

La dynamique de la fréquence fondamentale de la voix comme indicateur de la distance interlocuteur

Thibaut Fux, Gang Feng, Véronique Zimpfer

► **To cite this version:**

Thibaut Fux, Gang Feng, Véronique Zimpfer. La dynamique de la fréquence fondamentale de la voix comme indicateur de la distance interlocuteur. 3èmes Journées Perception Sonore, Jun 2017, Brest, France. 2017. hal-01559879

HAL Id: hal-01559879

<https://hal.univ-brest.fr/hal-01559879>

Submitted on 21 Jul 2017

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.

La Dynamique de la Fréquence Fondamentale de la Voix comme Indicateur de la Distance Interlocuteur

Thibaut FUX^{1,2,3}, Gang FENG², Véronique ZIMPFER³

- ¹ Institut de Recherche Biomédicale des Armées (IRBA), Brétigny-sur-Orge, France
² GIPSA-Lab INPG, Université Grenoble-Alpes, Grenoble, France
³ Institut de Recherches Franco-Allemand de Saint-Louis (ISL), Saint-Louis, France



Objectifs

- ▶ Caractériser la voix criées / l'effort vocal
- ▶ Déterminer les éléments perceptivement pertinents de l'effort vocal par l'Analyse-Synthèse
- ▶ Piste pour la transformation de voix modales en voix criées
- ▶ Piste pour la perception auditive de la distance d'un locuteur

Caractéristiques des voix criées

Tout est différent entre la voix modale et la voix criée

- ▶ **Paramètres Acoustiques**
 - ▶ Fréquences/Largeurs de bande des formants, Pente spectrale
- ▶ **Paramètres Prosodiques**
 - ▶ F0, I, Durée (phonèmes, mots, ...)
- ▶ **Qualité de Voix**
 - ▶ Quotient ouvert, HNR (Harmonic-to-noise ratio), ...
- ▶ **Hyphotèse**
 - ▶ **Les variations prosodiques sont perceptivement pertinentes [1,2,3].**

Corpus en Laboratoire



FIGURE 1: Photo et schéma du protocole d'enregistrement.

- ▶ Atténuation : 36dB (env. 60m)
- ▶ Voix parlée et criée
- ▶ Acoustique + Electroglottograph
- ▶ Retour sur la compréhension
- ▶ locuteurs : 3 hommes
- ▶ 51 logatomes
- ▶ 3 voyelles (/a/, /i/, /u/)
- ▶ 17 consonnes
- ▶ CV, C₁VC₁, V₁CV₁ et C₁V₁C₁V₁

Méthodologie

- ▶ Analyse par phonème : F0, Intensité, Durée
- ▶ Estimation de F0 sur le signal EGG (plus précis pour de larges ambitus)

Analyses Prosodiques de l'Effort Vocal

Résultats : Analyses

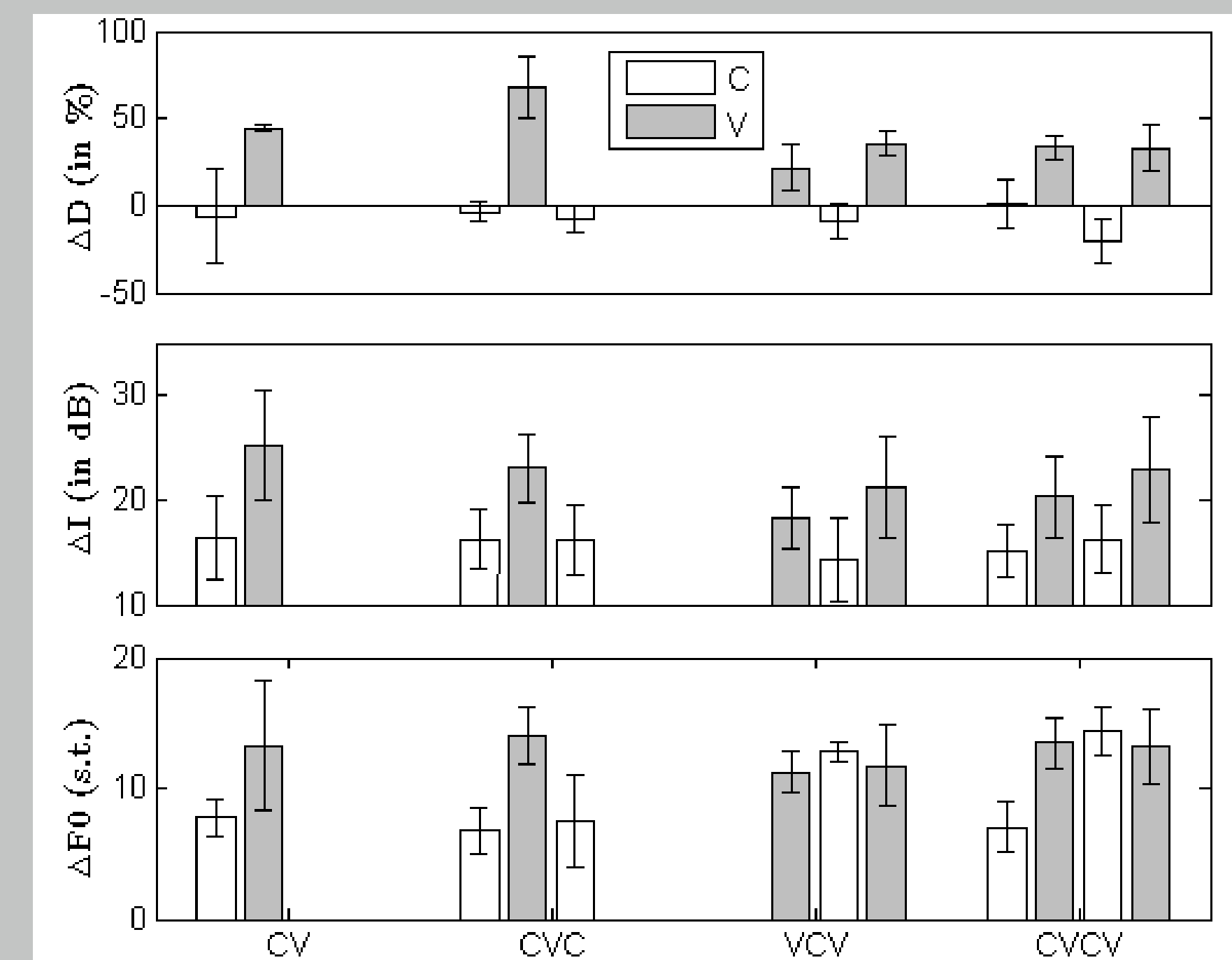


FIGURE 2: Variations de la fréquence fondamentale (F0), de l'intensité I et de la durée (D) entre voix parlées et voix criées en fonction de la structure du mot et de la position des phonèmes.

- ▶ F0 et I augmentent plus avec l'effort vocal pour les voyelles que pour les consonnes ; excepté pour la F0 des consonnes intervocaliques (i.e. VCV et CVCV).
- ▶ Forte augmentation de la durée des voyelles ; celle des consonnes tend à diminuer.
- ▶ Variations relatives plus marquées en monosyllabiques.
- ▶ Variation plus forte de l'Intensité de la 2ème consonne que pour la 1ère (typ. 2-3 dB).

L'effort Vocal impacte différemment les C et les V

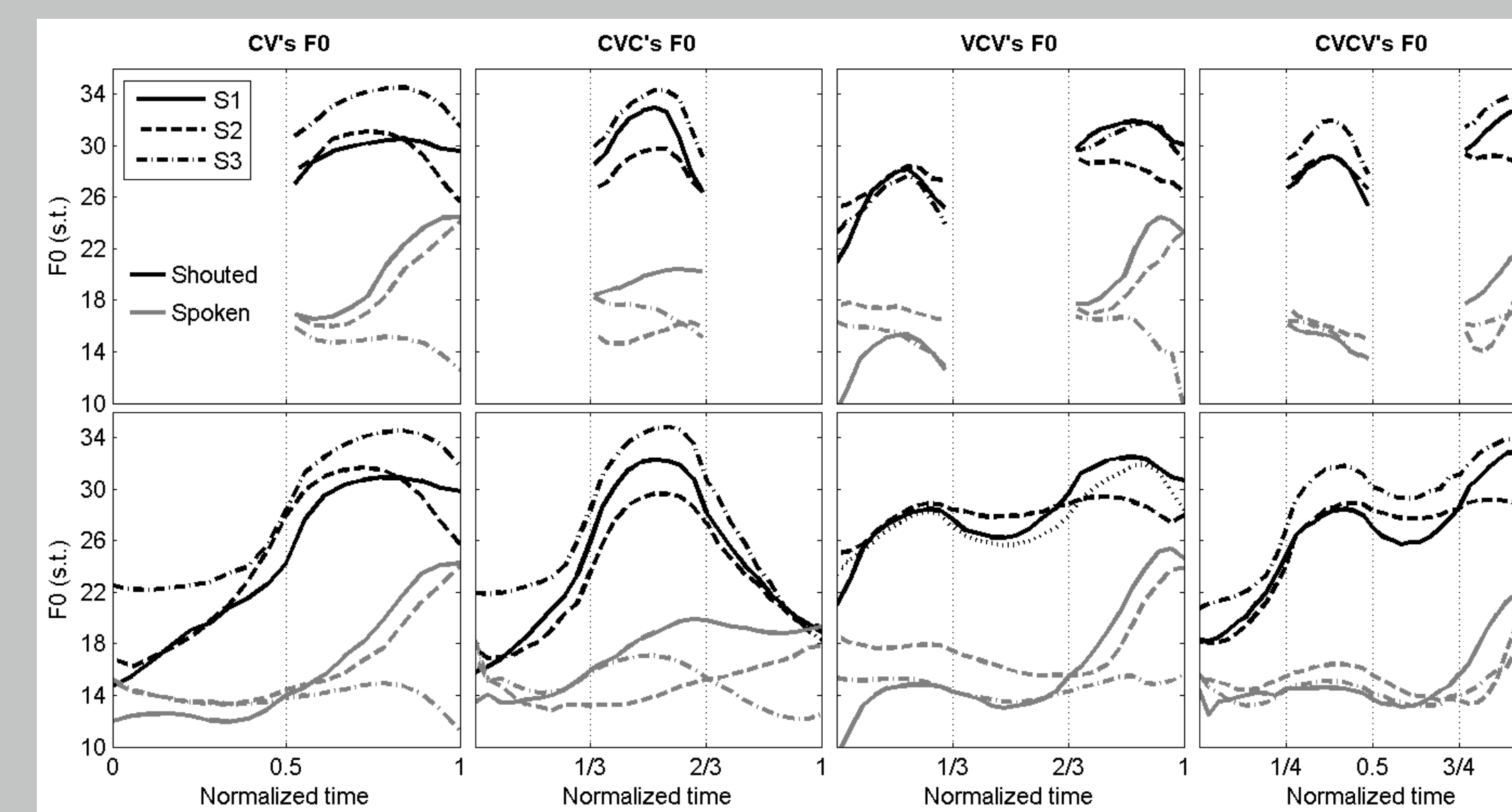


FIGURE 3: Contours F0 moyens des logatomes parlés et criés. La F0 est représentée en demi-ton (réf : 50Hz). Les contours sont tracés sur un axe normalisé en temps dans lequel chaque phonème occupe une portion égale de temps.

- ▶ Voix parlées de S1 et S2 sûrement affectées par l'effet de liste
- ▶ Patterns de F0 similaires en voix criées en fonction de la structure des logatomes
- ▶ 3 sous patterns visibles : (1) Consonne initiale/finale, (2) Consonne intervocalique (3) voyelle

Homogénéisation des contours de F0 avec l'effort vocal

Stylisation et Règles de Transformation de Voix Modales

L'aspect particulier de F0 est-il perceptivement important ?

- ▶ **F0** : Stylisation par des fonctions mathématiques partielles

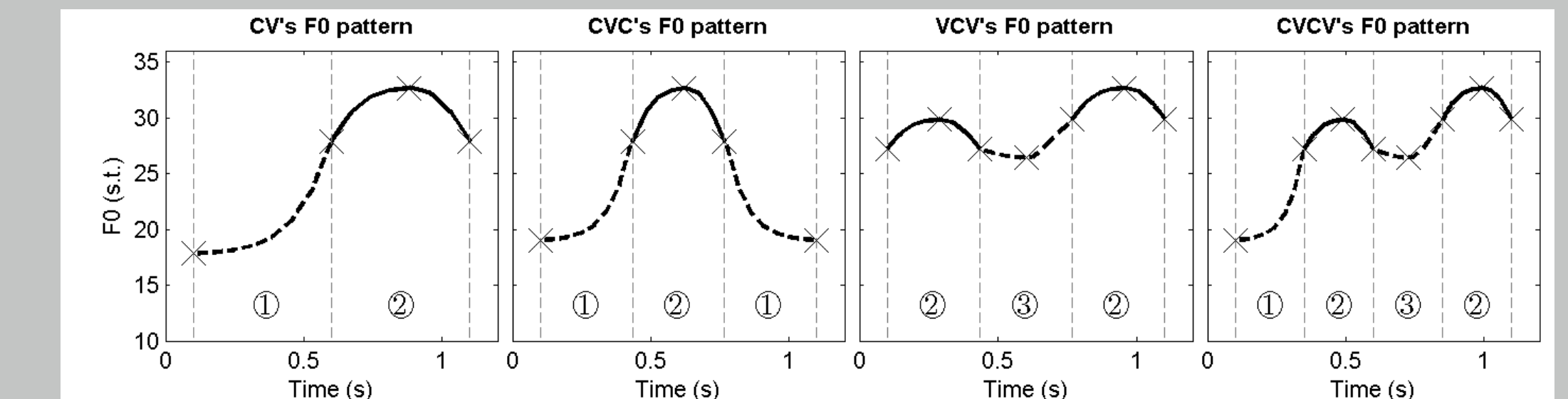


FIGURE 4: Résultat de la stylisation des contours de F0 pour les voix criées

- ▶ **Intensité** : +6dB pour la 1ère voyelle et +8dB pour la deuxième
- ▶ **Durée** : +40% pour les voyelles [4,5,6]
- ▶ **Formant** : F₁ multiple de F0 (Formant tuning) [7]
- ▶ **Pente Spectrale** : Réduite de 3dB/octave [8,9,10]
- ▶ Transformation par RELP-PSOLA [11] pour certain CV et CVCV uniquement (/pap/ /gag/ /ʃaʃ/ /vav/ /nan/ /lal/ /uas/ /papa/ /gaga/ /ʃaʃa/ /vava/ /nana/ /lala/ /uasa/)
- ▶ **2 Transformations** :
 - ▶ **"With"** : avec stylisation de F0
 - ▶ **"Without"** : uniquement en multipliant la F0 des voix parlées

Résultats : Test Perceptif

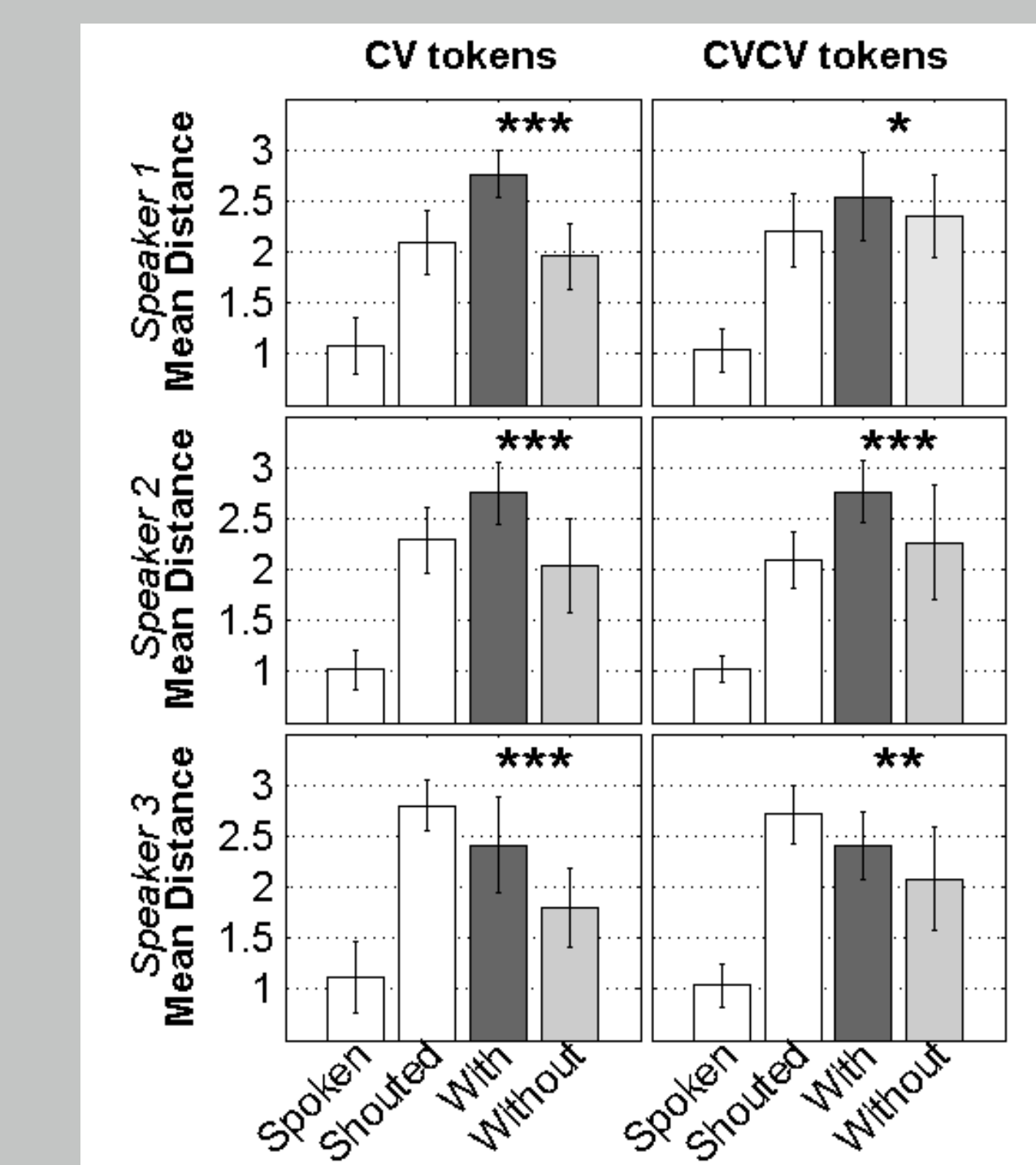


FIGURE 5: Résultat du test perceptif à choix forcé de perception auditive de la distance d'un locuteur. (proche=1 ; moyen=2 ; loin=3)

"With" et "Without" ne diffèrent que par la dynamique de F0. La valeur moyenne est identique.

- ▶ "With" et "Without" perçues éloignées (plus que la voix criée car stylisation de F0 sur S3)
- ▶ "With" perçues plus loin que "Without"

Conclusions

- ▶ Importance de la dynamique de F0 sur la perception de l'effort vocal et donc de la distance d'un locuteur
- ▶ La prosodie devrait être systématiquement considérée dans le cas d'étude sur l'effort vocal