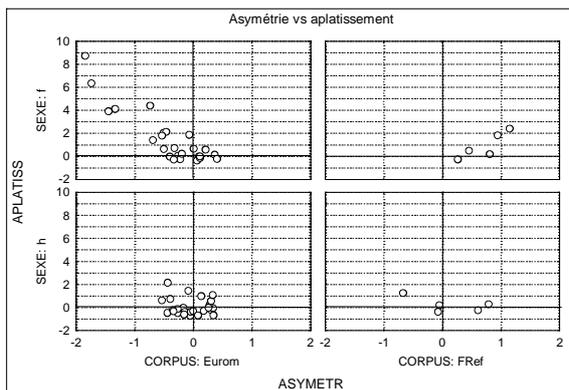


Figure 6. Asymétrie vs aplatissement

Dans le corpus de parole spontanée, on retrouve bien le groupement des hommes autour du point central, mais la tendance pour les locutrices est inverse : aplatissement avec asymétrie positive, indiquant une « queue » plus longue du côté aigu de la distribution.

7. CONCLUSION

Des résultats contradictoires et de nombreuses controverses, parfois sur fond de préjugés sexistes ont émaillé la recherche sur les différences entre sexes concernant la F_0 . L'étude de Tielen [Tie92] semblait avoir clos le débat en montrant, sur le néerlandais, qu'il n'y avait pas de différence d'étendue de F_0 selon le sexe. Toutefois, nous avons montré dans cette étude que la mesure utilisée par [Tie92] (et de nombreux autres auteurs), l'étendue à 90% (90% range), a justement pour caractéristique de gommer les différences que l'on souhaite observer. Nous montrons, sur un large corpus (cinq langues, 5 heures et demie de parole lue et spontanée, 60000 mouvements mélodiques, 60 locuteurs) qu'il existe bien des différences importantes concernant la F_0 entre sexes. Les voix de femmes ont une tendance très forte à l'asymétrie et à l'aplatissement (avec des variations importantes selon les langues), tandis qu'on n'observe aucune tendance analogue pour les voix d'hommes. Curieusement, la tendance est à l'asymétrie négative (caractéristique d'une longue « queue » de points dans le grave) pour les femmes en parole lue, alors qu'elle est à l'asymétrie positive (caractéristique d'une surabondance de points dans l'aigu) pour la parole spontanée, du moins pour le français.

Nous nous gardons de conclure de façon définitive sur les causes possibles des effets observés. Malgré la taille de nos données, il conviendrait, étant l'aspect sensible du débat, de multiplier les études avec d'autres locuteurs et d'autres situations, de façon à exclure la possibilité d'accidents statistiques dus à l'influence de facteurs sociolinguistiques corrélés avec les groupes de locuteurs tel que l'âge, l'usage du tabac, etc., ou des conditions d'enregistrement (chambre sourde, public, etc.). Nos observations montrent en tous cas clairement que, sur la question des rapports entre sexe et F_0 , les résultats antérieurs ne peuvent être tenus pour défini-

tifs, et que les études les mieux conduites peuvent contenir des artefacts méthodologiques amenant à des conclusions erronées.

BIBLIOGRAPHIE

- [Jes22] Jespersen O. (1922), *Language: its nature, development and origin*. Allen and Unwin, London.
- [Cam01] Campione E. (2001), *Étiquetage automatique de la prosodie dans les corpus oraux : algorithmes et méthodologie*. Thèse de l'Université de Provence, Aix-en-Provence.
- [Cam00] Campione E., Hirst D., & Véronis J. (2000), Stylistic and symbolic coding of F_0 : comparison of five models. In A. Botinis (Ed.), *Intonation: Models and Theories* (pp. 185-208). Dordrecht: Kluwer Academic Publishers.
- [Cha95] Chan D. et al. (1995), Eurom - A Spoken Language Resource for the EU, *Proceedings of the 4th European Conference on Speech Communication and Speech Technology (Eurospeech'95)* (Vol. 1, pp. 867-870). Madrid.
- [Her91] Hermes D., & Van Gestel J. (1991), The frequency scale of speech intonation. *Journal of the Acoustical Society of America (JASA)*, 90, 97-102.
- [Hir93] Hirst D. J., & Espesser R. (1993), Automatic Modelling of Fundamental Frequency using a quadratic spline function. *Travaux de l'Institut de Phonétique d'Aix-en-Provence (TIPA)*, 15, 75-85.
- [Ter00] Terken J., & Hermes D. (2000), The perception of prosodic prominence. In M. Horne (Ed.), *Prosody: Theory and Experiment. Studies Presented to Gösta Bruce* (pp. 89-127). Dordrecht: Kluwer Academic Publisher.
- [Tie92] Tielen M. (1992), *Male and Female Speech: an experimental study of sex-related voice and pronunciation characteristics*. Omslag: Josan Tielen.
- [Tra95] Traunmüller H., & Erikson A. (1995), The perceptual evaluation of F_0 -excursions in speech as evidenced in liveliness estimations. *Journal of the Acoustical Society of America (JASA)*, 97, 1905-1915.