

Ouverture de la glotte, Fo, intensité et simulations émotionnelles : le cas de la joie, la colère, la surprise, la tristesse et la neutralité.

Gendrot Cédric

Université Paris 3- La Sorbonne Nouvelle
Institut de Linguistique et de Générales Appliquées
19, rue des Bernardins - 75005 PARIS
CNRS UMR 7018
Tél.: ++33(0) 143 26-37 80
cgendrot@hotmail.com

Abstract

Voice open quotients have been measured and then compared to Fo and intensity in simulated emotional speech. We used electro-glottography on 3 male actors producing happiness, anger, sadness, surprise and neutrality from a specific corpus. Stimuli that were correctly identified in a perception test were selected and analysed, we found that :

Measurements of open quotient could significantly distinguish between 4 classes: happiness, anger, sadness and neutrality ($F=217.116$ and $p<0.0001$ for the emotion factor).

Open quotient significantly increased with intensity and Fo for anger, happiness and neutrality.

1. Introduction

Le but de notre approche est avant tout de mieux comprendre comment les caractéristiques acoustiques de la parole émotionnelle sont déterminées par des modifications physiologiques.

De nombreuses études sur les émotions se sont intéressées aux paramètres prosodiques classiques que sont la fréquence fondamentale (Fo), la durée et l'intensité. Longtemps, les chercheurs ont cru pouvoir caractériser les émotions grâce à une combinaison de ces trois facteurs, mais ont du se rendre à l'évidence que la qualité de la voix est un facteur non négligeable et un domaine qu'il faudrait approfondir [Fón83a; Sch89; Mon98]. Klatt et Klatt (1989) ont tenté de vérifier l'influence des paramètres de la source sur la qualité vocale en se basant sur l'observation de spectres ; cependant ces procédés sont toujours sujets à controverse.

La joie et la colère par exemple sont souvent caractérisées par des paramètres prosodiques similaires [Sch86]. Nous avons choisi d'aborder un aspect physiologique en analysant le temps ouvert relatif des cordes vocales pendant la phonation, le larynx étant grandement responsable des caractéristiques de la qualité de la voix [Lav80]. Nous avons eu recours à l'électro-glottographie pour vérifier l'hypothèse selon laquelle la joie et la colère peuvent être distinguées à

ce niveau, alors qu'elles sont confondues du point de vue des paramètres acoustiques.

2. Choix préliminaire du corpus

Fónagy [Fón71] notait que la perception de l'émotion était en partie dépendante du contenu sémantique de l'énoncé. Pour vérifier si le contenu sémantique de nos phrases pouvait influencer nos résultats, nous avons séparé notre corpus en plusieurs classes distinctes [Mur93] :

Type A : Les phrases émotionnellement indéterminées où plusieurs émotions pourraient être attribuées à la phrase après lecture, mais aucune émotion précise ne se dégage de manière significative.

par exemple : « c'est pas ce que je pensais »

Type B : Les phrases sémantiquement neutres pour lesquelles le sentiment dominant est celui de la neutralité, l'absence d'émotion véritable.

par exemple : « la bouteille est sur la table ».

Type C : Les phrases émotionnellement chargées : phrases pour lesquelles une émotion se dégage de manière significative à la lecture. Nous avons aussi choisi pour cette étude des phrases sémantiquement orientées vers nos quatre émotions.

par exemple colère : « je vais détruire ta maison »
joie : « j'ai gagné une voiture neuve »
tristesse : « j'ai perdu tout mon argent »
surprise : « je ne m'y attendais pas »

3. Protocole expérimental

3.a. Evaluation écrite du corpus

Nous nous sommes donc inspiré de ces 3 distinctions pour élaborer notre corpus avant sa première évaluation. 35 phrases heptasyllabiques, qui semblaient pouvoir être réparties dans 6 groupes, ont été choisies.

pour le type A: indéterminé

pour le type B: neutre

pour le type C: joie, colère, tristesse et surprise

Après une lecture silencieuse, 15 évaluateurs devaient alors décider pour chacune des phrases, lequel de ces 6 groupes

était le plus approprié : ce test à choix forcés représentant certaines limites notées dans la littérature [Dav64; Léo71], nous avons donc choisi de l'améliorer grâce à l'ajout de trois échelles bipolaires.

Ces échelles représentant les théories dimensionnelles de l'émotion furent élaborées par Osgood [Osg57]. Valence (agréable/désagréable), puissance (faible/fort) et activation (inactivité/tension) sont représentées sur trois axes gradués de -3 à +3. Les résultats obtenus par ces échelles tout au long de cette étude nous ont permis de corroborer et parfois de préciser ceux des tests à choix forcés.

Les 6 groupes ont été constitués selon le raisonnement suivant : une phrase ayant statistiquement (test dit du Chi2) recueilli un nombre majoritaire d'un seul choix sur les six possibles (joie, colère, tristesse, surprise, neutre et indéterminé) pouvait être retenue.

18 phrases ont ainsi été retenues pour le corpus que nous avons présenté à des acteurs :

Type A : 1 groupe de 3 phrases pour les variantes émotionnellement indéterminées

Type B : 1 groupe de 3 phrases pour les variantes émotionnellement neutres

Type C : 4 groupes de 3 phrases, 1 pour chaque émotion (joie, colère, tristesse et surprise).

2.b. Obtention des stimuli

Nous avons demandé à trois acteurs amateurs de simuler quatre émotions (joie, colère, tristesse et surprise) à partir de notre corpus précédemment choisi. Chaque phrase a également été lue par le locuteur de manière aussi neutre que possible. Pour ne pas créer de stimuli contradictoires en associant des émotions et des contenus sémantiques différents, les phrases émotionnellement chargées (de type C) ont seulement été produites avec l'émotion correspondant à leur contenu sémantique.

Nous nous sommes assuré de la validité émotionnelle de nos stimuli grâce à un test de perception. Pour ce test, 15 sujets ont dû identifier les émotions simulées par les acteurs.

émotion simulée	émotion perçue				
	joie	colère	tristesse	surprise	neutre
joie	58	3	12	19	8
colère	0	98	0	1	1
tristesse	0	0	76	11	13
surprise	22	18	7	40	13
neutre	0	6	16	2	76

tableau 1 : Matrice de confusion pour les simulations d'émotions (en pourcentages).

Seules les simulations identifiées à plus de 75% ont été retenues pour la suite de notre étude.

2.c. l'étude laryngographique :

Des recherches utilisant l'électro-glottographie ont suggéré que le quotient ouvert varie de manière significative en fonction de la qualité serrée, tendue (Oq petit), ou détendue (Oq élevé) de la voix.

Dans cette étude, le quotient Oq est défini comme le temps ouvert relatif, c'est à dire le rapport de la période ouverte de la glotte sur la période de voisement [Chi84 et Hen00]. Les prises de données laryngographiques ont été faites simultanément avec les enregistrements acoustiques. Nos trois acteurs portaient un collier positionnant deux électrodes placées à l'aide de l'expérimentateur. En dérivant le signal EGG (la formule de dérivée était la suivante : $dX(t)/dt = [x(i) - x(i-1)]/DT$), on obtient pour chaque période deux pics assez nets, de sens opposé. Le pic le plus marqué correspond au moment de fermeture glottique, tandis que l'autre pic peut être relié à l'instant d'ouverture glottique. La détermination de ces deux instants permet alors de connaître la valeur du quotient ouvert (Chi84 et Fou93). En effet, Oq sera égal au rapport d1/d2 (figure 1). Notons que cette méthode de mesure du quotient ouvert connaît certaines limites du fait que l'on observe parfois une indétermination du pic associé à l'ouverture glottique, ou bien un double pic associé à la fermeture des cordes vocales.

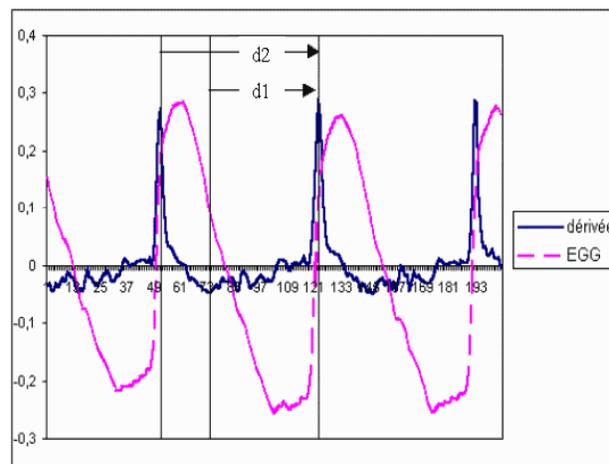


Figure 1 : Graphe de la courbe EGG et de sa dérivée (les échelles ont été modifiées pour une meilleure visibilité). Oq est calculé selon la formule : $Oq = d1/d2$

3. Calculs des quotients ouverts.

Les mesures de quotients ouverts fourniront-elles des indices pertinents, c'est à dire des valeurs moyennes

statistiquement différentes selon la simulation émotionnelle? Les variations seront elles liées aux variations de Fo ou d'intensité ?

3.a. Valeurs moyennes de Oq

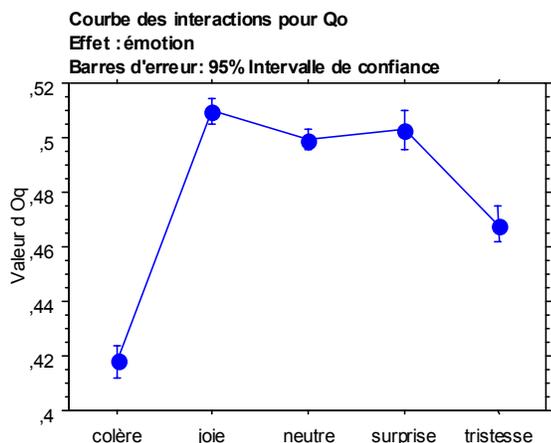


Figure 2: Valeurs moyennes du quotient ouvert Oq pour chaque émotion.

Des Anovas à deux facteurs ont permis d'établir des différences significatives entre les quotients ouverts correspondant à la joie, la colère, la tristesse et la neutralité (F=217.1 et p<0.0001 pour le facteur émotion).

Test PLSD de Fisher pour Qo			
Effet : émotion			
Niveau de significativité : 5 %			
	Diff. moy.	Diff. crit.	Valeur p
colère, joie	-,092	,007	<,0001
colère, neutre	-,082	,007	<,0001
colère, surprise	-,085	,009	<,0001
colère, tristesse	-,050	,008	<,0001
joie, neutre	,010	,007	,0021
joie, surprise	,007	,009	,1178
joie, tristesse	,042	,008	<,0001
neutre, surprise	-,003	,009	,4758
neutre, tristesse	,031	,008	<,0001
surprise, tristesse	,035	,010	<,0001

Tableau 2: Test PLSD de Fisher effet « émotion » (quotients ouverts).

Nous avons comparé les valeurs moyennes de Oq pour chacune de nos quatre émotions à celles des phrases neutres.

L'écart de Oq le plus fort est observé pour la paire colère / neutre (-0,082 pour p<0.0001). L'écart de Oq calculé pour la paire neutre / surprise est le plus faible et non significatif (-0,003 pour p=0.475). Quant aux paires joie / neutre et neutre / tristesse, les écarts observés sont intermédiaires mais significativement différents (0,01 et 0,031 pour p=0.002 et p<0.0001 respectivement). Notons que ce paramètre permet donc une distinction très nette entre la colère et la joie (-0.092 pour p<0.0001 alors qu'une étude acoustique ne révèle pas toujours de différences significatives).

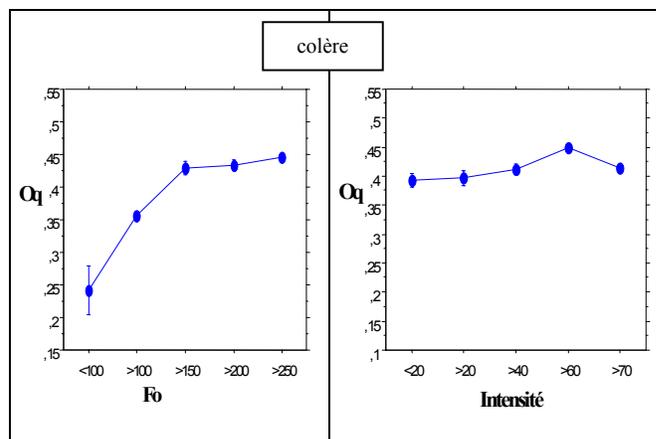
L'influence des variations sémantiques de notre corpus n'a pu être validée pour cette étude. Nous suggérons que l'utilisation de courtes simulations a pu fausser ce type de résultats, ces derniers étant sans doute très dépendants de la situation de communication. Quelques tendances ont pu être remarquées quant à des variabilités inter-locuteurs, il reste difficile d'en tirer des conclusions de par l'utilisation d'un faible nombre de locuteurs.

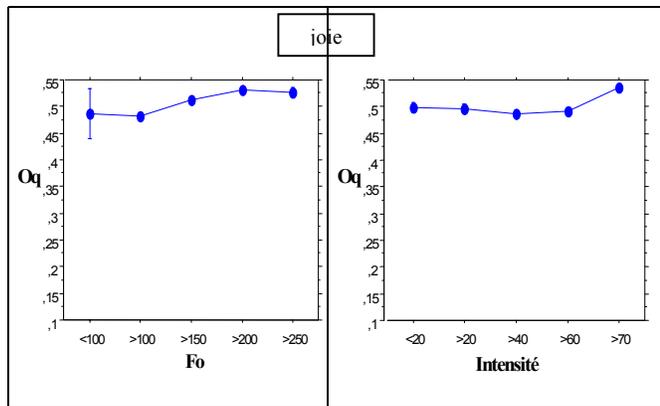
3.b. Variation des valeurs de Oq en fonction des contours de Fo et d'intensité.

Nous avons tenté de vérifier l'effet des deux variables acoustiques, Fo et intensité, sur nos valeurs de quotients ouverts calculées en fonction de chaque émotion.

Pour ce faire, nous avons obtenu pour chaque phrase, grâce au logiciel Praat [Boe99], des courbes de Fo et d'intensité que nous avons ainsi pu comparer aux courbes de quotients ouverts calculées préalablement.

Cet effet s'est révélé significatif pour la colère, la joie et la voix neutre. Globalement les valeurs de Oq augmentent avec celles de la Fo et de l'intensité (figures 3 à 6). On remarque que l'effet est le plus significatif pour les deux émotions considérées à forte activation que sont la colère et la joie (F=40.9 et 17.3 respectivement pour la Fo ; F=7.8 et 8.6 pour l'intensité, p<0.0001).





Figures 3, 4, 5 et 6 : Effet de la Fo en Herz (à gauche) et de l'intensité relative en dB (à droite) sur le quotient ouvert calculé en fonction de l'émotion.

4. Conclusion

Nous avons montré dans cette étude que les quotients ouverts calculés à partir de courbes électro-glottographiques se sont révélés être significativement différents pour la joie, la colère, la tristesse et la neutralité. Ils permettent de distinguer, par exemple entre la joie et la colère alors que les paramètres acoustiques Fo, durée et intensité ne permettaient pas toujours une distinction aisée. D'autres paramètres acoustiques, tel que la pente spectrale, qui est en relation avec les mouvements de la glotte, vont être étudiées.

Nous avons pu mettre en évidence une relation entre deux facteurs prosodiques (Fo et intensité) et nos valeurs de quotients ouverts pour la joie, la colère et la neutralité. Ces analyses doivent être approfondies avec un plus grand nombre de locuteurs afin de déterminer précisément une éventuelle influence de la focalisation sur la variation des valeurs de quotients ouverts.

Bibliographie

[Boe99] Boersma, P. and Weenink, D. Praat, a system for doing phonetics by computer. Institute of Phonetic Sciences of the University of Amsterdam, report 132-182.

[Chi84] Childers D.G., Smith A.M., Moore G.P. (1984), "Relationships between Electroglotto-graph, Speech, and Vocal Cord Contact", *Folia Phoniatr.*, 36, 105-118.

[Dav64] Davitz, J. R. (1964), Personality, Perceptual and Cognitive Correlates of Emotional Sensitivity, dans Davitz, J. R. (éd.), *The communication of Emotional Meaning*, 57-68, McGraw-Hill, New York.

[Fon71] Fónagy, I. (1971a), Double coding in speech, *Semiotica* 3, 189-222.

[Fou93] Fourcin, A.J. (1993), "Normal and pathological speech: phonetic, acoustic and laryngographic aspects". In Singh W., Soutar D. (eds.), *Functional surgery of the larynx and pharynx*, Butterworth Heinemann, London.

[Hen00] Henrich, N., d'Alessandro, C., Castellengo, M. (2000). Mesures électroglottographiques du quotient d'ouverture glottique en voix parlée et chantée, XXIIIèmes Journées d'Etude sur la Parole, Aussois, 19-23 juin 2000.

[Kla89] Klatt, D. and Klatt, L. (1989), Analysis, synthesis, and perception of voice quality variations among female and male talkers, *J. Acoust. Soc. Am.* 87, (2).

[Lav80] Laver, J. (1980), *The phonetic description of voice quality*, Cambridge University Press.

[Leo71] Léon, P. R. (1971), *Essais de Phonestylistique*, Didier, Paris, Montréal, Bruxelles.

[Mon99] Montero, J., Gutierrez-Arriola, J., Colas, J. & Pardo, J. M., (1999), Analysis and modelling of emotional speech in Spanish, *Proceedings of ICPhS*, San Francisco, U.S.A.

[Mur93] Murray, I. R. & Arnott, J. L. (1993), Toward the simulation of emotion in synthetic speech : a review of the literature on human vocal emotion, *J. Acoust. Soc. Am.* 93(2), 1097-1108.

[Osg57] Osgood, C. E., Suci, G. J. & Tannenbaum, P. H. (1957), *The measurement of meaning*, University of Illinois Press.

[Sch86] Scherer, K. R. (1986), Vocal Affect Expression: A Review and a Model for Future Research, *Psychological Bulletin* 99(2), 143-165.

[Sch89] Scherer, K. R. (1989). Vocal measurement of emotion. In R. Plutchik, & H. Kellerman (Eds.), *Emotion: Theory, research, and experience. Vol. 4. The measurement of emotion.*(pp.233-260) New York: Academic Press.