

LA VARIABILITE INTER-LOCUTEUR,  
ETUDE SUR LES REALISATIONS ACOUSTIQUES DE /e, ε/

A. Bonneau

CRIN-INRIA, Nancy, France.

ABSTRACT

This paper deals with the different sources of acoustic variability and particularly with across-speaker variation. We propose a speaker's normalization procedure based on a minimal training and implementing expert knowledge.

We test our procedure on two french vowels (/e, ε/) uttered by 13 male speakers.

INTRODUCTION

Nous présentons ici une étude des variations qui agissent simultanément sur la parole afin de déterminer leurs manifestations acoustiques et leurs influences réciproques. Nous avons choisi de nous concentrer essentiellement sur la variation inter-locuteur et sur l'influence du contexte sur celle-ci.

La variation inter-locuteur est décomposée en variation d'origine physiologique et en variation d'origine articulatoire. Nous proposons une méthode d'évaluation de la variation d'origine physiologique, fondée sur un apprentissage minimal grâce à l'apport de connaissances.

Nous avons choisi de commencer notre étude par l'analyse de deux voyelles antérieures du français, /e/ et /ε/, pour deux principales raisons: d'une part /e/ constitue une bonne voyelle d'apprentissage et d'autre part il nous semble particulièrement intéressant d'étudier deux voyelles dont l'opposition est neutralisable en français.

1. VARIABILITE INTER-LOCUTEUR

Nous tenterons d'évaluer les effets de la variation inter-locuteur d'origine physiologique et nous évoquerons les

conséquences de la variation inter-locuteur d'origine articulatoire.

1.1 Tentative d'évaluation et de normalisation de la variation d'origine physiologique

Nous désirons déterminer les répercussions acoustiques des différences physiologiques entre locuteurs ou, plus précisément, déterminer les variations des fréquences formantiques en fonction de la taille du conduit vocal.

Si on multipliait toutes les dimensions du conduit vocal par un même facteur, les fréquences formantiques caractéristiques de ce conduit vocal -abstraction faite des conséquences de la radiation aux lèvres- seraient également multipliées par un même facteur, l'inverse du premier.

Le rapport moyen entre les fréquences formantiques des locutrices et celles des locuteurs, toutes voyelles et tous formants confondus, est de 1.17 ce qui correspond approximativement à une différence de taille de 3 cm entre les deux conduits vocaux des deux sexes. Les écarts observés ne sont pas les mêmes quels que soient le formant et la voyelle considérés. Ainsi selon les calculs de Peterson et Barney [1]; soit Fh et Ff représentant respectivement les moyennes masculines et féminines:

Ff/Fh est égal à 1.01 pour le premier formant de /u/ et 1.23 pour le troisième formant de cette même voyelle.

La principale explication de ce phénomène est l'existence de différences physiologiques entre les configurations vocales des hommes et celles des femmes. En effet la longueur du pharynx (Fant [2]) et l'ouverture relative au point de constriction maximale (Traünmüller [3]) sont relativement plus grandes chez les hommes.

Une évaluation des variations fréquentielles liées aux différences physiologiques nécessiterait la réponse à la question suivante: comment s'effectue

le passage d'une configuration vocale masculine typique à une configuration vocale féminine typique? S'effectue-t-il de manière continue, en fonction de la taille du conduit vocal, que le locuteur soit féminin ou masculin? Ou de manière discontinue, avec une rupture du continuum à la mue, causée par la baisse du larynx chez les hommes (Traünmüller [3])? Cette dernière solution nous apparaît la plus plausible, mais nous analyserons néanmoins les deux éventualités. Remarquons auparavant qu'au-delà de cette interrogation d'ordre physiologique, c'est également le problème de la portée de la normalisation qui est posé ici: s'appliquera-t-elle à tous les locuteurs, ou doit-on normaliser séparément les voyelles des hommes et celles des femmes?

1.1.1 Normalisation formantique conjointe des locuteurs et des locutrices

Nous faisons l'hypothèse que la longueur relative du pharynx et l'ouverture relative au point de constriction maximale dépendent uniquement de la taille du conduit vocal. Supposons que les fréquences formantiques varient linéairement en fonction de cette taille et que les différences physiologiques constituent l'unique source de variation fréquentielle. Cette variation peut être simplement évaluée par la position relative des fréquences formantiques d'un locuteur par rapport aux fréquences moyennes des hommes et des femmes. Cette position relative est identique pour tous les formants de toutes les voyelles d'un locuteur puisqu'elle indique, selon nos hypothèses, la taille de son conduit vocal.

A partir d'une seule fréquence formantique d'une seule voyelle d'un locuteur, on pourrait donc prédire les fréquences formantiques de toutes les voyelles de ce locuteur.

Mais d'autres facteurs de variations fréquentielles sont à prendre en considération, comme l'articulation spécifique à chaque locuteur, que nous ignorons et dont les conséquences acoustiques sont plus difficiles à évaluer. Nous laissons de côté pour l'instant les variations dues à la vitesse d'élocution.

Etant données nos précédentes hypothèses (continuum des configurations vocales, linéarité des variations fréquentielles), les représentations acoustiques de chaque

voyelle V dans le plan ou l'espace formantique se répartissent, sous l'effet des différences physiologiques, le long d'une droite définie par les moyennes masculines et féminines de V -appelons chacune des droites ainsi définies, il en existe une par voyelle, droite "physiologique"-.

La méthode la plus simple pour évaluer l'effet des différences physiologiques en dépit des conséquences acoustiques des autres sources de variation consiste à effectuer une projection perpendiculaire d'une image acoustique d'une voyelle V sur la droite "physiologique" de V et à se reporter au critère de la position relative, en le modifiant légèrement puisque deux formants au moins sont désormais nécessaires pour notre évaluation. Cette méthode, proposée par F Lonchamp (Bonneau [4]), suppose que l'effet des variations articulatoires est négligeable le long des droites "physiologiques". La méthode d'évaluation exposée ici nous permet de proposer une procédure de normalisation des locuteurs féminins et masculins fondée sur un apprentissage minimal.

A partir des fréquences formantiques d'une voyelle V prononcée par un locuteur l ainsi que des fréquences formantiques moyennes des hommes et des femmes pour cette même voyelle V, nous déterminons un paramètre de normalisation, appelons-le "dist", qui tient compte implicitement de la taille du conduit vocal de l. Détaillons la procédure:

- nous enregistrons trois répétitions de /ε/ en contexte labial, émises par un locuteur

l. En théorie, une seule répétition de /ε/ suffit mais trois répétitions au moins sont nécessaires afin de minimiser les erreurs dues à une prononciation déviante. /ε/ nous semble être une bonne voyelle d'apprentissage (Bonneau [4])

- nous projetons l'image acoustique de /ε/ -la moyenne des trois répétitions- sur la droite qui relie les moyennes masculines

(Fhε) et féminines (Ffε) de /ε/.

- nous calculons le paramètre de normalisation physiologique "dist" qui indique la position relative de notre projection par rapport à Ffε et Fhε.

- toutes les voyelles de /ε/ émises par l sont normalisées par l'application de la

formule:  
soit  $F_{lij}$  représentant la fréquence du ième formant de la voyelle  $j$  prononcée par le locuteur  $l$  et  $F_{lij}$  sa normalisation,  
 $F_{lij} = F_{lij} - \text{dist} * F_{hij} F_{fij}$ .

Notre normalisation consiste donc à déplacer chaque représentation acoustique d'une voyelle quelconque parallèlement à la droite "physiologique" qui correspond à cette voyelle.

### 1.1.2 Normalisation formantique séparée des locuteurs et des locutrices

Une normalisation commune des locuteurs et des locutrices n'est plus possible. Si on se reporte au cas simple évoqué plus haut - c.à.d pour une configuration vocale donnée, toutes les dimensions du conduit vocal et les fréquences formantiques sont multipliées par un même facteur quand la taille du conduit vocal varie-, on peut évaluer simplement les variations fréquentielles des locuteurs masculins à partir des moyennes formantiques masculines, et celles des locutrices à partir des moyennes formantiques féminines.

Afin de limiter les erreurs d'évaluation causées par l'articulation spécifique à chaque locuteur, nous proposons à nouveau une procédure d'évaluation semblable à la précédente, c.à.d qui comporte une projection des données d'apprentissage sur une droite "physiologique". Deux droites "physiologiques" sont ici nécessaires par

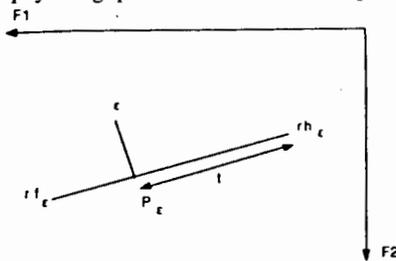


Figure 1

Détermination du paramètre de normalisation  $t$ ,

$rh_{\epsilon}$  moyenne masculine de  $\epsilon$

$r_{\epsilon}$  moyenne féminine de  $\epsilon$

$\epsilon$ : représentation acoustique de cette voyelle pour le locuteur sur lequel on effectue la normalisation,  $P_{\epsilon}$  sa projection sur la droite des moyennes.

voyelle, une pour les références féminines et une pour les références masculines.

Il se peut que les différences physiologiques n'aient pas des conséquences aussi triviales et qu'il faille réévaluer celles-ci. Si nous conservons l'hypothèse simple d'une variation linéaire des fréquences formantiques en fonction de la taille du conduit vocal, nous devons proposer de meilleures droites "physiologiques". Cette tâche soulève quelques problèmes que nous n'avons pas la place d'évoquer ici.

### 1.2 Sources d'erreurs

Ce qui précède est une version simplifiée des conséquences probables de la variation inter-locuteur.

Citons quelques phénomènes qui peuvent perturber, selon leur ampleur, la bonne estimation de nos paramètres de normalisation:

- la compensation articulatoire qui affecte notamment l'ouverture relative du conduit vocal,

- l'influence très forte de certains contextes, par exemple l'influence des dentales sur l'articulation de /u/, qui peut remettre en cause la validité des droites "physiologiques",

- les variations fréquentielles entre locuteurs d'origine articulatoire, si elles ne sont pas négligeables le long de nos droites physiologiques, et qui sont d'autant plus difficiles à cerner qu'elles peuvent changer pour un même locuteur avec le contexte.

Le bien-fondé des procédures de normalisation de la variation inter-locuteur d'origine physiologique est très délicat à établir puisque nous ne connaissons ni la taille réelle du conduit vocal du locuteur ni son articulation spécifique. Que normalisons-nous réellement, la variation d'origine physiologique, articulatoire ou une partie des deux?

## 2. METHODOLOGIE

Le corpus d'évaluation est constitué de deux phrases qui comportent les réalisations des deux voyelles / $\epsilon$ ,  $e$ /, une voyelle par phrase, en syllabe accentuée et dans des contextes consonantiques symétriques : labial, dental, palatal et uvulaire. Voici ces phrases, nous avons séparé les voyelles étudiées par un espace.

"A Papeete, son fr è re a mangé cette f è ve faite en li è ge"

"Vous ser e z raccord é si vous pouv e z pay e r chaque mois"

/ $\epsilon$  / et /  $e$  / apparaissent dans des contextes où leur prononciation est bien déterminée en français, de ce fait ces contextes ne sont pas tout-à-fait

comparables: syllabe fermée pour / $\epsilon$  / et syllabe ouverte suivie d'une frontière morphologique pour /  $e$  /. Treize locuteurs masculins ont enregistré le corpus, que nous espérons compléter par les données d'autres locuteurs et surtout d'autres locutrices pour la présentation de ce papier au congrès. Le signal a été échantillonné à 16 Khz, et les fréquences formantiques ont été mesurées manuellement sur grand écran.

## 3. RESULTATS

Signalons d'abord que l'opposition /  $e$ ,  $\epsilon$  / est respectée puisque les réalisations acoustiques de ces deux voyelles sont bien distinctes dans un contexte donné. D' autre part l'effet de chaque contexte consonantique est parfaitement prévisible quel que soit le locuteur.

Nous avons testé deux procédures de normalisation:

- la normalisation conjointe des locuteurs et des locutrices,

- la normalisation séparée des locuteurs et des locutrices à partir des moyennes masculines et féminines.

Voici les résultats obtenus avec la première méthode, nous avons mis entre parenthèses les résultats obtenus avec la deuxième méthode quand ils sont différents des précédents. Trois formants ont été utilisés pour l'apprentissage.

- 22% (30%), 27% (33%), 70% (76%) pour F1, F2, F3 de / $\epsilon$ /,

- 5%, 51%, 66% pour F1, F2, F3 de / $e$ /.

Nous ne constatons pas de grandes différences entre les méthodes qui normalisent les voyelles des hommes et les femmes ensemble ou séparément, à une exception près: le premier formant de

/  $\epsilon$  / en contexte palatal, mieux normalisé par la deuxième méthode; pour /  $e$  /, les droites "physiologiques" sont sensiblement identiques pour les deux méthodes-.

Il sera intéressant de vérifier ces résultats sur les formants des voyelles d'arrière

ouvertes.

L'apprentissage effectué avec la voyelle /

$\epsilon$  / se révèle aussi performant pour la normalisation de /  $e$  / que l'apprentissage effectué avec /  $e$  / même. Là encore, il sera intéressant de confirmer ces résultats avec d'autres voyelles.

Les résultats obtenus par voyelle et par formant semblent satisfaisants si on considère que seules des voyelles émises par des locuteurs masculins ont pour l'instant été normalisées. L'emploi de barks à la place des Hertz améliore légèrement les résultats obtenus pour F1 en rééquilibrant pour chaque formant les écarts fréquentiels entre les moyennes ou références acoustiques.

L'articulation spécifique à chaque locuteur varie avec le contexte et s'exprime le long de nos droites "physiologiques" ce qui perturbe les

résultats obtenus pour F2 /  $\epsilon$  /. Il est difficile de commenter les résultats obtenus pour F1 /  $e$  /, la variation inter-locuteur étant relativement faible avant la normalisation.

## 4. REFERENCES

- [1] PETERSON, G.E. and BARNEY, H.L. (1952), "Control Methods Used in a Study of the Vowels", JASA, vol 24, pp175-184.
- [2] FANT, G. (1973), "A note on vocal tract size factors and nonuniform F-Pattern scalings", Speech Sounds and Features, MIT Cambridge, USA, pp 84-93.
- [3] TRAUNMÜLLER, H. (1984), "Articulatory and perceptual factors controlling the age- and sex-conditioned variability in formant frequencies of vowels", Speech Communication, vol 3, numéro 1, pp 49-62.
- [4] BONNEAU, A., FOHR, D., LONCHAMP, F. (1989), "Normalisation formantique des locuteurs féminins et masculins à l'aide de connaissances et d'un apprentissage minimal" Actes du séminaire sur la variabilité du locuteur, Luminy, juin 1989.