Représentation et Estimation de la Force de Voix à partir du Spectre Moyen à Long Terme

Jean-Sylvain Liénard

Représentation et Estimation de la Force de Voix à partir du Spectre Moyen à Long Terme

Jean-Sylvain Liénard
LIMSI, 91405 Orsay cedex, France
jean-sylvain.lienard@limsi.fr

RESUME

La présente étude vise à retrouver par le calcul le niveau sonore émis par un locuteur, à partir de la seule enveloppe du spectre à long terme. Les données utilisées consistent en un ensemble de Spectres Moyens à Long Terme en tiers d'octave, étalonnés en niveau sonore et comportant une grande variabilité en fonction du genre du locuteur, de son âge et du degré d'effort vocal requis. La représentation visuelle des spectres montre qu'il est plus cohérent de les regrouper en fonction du niveau émis qu'en fonction du degré d'effort vocal requis. Une procédure de comparaison est appliquée à l'ensemble des spectres, après normalisation à une valeur commune, arbitraire, de leur niveau sonore. Les résultats indiquent que la forme du spectre est suffisante pour retrouver le niveau sonore émis, avec une marge d'erreur statistique inférieure à 5 dB.

ABSTRACT

Representing and Recovering Voice Strength from the Long Term Average Spectrum

The goal of the study is to recover the Sound Pressure Level emitted by a speaker, from the single long term spectrum envelope. The data consists of a set of 1/3rd octave Long Term Average Spectra, calibrated in sound level and exhibiting a large variability according to the speaker's gender, age and requested vocal effort degree. The visual representation of the spectra shows that it is more coherent to group them according to the emitted sound level than from the requested vocal effort degree. A comparison procedure is then applied to the data, after normalization of the spectra to a common, arbitrary value of their sound level. The results indicate that the single spectral envelope is sufficient to recover the emitted sound level, within a statistical margin of error smaller than 5 dB.

MOTS-CLES : Effort Vocal, Force de Voix, Spectre Moyen à Long Terme.
KEYWORDS: Vocal Effort, Voice Strength, Long Term Average Spectrum.

1 Introduction

L'Effort vocal (EV) joue un rôle essentiel dans l'interaction orale: le parleur ajuste l'intensité de sa voix, selon la situation, de façon à se faire bien comprendre par son interlocuteur. Ce faisant il modifie notablement les structures spectro-temporelles du signal qu'il émet. Ces modifications sont reconnues par l'interlocuteur, qui peut ainsi juger de l'intensité émise, indépendamment du

10.21437/JEP.2018-71

niveau sonore parvenant à son oreille. Si l'on se place hors du contexte de l'interaction orale, ces modifications apparaissent comme une variabilité indésirable, qui complique la recherche des structures phonétiques du signal de parole ainsi que son traitement automatique.

Les travaux menés antérieurement dans la recherche des effets de l'EV sur les structures de la parole (HANSON, 1997; HUBER et al., 1999; LIENARD and DI BENEDETTO, 1999; TRAUNMULLER and ERIKSSON, 2000) ou dans la perspective de retrouver le niveau émis (LIENARD and BARRAS, 2013; LIENARD, 2014) portent sur des voyelles isolées ou de très courtes phrases et ne s'appliquent guère à la voix "conversationnelle" (de très faible à très forte) utilisée majoritairement dans les situations ordinaires de l'interaction orale. Ces travaux s'appliquent encore moins aux extrêmes que sont les voix criées ou chuchotées, qui correspondent à des situations exceptionnelles dans lesquelles on doit accepter une perte d'intelligibilité.

En voix conversationnelle les déformations spectro-temporelles induites par les variations d'effort vocal peuvent être considérées comme des variations de timbre. Le timbre individuel du locuteur, ou celui qu'il donne à sa voix pour faire passer telle ou telle émotion ou expression stylistique, se traduit en partie par des variations de force de voix (RILLIARD et al., 2018).

L'étude vise à mettre en évidence les modifications spectrales causées par l'effort vocal et à en déduire une estimation de l'intensité fournie par le locuteur. Cette démarche se heurte à deux difficultés. La première est que la notion d'EV elle-même est mal définie, avec des qualificatifs tels que voix modale, normale, faible, forte, criée, confidentielle, sourde, feutrée, stridente, etc., qui relèvent tout autant de la dimension de timbre que de celle d'intensité. La seconde tient à la mesure de l'intensité de la voix, nécessaire pour une approche objective du problème.

L'intensité est souvent négligée en phonétique, comme en traitement automatique de la parole. Au mieux elle est utilisée en valeur relative et caractérisée par sa variation au long d'une même séquence orale. Nous nous intéressons ici à l'intensité sonore dans l'absolu, évaluée sur une durée de plusieurs secondes afin de moyenner les variations syllabiques et prosodiques. La distinction entre intensité émise (par le parleur) et intensité reçue (par l'auditeur ou par le microphone) est essentielle. Pour réduire le risque d'ambiguïté ainsi que l'imprécision de la notion d'Effort Vocal nous emploierons le terme de Force de Voix (FDV) pour désigner l'intensité moyenne émise par seconde de signal (niveau équivalent L_{EQ}, en décibels).

Il n'existe pas aujourd'hui de base de données publique de grande dimension qui permette d'associer à une séquence de parole une mesure fiable de la FDV. Pourtant une telle base de données a été ralisée en 1977 par Pearsons et al. dans le but d'élaborer des normes d'intelligibilité dans le bruit (PEARSONS et al., 1977). Les enregistrements sonores ont été égarés par la suite, mais les relevés de mesure (Spectres Moyens à Long Terme, SMLT) ont été retrouvés, numérisés et mis à disposition de la recherche par Anthony Nash (NASH, 2014). C'est sur ces relevés que porte la présente étude, qui comporte deux parties principales. La première consiste en une analyse qualitative des données, à partir de leur représentation graphique. La seconde vise à retrouver la FDV par le calcul, à partir de la forme du spectre après recalage de tous les SMLT à une valeur commune arbitraire de leur FDV.

2 Représentation de l'Effort Vocal

Dans cette partie les données sont représentées graphiquement dans le but de faire apparaître qualitativement les relations entre le SMLT et l'EV. Ces relations, relativement floues si l'on décrit l'EV par la consigne d'effort vocal donnée aux locuteurs, s'avèrent beaucoup plus nettes si l'EV est représenté par la mesure effective de la FDV.

2.1 Les données de Pearsons et al.

L'objectif de l'équipe de Pearsons était de définir le niveau de bruit maximum tolérable dans des lieux publics ou privés tels que trains, avions, hôpitaux, écoles, appartements, sans compromettre l'intelligibilité de la parole. Des mesures extensives et soignées du niveau sonore ont été effectuées dans ces lieux. De plus, un corpus de parole a été enregistré dans une chambre anéchoïque en conditions contrôlées (à 1 m dans l'axe de la bouche du sujet, avec le microphone du sonomètre lui-même). Le matériau de parole était une phrase sans signification, phonétiquement équilibrée ("Joe took father's shoe bench out; she was waiting at my lawn"), traditionnellement utilisée dans les tests de qualité des systèmes téléphoniques. Il était demandé aux 97 locuteurs non-professionnels (48 hommes, 37 femmes, 12 enfants de moins de 13 ans) de prononcer cette phrase de manière répétitive pendant au moins 10 secondes, selon 4 consignes vocales: "normal", "raised" (appuyé), "loud" (fort), "shout" (crié). De plus une conversation informelle entre le sujet et l'opérateur distant de 1 m s'ajoutait aux précédents enregistrements et recevait le qualificatif de "casual" (détendu, décontracté). Il s'agissait dans ce cas de voix relativement faible et le contenu phonétique n'était pas spécifié. Le nombre total d'enregistrements s'élevait à 482.

Les sons enregistrés étaient traités ensuite par un analyseur (banc de 24 filtres en tiers d'octave, dont les fréquences centrales allaient de 50 Hz à 10 kHz). L'énergie par canal, sommée tout au long de la séquence, était conservée et divisée par la durée, aboutissant à deux mesures du niveau équivalent (L_{EQ}), en dB (SPL) et en dBA (pondéré). L'ensemble des 24 valeurs constitue le SMLT. Divers quantiles de la distribution d'énergie dans chaque canal (énergie mesurée toutes les 0,1 s) sont également fournis avec les valeurs spectrales moyennes.

Le SMLT est utilisé dans les domaines du bruit et de la voix, le plus souvent en échelle de fréquence linéaire (LÖFQVIST, 1986; NORDENBERG and SUNDBERG, 2004). S'il est établi sur une durée courte, de l'ordre de quelques syllabes, le SMLT s'avère extrêmement variable. Mais il se stabilise dans la durée et devient indépendant du contenu phonétique de la séquence orale enregistrée. Il reflète alors certaines caractéristiques de la voix du locuteur. La durée requise est normalement d'une quarantaine de secondes mais une durée plus courte (10 à 20 secondes) est suffisante pour comparer entre elles des séquences orales de même contenu phonétique.

Le travail de Pearsons et al. a été récemment reproduit (CUSHING et al., 2011) sur les mêmes bases (enregistrement étalonné en chambre sourde, 50 locuteurs anglophones, même dispositif d'analyse en 24 canaux). Une définition plus précise de la consigne vocale a été mise en œuvre, et une catégorie "hushed" (voix feutrée) a remplacé la catégorie "casual", produisant des niveaux d'environ 5 dB plus faibles. Ces données ne sont pas publiques. Les auteurs estiment que leurs mesures sont en moyenne très proches (à moins de 2 dB près) de celles de Pearsons et al, confirmant tout l'intérêt de celles-ci en tant que référence pour les études de l'effort vocal.

2.2 Variations individuelles du SMLT

La figure 1 montre deux exemples des SMLT produits par un locuteur (à gauche) et une locutrice (à droite) de même classe d'âge, selon les 5 consignes vocales proposées; par ordre de FDV croissante: "casual" (trait vert), "normal" (noir), "raised" (bleu), "loud" (violet) et "shout" (rouge). Le niveau global en dB SPL est indiqué avec la même couleur sur chaque figure. Chaque canal est désigné par son numéro; pour faciliter l'interprétation fréquentielle des tracés certains traits verticaux sont renforcés; les abscisses 4, 7 et 21 correspondent respectivement à 100, 200 et 5000 Hz, les abscisses 11 et 17 à 500 et 2000 Hz et l'abscisse 14 correspond à 1000 Hz. La zone des fondamentaux usuels se trouve entre les abscisses 3 (80 Hz) et 9 (320 Hz) et la zone du second

formant est centrée sur les abscisses 13 à 17 (800 à 2000 Hz). Il faut noter qu'un maximum du SMLT ne représente pas la fréquence d'un fondamental ou d'un formant, mais la fréquence moyenne autour de laquelle il évolue tout au long de la séquence. Cette grandeur peut être qualifiée de "fréquence dominante" (F0d, F2d etc).

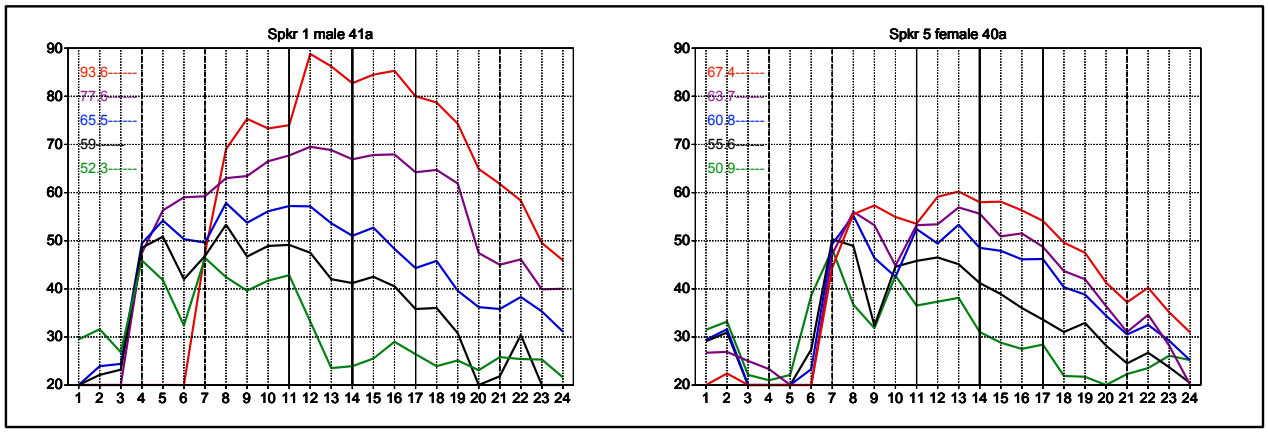

FIGURE 1: SMLT d'un locuteur (à g.) et d'une locutrice (à dr.) pour les 5 consignes vocales.

Ces deux exemples illustrent la grande variabilité des données. Pour une même consigne vocale deux locuteurs peuvent produire des FDV très différentes, variant de 25 dB et plus. Ainsi la production "shout" de la locutrice a une FDV de 67,4 dB et se trouve proche du mode "raised" du locuteur (à 65,5 dB) alors que le mode "shout" de ce dernier atteint 93,6 dB. Parallèlement à ces différences de FDV, les tracés des SMLT montrent une différence importante en basse fréquence, F0d se trouvant majoritairement au-dessous du canal 6 (160 Hz) pour le locuteur et au dessus pour la locutrice. Ceci à l'exception du mode "shout", pour lequel le locuteur atteint une F0d de l'ordre de 320 Hz, voisin de celui de la locutrice dans le même mode.

La progression d'un mode à l'autre s'accompagne de déformations d'ensemble: F0d se déplace vers l'aigu et le maximum spectral se déplace de F0d jusque vers 1000 Hz. La pente du spectre varie beaucoup selon la FDV. Pour les FDV moyennes l'amplitude décroît régulièrement de -6 à -8 dB/octave dans la partie haute du spectre. Le spectre des voix très fortes évolue de manière différente, la pente passant d'une valeur faible (0 à -3 dB/octave) dans la zone centrale (canaux 11 à 18) à une valeur forte (-12 à -18 dB/octave) dans l'aigu. Ces observations peuvent être mises en rapport avec les études théoriques des modèles d'onde glottique (DOVAL et al., 2006).

La grande différence entre les deux cas illustrés par la figure 1 n'est nullement exceptionnelle. En fait, les locuteurs ne respectent pas toujours la consigne vocale qui leur a été proposée. D'une manière générale les locutrices et les enfants ont tendance à produire des FDV maximales plus faibles que celles des locuteurs masculins. Il s'ensuit que la consigne vocale n'est pas un critère fiable de la FDV. Par contre, hors la zone de F0d, on observe une grande ressemblance des SMLT correspondant à une même FDV. Ceci permet de poser l'hypothèse selon laquelle la FDV, intensité émise, peut être estimée quantitativement à partir du seul profil du SMLT, indépendamment de l'intensité reçue.

2.3 A propos de la dynamique des enregistrements étalonnés en niveau

Les relevés indiquent que le rapport signal/bruit des enregistrements est proche de 100 dB ("dynamique d'enregistrement"). Cet intervalle doit être clairement distingué de l'intervalle séparant le niveau des voix les plus faibles de celui des voix les plus fortes ("dynamique de la voix"), qui excède rarement 50 dB. Dans la zone "conversationnelle" allant de "casual" à "loud"

l'intervalle est habituellement de 25 à 30 dB ("dynamique conversationnelle"). Il faut aussi prendre en considération le rapport signal/bruit propre à chaque SMLT ("dynamique propre" du SMLT), c'est-à-dire la différence de niveau entre le maximum spectral et le bruit de fond. La figure 1 montre un rebond d'intensité dans les canaux 1 et 2 (50 et 63 Hz), dans une zone où la voix n'a aucune énergie. Il s'agit de bruits parasites, qui limitent la dynamique propre des séquences les plus faibles et peuvent interférer avec le processus de normalisation et seuillage décrit plus loin.

2.4 Deux distributions selon le genre et l'âge des locuteurs

L'étude de Pearsons et al. permet de distinguer 3 catégories de locuteurs: hommes, femmes, enfants de moins de 13 ans. Les voix d'enfants et de femmes sont proches et il est légitime de les considérer ensemble, ce que nous ferons dans la suite. L'étude de certains indices comme les barycentres de diverses parties du SMLT montre l'existence de deux distributions distinctes. La figure 2 (à g), obtenue à partir des SMLT normalisés et seuillés entre 0 et 50 dB (cf section 3 ci-après), représente la fréquence du centre de gravité spectral en fonction de la FDV. Cette fréquence est exprimée en termes de numéros de canal (le canal 14 est à 1000 Hz).

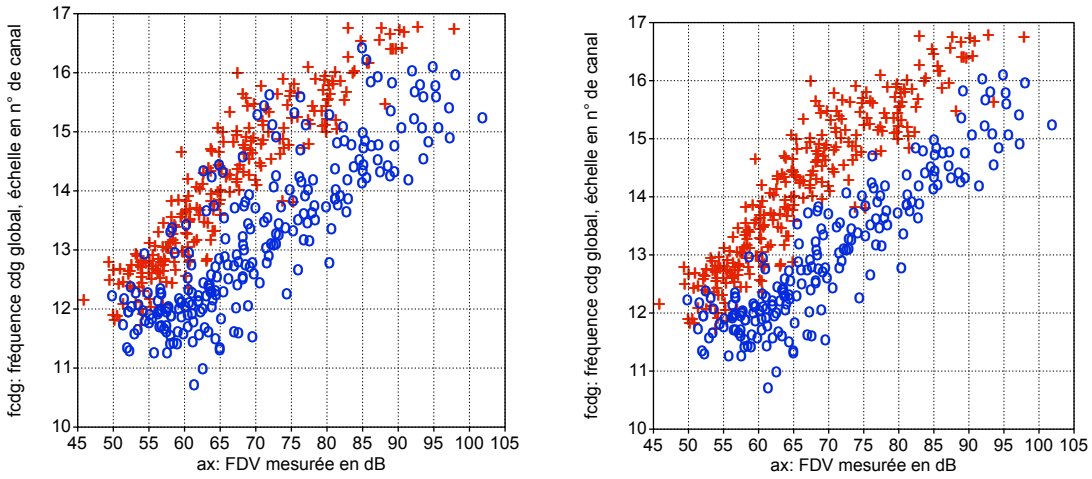

FIGURE 2: position du centre de gravité du SMLT en fonction de la force de voix.
En bleu: voix m adultes. En rouge: voix f et enfants <13 ans (à g), <16 ans (à dr).

On constate l'existence de deux nuages de points distincts (voix f et e en rouge, m en bleu) quasiment parallèles et croissants: à FDV constante le passage d'un groupe à l'autre entraîne un décalage du centre de gravité de 1,5 à 2 tiers d'octave. Un décalage comparable, de l'ordre d'une octave, s'observe à propos de la fréquence fondamentale dominante. Ces observations apparaissent massivement dans le SMLT à cause de l'échelle logarithmique qui dilate la zone basse du spectre au détriment de la zone haute.

Par ailleurs le rapport Pearsons définit les catégories "garçons" et "filles" par un âge maximum de 12 ans. La figure 2 (à g.) montre que de nombreux points provenant de voix masculines apparaissent dans le groupe des voix féminines. Il s'agit de jeunes locuteurs masculins. Du point de vue vocal il est donc plus réaliste de placer la limite à 15 ans, de façon à ce que le groupe des voix m ne comprenne que des hommes adultes, dont la mue est achevée. En procédant ainsi une trentaine de points initialement affectés au groupe masculin se trouvent réintégrés dans le groupe féminin (fig 2, dr.). Dans la suite nous considérerons donc seulement 2 groupes de locuteurs, le groupe m (hommes adultes >= 16 ans) et le groupe f+e (femmes adultes + enfants < 16 ans).

3 Estimation de la Force de Voix à partir du SMLT normalisé

Dans la section précédente il a été montré que la FDV est un meilleur descripteur du SMLT que ne l'est la consigne vocale donnée au sujet. Il s'agit maintenant de déterminer dans quelle mesure la FDV peut être prédite à partir du SMLT quand on retire de ce dernier l'information de niveau sonore absolu. La procédure adoptée consiste à normaliser tous les SMLT à un même niveau arbitraire, à comparer chacun à tous les autres, sauf ceux provenant du même locuteur, et à considérer la FDV du plus proche voisin comme l'estimation recherchée. La distance utilisée est de type L2 (moyenne quadratique des différences entre les valeurs spectrales des deux SMLT normalisés). Le résultat est exprimé sous deux formes: marge statistique d'erreur à 1 écart-type (dans une distribution normale 68% des observations se trouvent à moins de 1 écart-type de la moyenne), et coefficient de corrélation entre valeurs mesurées et valeurs estimées de la FDV.

3.1 Estimation de la FDV sur les données normalisées entre 0 et 50 dB

Tous les SMLT sont recalés en amplitude à un même niveau global de 50 dB. Les valeurs spectrales faibles sont limitées par un seuillage à 0 dB. Seuls les 20 canaux compris entre 3 (80 Hz) et 22 (6,3 kHz) sont pris en compte. La raison en est que l'utilisation de ces canaux extrêmes reviendrait à identifier les SMLT des voix faibles par leur faible recul du bruit de fond, qui dépend essentiellement des conditions de prise de son et non de la voix elle-même.

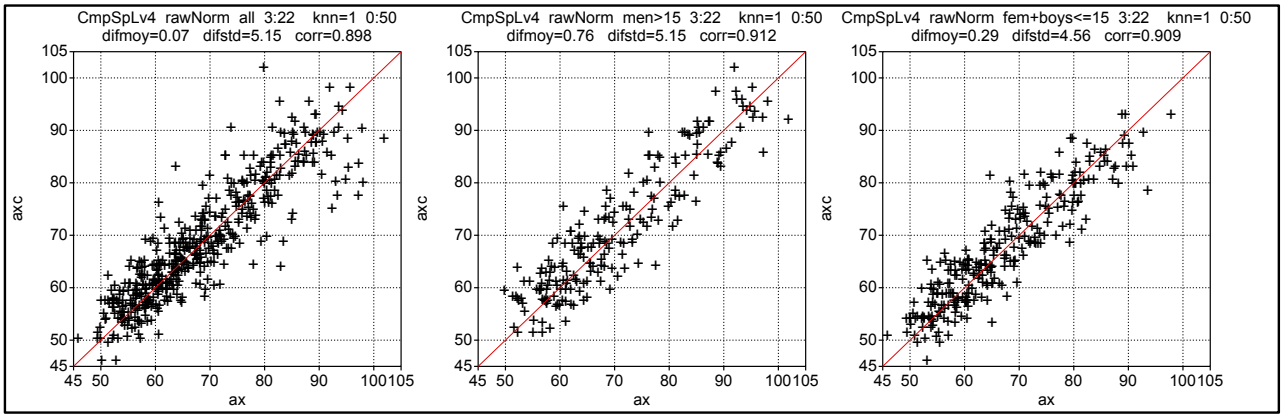

FIGURE 3: FDV estimée (axc), en fonction de la FDV mesurée (ax), pour le corpus total (fig 3a), pour les voix d'hommes adultes (3b) et pour les voix de femmes et d'enfants (3c).

Sur le corpus total (figure 3a) on observe une forte corrélation (0.898) et une marge statistique d'erreur de 5,15 dB. La distribution est mieux groupée dans la première moitié de l'échelle: les voix très fortes ou criées donnent des SMLT plus différents entre eux que les voix faibles ou moyennes. L'essentiel des erreurs grossières (à plus de 1 écart-type) concerne les voix très fortes ou criées, de FDV supérieure à 75 dB.

Les sous-corpus m (figure 3b) et f+e (figure 3c) donnent des résultats voisins, les marges d'erreur passant respectivement à 5,15 et 4,56 dB et les corrélations à 0,912 et 0,909. Les erreurs grossières sont moins nombreuses que dans le corpus total, ce qui suggère que la plupart d'entre elles sont imputables à la ressemblance des voix f+e moyennement fortes et des voix m très fortes ou criées.

Les résultats se dégradent peu quand on réduit la dynamique: la réduction à 30 dB au lieu de 50 dB augmente la marge d'erreur de 0,18 dB pour le corpus total, de 0,07 dB pour le sous-corpus m et de 0,48 dB pour le sous-corpus f+e. Il convient de rappeler que la dynamique est comptée à partir

du niveau global, lui-même supérieur d'une dizaine de dB au maximum spectral (ceci selon la forme du spectre). Entre le maximum spectral et le seuil on n'a guère plus d'une vingtaine de dB; ceci garantit que les performances obtenues reposent essentiellement sur les maxima spectraux et sont indépendantes du recul du bruit de fond des enregistrements.

3.2 Résultats avec plusieurs plus proches voisins

La quasi-symétrie des nuages de points autour de la diagonale est due au fait que l'on choisit comme valeur estimée de la FDV celle du SMLT le plus proche. L'ensemble des données n'étant pas infini on retrouve souvent en correspondance étroite la même paire de SMLT, l'un en test et l'autre en référence et réciproquement. La prise en compte de la distance moyenne des 4 plus proches voisins, pondérés selon leur classement, entraîne une amélioration globale (figure 4).

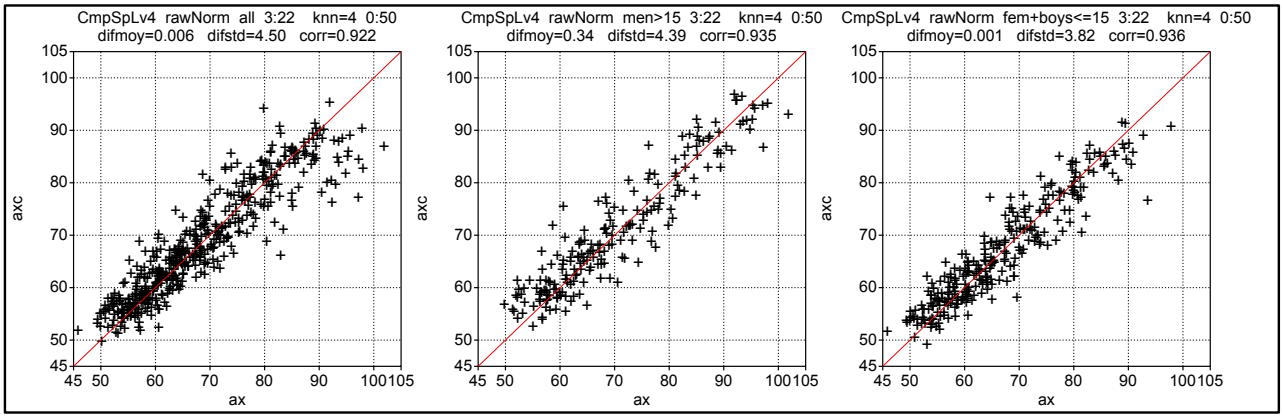

FIGURE 4: comme figure 3, mais la FDV est estimée à partir des 4 plus proches voisins.

Les marges d'erreur passent respectivement à 4,50, 4,39 et 3,82 dB et les coefficients de corrélation à 0,922, 0,935 et 0,936. Le nombre de valeurs surestimées (i.e. pour lesquelles la valeur estimée axc est très supérieure à la valeur mesurée ax) diminue et il subsiste pour le corpus total un certain nombre de valeurs sous-estimées. Globalement ces résultats valident l'hypothèse formulée plus haut, selon laquelle la FDV peut être prédite à partir de l'enveloppe du SMLT.

4 Discussion

Les données de Pearsons ne permettent pas de traiter complètement le problème d'estimation de l'EV. Il manque des nuances supplémentaires en voix faible, les mesures en mode "casual" ne portent pas sur le même texte que les autres, la prise de son à 1 m limite la dynamique des voix faibles, et surtout la non-disponibilité des enregistrements sonores empêche toute étude spectro-temporelle. Mais malgré leur ancienneté et leurs manques, ces données sont précieuses car elles constituent un matériau d'étude étalonné, indépendant du contenu phonétique, qui permet de relier directement la FDV à la forme du SMLT.

Les résultats sont encourageants, voire surprenants par leur relative précision, dans la mesure où l'information contenue dans le SMLT en 24 bandes - en fait 20 bandes dans les expériences rapportées ci-dessus - est très pauvre. Certaines erreurs sont imputables à la ressemblance spectrale entre la voix forte féminine et la voix masculine très forte ou criée, la F0d étant dans les deux cas de l'ordre de 300 Hz: s'il existe des différences significatives, celles-ci n'apparaissent pas suffisamment dans les SMLT; leur étude demanderait des données plus complètes. Il faut cependant souligner le fait que la voix criée pose des problèmes particuliers (ROSTOLLAND,

1982; JUNQUA, 1992; FUX et al., 2010), qu'elle est d'un emploi peu fréquent, et qu'il serait légitime de ne pas la considérer sur le même plan que la voix conversationnelle.

L'échelle logarithmique du SMLT en tiers d'octaves donne une grande importance aux fréquences inférieures à 1 kHz et sépare, au moins pour les FDV moyennes, le fondamental de l'harmonique 2. La différence d'amplitude entre ces deux composantes est depuis longtemps reconnue comme liée à l'EV, mais son utilisation en tant qu'indice est rendue problématique par la présence du premier formant F1 dans la même zone fréquentielle, de manière variable selon la voyelle, le genre du locuteur, et bien sûr la FDV. Le SMLT, qui intègre en 10 à 20 secondes une cinquantaine de syllabes, donne une représentation moyennée et acoustiquement stable de cette zone de fréquence, ce qui peut expliquer sa pertinence dans le problème d'estimation de la FDV.

Dans la suite il faudra confirmer les résultats et tenter de les étendre à d'autres ensembles de données, diverses langues, divers contenus phonétiques, diverses catégories de locuteurs. La limitation actuelle se trouve dans l'absence de bases de données étalonnées en termes de force de voix. La perspective, à terme, est de pouvoir affecter une valeur de la FDV à tout enregistrement de voix dont les conditions de prise de son (microphone, distance bouche-micro, gain) sont indéfinies, ce qui est le cas général.

L'estimation quantitative de la FDV concerne en particulier la recherche en acoustique phonétique. La variation de FDV produit d'importantes modifications spectro-temporelles du signal oral, qui perturbent la recherche d'invariants acoustiques associés aux éléments phonétiques et prosodiques de tous niveaux. Il faut prendre en compte la FDV, et donc pouvoir évaluer celle-ci. La même remarque vaut pour la prosodie, ainsi que pour le timbre individuel du locuteur ou les nuances liées à son expression.

Elle concerne également le traitement automatique de la parole et de la voix. Qu'il s'agisse de reconnaissance de la parole, du locuteur, ou du tour de parole, tous les systèmes se heurtent à la variabilité acoustique due à l'EV, qui fait chûter les performances. L'estimation préalable de la FDV permettrait d'en compenser les effets dans les processus d'apprentissage.

5 Conclusion

La notion qualitative d'effort vocal est une importante source de variabilité dans les sciences de la voix et de la parole. Des données de métrologie acoustique datant de 1977, réhabilitées récemment, permettent de démontrer l'intérêt du spectre moyen à long terme pour retrouver par le calcul, à moins de 5 décibels près, l'intensité acoustique émise par le locuteur, appelée Force de Voix. Dans le futur, la connaissance de cette grandeur devrait permettre d'expliquer, voire de compenser la variabilité qu'elle produit dans le signal oral. Pour progresser dans cette direction il est nécessaire de disposer de bases de données étalonnées en niveau sonore, qui malheureusement n'existent pas à ce jour.

Remerciements

Un grand merci à Brian Katz (LAM, UPMC, Paris) qui nous a signalé l'existence de l'étude de Pearsons et al., à Albert Rilliard (LIMSI) pour de nombreuses conversations fructueuses, et à Anthony Nash (Charles Salter Associates, San Francisco) qui a réhabilité les données de Pearsons et les a mises à disposition de la communauté scientifique.

Références

CUSHING I.R., LI F.F., COX T.J., WORALL K., JACKSON, T. (2011): "Vocal effort levels in anechoic conditions", Applied Acoustics, 72, 695-701.

DOVAL, B., D'ALESSANDRO, C. HENRICH, N. (2006). "The spectrum of glottal flow models", Acustica united with Acta Acustica, 92:1026-1046.

FUX, T., FENG, G., ZIMPFER, V. (2010). "Le rôle de la prosodie dans la perception de l'effort vocal", 10ème Congrès Français d'Acoustique, Lyon.

HANSON, H. (1997). "Glottal characteristics of female speakers: acoustic correlates", J. Acoust. Soc. Am. 101 (1), 466-481, 1997.

HUBER, J.E., STATHOPOULOS, E.T., CURIONE, G.M., ASH T.A. AND JOHNSON, K. (1999). "Formants of children, women, and men: the effects of vocal intensity variation", J. Acoust. Soc. Am. 106 (3), 1532-1542.

JUNQUA, J.-C. (1992). "The Lombard reflex and its role on human listeners and automatic speech recognizers", J. Acoust. Soc. Am. 93, 510-524.

LIENARD, J.S. AND DI BENEDETTO, M.G. (1999). "Effect of vocal effort on spectral properties of vowels", J. Acoust. Soc. Am. 106 (1), 411-422.

LIENARD J.S. AND BARRAS C. (2013). “Fine-grain voice strength estimation from vowel spectral cues”, InterSpeech, Lyon.

LIENARD J.S. (2014). “Etude des voyelles et de la force de voix par analyse discriminante”, JEP 2014, Le Mans.

LÖFKVIST A. (1986), "The long-time-average spectrum as a tool in voice research". Journal of Phonetics 14:472

NASH, A. (2014): "An electronic database of speech sound levels", Inter-Noise, Melbourne, 2014.

NORDENBERG M. AND SUNDBERG J. (2004), "Effect on LTAS of vocal loudness variation", Logoped Phoniatr Vocol 29, 183:191

PEARSONS K.S., BENNETT R.L., FIDELL S. (1977): "Speech levels in various noise environments", (Report No. EPA-600/1-77-025), U.S. Environmental Protection Agency, Washington DC.

RILLIARD, A., D'ALESSANDRO, C. AND EVRARD, M. (2018). " Paradigmatic variation of vowels in expressive speech: Acoustic description and dimensional analysis", J. Acoust. Soc. Am. 143, 109-122.

ROSTOLLAND, D. (1982). "Acoustic features of shouted voice", Acustica, vol 50, 118-125.

TRAUNMULLER, H. AND ERIKSSON, A. (2000). "Acoustic effects of variation in vocal effort by men, women and children", J. Acoust. Soc. Am. 107 (6), 3438-3451.