

Rôle de la qualité de la voix dans la simulation des émotions : une étude perceptive et physiologique

Gendrot Cédric

Laboratoire de Phonétique et de Phonologie CNRS UMR 7018

I.L.P.G.A. Paris III - 19, rue des Bernardins - 75005 PARIS

cgendrot@univ-paris3.fr

Mots-clés : simulations émotionnelles, qualité de la voix, quotient d'ouverture, prosodie, stratégies.

Key words : Simulated emotions, voice quality, open quotient, prosody, strategies

Résumé

La distribution spectrale et les variations de qualité de voix sont étudiées quant au rôle qu'elles apportent dans la perception de quatre émotions simulées: surprise, colère, joie, et tristesse . Les résultats de re-synthèse de plusieurs paramètres prosodiques (Fo, durée et intensité) suggèrent que l'ampleur de la contribution de la qualité de la voix varie en fonction de l'émotion simulée. La qualité de voix aurait une contribution minimale pour la surprise. Au contraire, elle permet une bonne identification de la colère. La combinaison de la qualité de voix et des autres facteurs prosodiques (Fo, durée) semble propice à une bonne perception de la joie et la tristesse.

Une étude électro-glottographique a permis d'étudier directement le comportement des cordes vocales durant la production de ces quatre types d'émotion. Elle a permis de révéler des différences significatives concernant le quotient d'ouverture (Oq) en fonction de l'émotion et notamment une distinction très nette entre la joie et la colère, que des études se basant sur des paramètres acoustiques peinaient à différencier.

Abstract

Voice quality has been investigated in this study on the role it plays in the perception of four simulated emotions: happiness, anger, sadness and surprise. The modelling of the following parameters: fundamental frequency, duration and intensity allowed us to consider the contribution of voice quality as varying according to specific emotions. It seems crucial in conveying anger while almost unimportant in surprise. Intermediate results have been found concerning happiness and sadness.

We have also intended to study voice source quality parameters by the means of the Electro-GlottoGraph. The results showed significant differences in open quotients and revealed the greatest gap between anger and happiness, two emotions reckoned as being difficult to distinguish acoustically.

Introduction

De nombreuses études sur les émotions se sont intéressées aux paramètres prosodiques classiques que sont la fréquence fondamentale, la durée et l'intensité. Longtemps, les chercheurs ont cru pouvoir caractériser les

émotions grâce à une combinaison de ces trois facteurs, mais ont dû se rendre à l'évidence que la qualité de la voix est un facteur non négligeable et un domaine qu'il faudrait approfondir (Fónagy, 1983a ; Scherer, 1989).

Depuis Laver (1991), il est devenu commun de décrire la qualité de voix en utilisant le terme " setting " : " There is an alternative, wider approach to the task of articulatory description, that concerns itself with both differences and similarities in vocal performance in speech and sees individual segments as momentary actions superimposed on a long term SETTING of the vocal apparatus. The setting accounts for the similarities and the segments for the differences, as it were. A setting gives a background, auditory colouring to sequences of short-term segmental articulations. "

En considérant le modèle source-filtre de production de la parole de Fant (1960), la génération d'une source audible d'énergie acoustique (à partir d'un flux d'air) est susceptible d'être modifiée par les mouvements articulatoires du reste de l'appareil vocal. Il est ainsi aisé de comprendre la distinction entre les trois cas suivants : (1) la production du même phonème avec la même articulation peut être perçue avec une coloration de voix différente, à cause d'une différence dans la source (tension et mode de vibration des cordes vocales) – il s'agit là d'une différence de phonation, (2) une articulation différente (labialisation, pharyngalisation, ou encore crispation des articulateurs supraglottiques, etc.) peut modifier les caractéristiques du filtre (avec un signal de source inchangé) – il s'agit là d'une différence de " timbre ", ou encore (3) une combinaison entre une différence de phonation et de timbre. Notons que la notion de " qualité de voix " réfère en général à la source (phonation), mais que le terme de " timbre " est ambigu et ne sera par conséquent pas employé ici.

Montero, Gutierrez-Arriola, Colas et Pardo (1999) suggèrent ainsi que la qualité de la voix est un facteur plus important que les autres paramètres prosodiques pour certaines émotions. Ils considèrent des émotions telles que la joie et la colère de " segmentales " puisque leur modélisation à partir de paramètres prosodiques est particulièrement peu naturelle.

Un problème récurrent mentionné dans la littérature reste la similarité entre les paramètres acoustiques conventionnels (Fo, durée et intensité) entrant en jeu dans l'expression de certaines paires d'émotions, comme notamment la joie et de la colère (Scherer, 1986). La qualité de la voix serait-elle, dès lors, un facteur qui permette de différencier de manière certaine ces deux émotions ? C'est précisément le rôle de la qualité de voix, et sa contribution à la perception de l'émotion dans la parole, qui nous intéresse dans cette étude. Cependant, l'analyse, la synthèse et la reconnaissance des émotions sont encore freinées par le manque d'avancement technologique concernant la synthèse de la qualité de voix : ainsi, les systèmes de synthèse et de reconnaissance vocale n'intègrent pas la variation stylistique et émotionnelle de la parole (Scherer, 1998).

Dans cet article, nous tentons de quantifier la contribution de la qualité de la voix à la perception de l'émotion. Puisqu'il est difficile de synthétiser et de moduler la qualité de la voix, nous avons choisi de l'isoler en modulant les facteurs prosodiques classiques que sont la fréquence fondamentale, la durée et l'intensité. Ce raisonnement nous permet ainsi de suggérer l'influence de la qualité de la voix par rapport à l'influence de ces trois paramètres prosodiques.

Par ailleurs, l'apport de données physiologiques nous semble être un complément indispensable à toute étude sur les émotions, fournissant ici une preuve directe d'un comportement différent des cordes vocales. C'est pourquoi, dans la deuxième partie de cette recherche, nous recourons à des courbes laryngographiques (EGG) décrivant précisément le comportement des cordes vocales pendant la phonation.

I. Méthode et recueil des données

1.1. Le corpus

Fónagy (1971a) souligne que l'identification d'une émotion est en partie dépendante du niveau linguistique. Il nous paraît donc intéressant de tester l'hypothèse selon laquelle la contribution de la qualité de la voix dans l'expression d'une émotion peut varier en fonction de la contribution lexicale à la phrase émotionnelle. Il nous fallait regrouper les phrases de notre corpus en plusieurs classes distinctes. Pour élaborer notre corpus, nous nous sommes basés sur des groupements proposés, notamment, par Murray & Arnott (1993) :

Type A : les phrases émotionnellement indéterminées du type “ ce n'est pas ce que je pensais ”, où plusieurs émotions pourraient être attribuées à la phrase par sa simple lecture, mais aucune émotion précise ne se dégage de manière significative.

Type B : les phrases sémantiquement neutres pour lesquelles le sentiment dominant est celui de la neutralité, l'absence d'émotion véritable (“ la bouteille est sur la table ”).

Type C : les phrases émotionnellement chargées pour lesquelles une émotion se dégage de manière significative à la lecture. Nous avons ainsi choisi pour cette étude des phrases sémantiquement orientées vers quatre émotions : la colère (“ je vais détruire ta maison ”), la joie (“ j'ai gagné une voiture neuve ”), la tristesse (“ j'ai perdu tout mon argent ”) et la surprise (“ je ne m'y attendais pas ”).

Trente-cinq phrases heptasyllabiques ont été sélectionnées sur la base des six catégories ainsi déterminées:

1. émotionnellement indéterminées
2. neutres
3. émotionnellement chargées
 - a. colère
 - b. joie
 - c. tristesse
 - d. surprise

Quinze personnes ont ensuite été invitées à effectuer une lecture des 35 phrases, présentées dans un ordre aléatoire, et ce, afin de déterminer à laquelle des six catégories chacune de ces propositions correspondait le mieux. Ce test perceptif présentant, comme toute épreuve de ce type, certaines limites fort bien décrites dans la littérature (Davitz, 1964 ; Léon, 1971), nous avons choisi de renforcer sa validité par la passation d'un test écrit basé sur trois échelles bipolaires représentant les théories dimensionnelles de l'émotion élaborées par Osgood en 1957, où la valence -agréable/désagréable-, la puissance - faible/fort- et l'activation -inactivité/tension- sont représentées sur trois axes gradués de -3 à +3. Le but de ce test écrit est de déterminer dans quelle mesure la sémantique de l'énoncé peut contribuer à l'expression émotionnelle dans notre corpus.

Nous avons réalisé pour chacune de ces phrases un test statistique (chi-carré) qui s'applique à des valeurs non paramétriques représentées par le nombre de réponses pour cette expérience. Il consiste à déterminer si la distribution des réponses obtenues dans notre test est, ou non, homogène.

Les groupes de phrases “ sémantiquement neutres ” et “ émotionnellement chargées ” ont été constitués selon le raisonnement suivant : une phrase ayant statistiquement recueilli un nombre majoritaire d'un seul choix sur les cinq possibles (joie, colère, tristesse, surprise ou neutre) pouvait être retenue. Le groupe de phrases “ émotionnellement indéterminées ” a été sélectionné pour son manque de significativité.

À la suite de cette procédure, 18 phrases ont ainsi été retenues pour la suite de notre étude :

- quatre groupes de trois phrases pour chaque émotion (joie, colère, tristesse et surprise) (12 phrases)
- deux groupes de trois phrases pour les variantes émotionnellement neutres et émotionnellement indéterminées. (6 phrases)

1.2. Le test de perception

Nous comparons ensuite l'évaluation des stimuli écrits (Scherer, 1986 ; Chung, 2000) à celle de ces mêmes stimuli produits par des acteurs afin de poursuivre une démarche cohérente : tester la contribution lexicale à la phrase émotionnelle.

Nous avons ainsi demandé à trois acteurs amateurs de simuler quatre émotions (joie, colère, tristesse et surprise) à partir de notre corpus. Chacune des 18 phrases a également été lue par le locuteur de manière aussi neutre que possible. L'émotion étant omniprésente dans la parole naturelle, suivant différents degrés d'intensité, l' “ émotion neutre ” correspond donc à un état d'âme caractérisé par une intensité émotionnelle la plus faible possible. Les différentes productions des locuteurs ont donc été réparties en 5 groupes : joie, colère, tristesse, surprise et neutre.

Nous nous sommes assuré de la “ validité émotionnelle ” de nos stimuli par à un test de perception et par une courte étude acoustique (calculs de valeurs moyennes de Fo, maxima et minima de Fo, durée et intensité) que nous ne mentionnerons pas ici. Pour ce test de perception, 15 sujets ont dû identifier les émotions simulées par les acteurs. Seules les émotions reconnues à plus de 75% ont été retenues pour la suite de notre étude.

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).

II. Re-synthèses

Les simulations d'émotions identifiées comme correctes ont été re-synthétisées de manière à pouvoir isoler la qualité de la voix. Les paramètres prosodiques des énoncés "émotionnels" d'une phrase (Fo, durée et intensité) ont été transférés aussi précisément que possible sur ceux de la même phrase prononcée de manière neutre (modification 1), et inversement depuis des énoncés neutres d'une phrase sur ceux de la même phrase avec simulation d'une émotion (modification 2) (Ladd, 1985 ; Carlson, 1992). Pour une meilleure lisibilité, nous avons représenté les modifications appliquées à nos phrases sur la figure 1. Pour cette expérience, nous entendons par qualité de voix le "setting" restant inchangé¹ après modification de la fréquence fondamentale, de la durée et de l'intensité relative.

Figure 1

III. Résultats des tests de perception à partir des phrases re-synthétisées

Nos stimuli re-synthétisés (modifications 1 et 2) ont finalement été évalués dans un second test de perception, et comparés avec les simulations originales.

Leur comparaison nous a permis de déduire dans quelle mesure la qualité de la voix de l'énoncé a contribué à l'expression émotionnelle de nos locuteurs. Nous avons également vérifié le rôle de la Fo, de la durée et de l'intensité, pris de manière groupée. Pour rappel, les émotions étudiées sont la joie, la colère, la tristesse et la surprise ; celles-ci seront comparées entre elles mais également à leur version dite "neutre".

Figures 2, 3, 4 et 5

Les tests ont été effectués (à nouveau des chi-carré) sur les réponses fournies à partir de stimuli émotionnels oraux. Les résultats sont analysés ci-dessous.

La surprise est principalement véhiculée par les facteurs prosodiques classiques puisque cette émotion est majoritairement identifiée pour la modification 1 ($\chi^2=54.7, p<0.0001$). Au contraire, pour la colère, la qualité de la voix à elle seule permet une identification très significative ($\chi^2=323.3, p<0.0001$ pour la modification 2). Pour la tristesse et la joie, on ne peut pas conclure à telle prépondérance de la qualité de la voix ou des facteurs prosodiques. La combinaison des deux semble indispensable à une bonne perception des émotions.

L'influence des variations sémantiques appliquées aux phrases de notre corpus n'a pu être validée pour cette étude (phrases reconnues à la lecture comme émotionnellement versus phrases émotionnellement chargées).

¹ La qualité de voix varie (perceptivement tout au moins) avec la Fo, et l'intensité, car elle dépend de la tension, de la longueur des cordes vocales, de la configuration de la glotte, du larynx, de la mucosité, etc...). Les re-synthèses pourraient modifier les stimuli; une certaine prudence est de rigueur quant aux conclusions de cette partie. Nous comptons cependant sur la perception subjective des évaluateurs (Helmholtz, 1863). C'est aussi

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. Pour la suite de cette étude, ce facteur ne sera donc pas pris en compte dans nos mesures statistiques.

IV. Etude laryngographique :

4.1. Intérêt

Le larynx permet de générer un flux sonore à partir d'un flux d'air; cette étape qui est nommée " phonation " est grandement responsable des caractéristiques de la qualité de la voix (Laver, 1994). De nombreuses études théoriques ont proposé des modèles paramétriques de l'onde glottique (modèles incluant la période de voisement (T_0), le quotient d'ouverture (O_q), la vitesse de fermeture des cordes vocales (S_q)). Certaines suggèrent qu'il est possible de distinguer l'effet des paramètres de la source sur des spectres (Klatt & Klatt, 1989 ; Doval & d'Alessandro, 1997) mais ces conclusions sont sujettes à controverse.

Le but de notre approche est destinée à corroborer les résultats fournis par notre expérience perceptive et à mieux comprendre comment les caractéristiques acoustiques de la parole émotionnelle sont déterminées par des modifications physiologiques.

Dans cet article, le quotient O_q est défini² comme le temps ouvert relatif, c'est à dire le rapport de la période ouverte de la glotte sur la période de voisement (Henrich, d'Alessandro & Castellengo, 2000). Du point de vue physiologique, les études laryngographiques ont suggéré que O_q varie de manière significative en fonction de la qualité serrée, tendue (O_q petit), ou détendue (O_q élevé) de la voix.

La seule étude fournissant à notre connaissance des résultats chiffrés, d'après une analyse laryngographique sur les émotions (Johnstone & Scherer, 1999), n'a montré aucune différence significative pour les quotients d'ouverture mais une variation du quotient de fermeture correspondant fortement à la variation de la période fondamentale en fonction des émotions.

4.2. Méthode

L'EKG renseigne sur le mouvement et le mécanisme vibratoires des cordes vocales : en effet le signal électroglottographique, proportionnel à l'impédance électrique du cou, varie en fonction de la surface de contact des cordes vocales. Nous avons décidé de calculer O_q ³ : en dérivant⁴ le signal EGG, 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

pourquoi notre étude laryngographique est indispensable pour venir confirmer ou infirmer nos premières conclusions.

² D'autres définitions existent, notamment " le rapport de la période d'ouverture de la glotte sur la période de voisement ".

³ On le trouve aussi parfois en français sous le nom plus approprié de " quotient ouvert " ou en anglais " open quotient ".

⁴ Le formule de dérivée était la suivante : $dX(t)/dt = [x(i) - x(i-1)]/DT$.

instants permet alors de connaître la valeur du quotient d'ouverture. En effet, O_q sera égal au rapport $d1/d2$ (voir figure 6).

Les prises de données laryngographiques ont été réalisées simultanément avec les enregistrements acoustiques de notre première partie. Nos trois acteurs portaient un collier positionnant deux électrodes placées à l'aide de l'expérimentateur.

Figure 6

Il faut aussi noter que cette méthode de mesure du quotient d'ouverture a des limites. On observe parfois une indétermination du pic associé à l'ouverture glottique, ce qui se traduit par des valeurs surprenantes de O_q (Henrich et al, 2000).

4.3. Valeurs moyennes de quotients d'ouverture.

Les questions abordées dans cette partie concernent les valeurs de O_q calculées pour les phrases des groupes émotionnellement neutres et indéterminés, qui nous serviront de référence. Pourra-t-on distinguer à partir des données EGG des différences entre ces valeurs de référence et certaines émotions, différences qu'il est difficile de distinguer à partir du signal acoustique ? Observera-t-on les résultats suggérés dans notre partie précédente ? Les variations seront elles liées aux variations de F_0 ou d'intensité ?

Figure 7

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

Comparons les valeurs moyennes de O_q pour chacune de nos quatre émotions à celles des phrases neutres :

- L'écart de O_q le plus fort est observé pour la paire colère / neutre (-0,082 pour $p<0.0001$).
- L'écart de O_q 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 (0,01 et 0,031 pour $p=0.002$ et $p<0.0001$ respectivement).

On peut donc en conclure que

- l'influence de la qualité de la voix est la plus importante pour la colère.
- le peu d'écart entre le O_q de la voix neutre et celui de la surprise corrobore le peu d'influence de la qualité de la voix notée dans la conclusion de l'expérience perceptive.
- O_q permet 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.

4.4. Variation des valeurs de O_q en fonction des contours de F_0 et d'intensité.

Nos résultats suggèrent que la qualité de la voix tend à varier au cours d'une phrase et que certaines de nos phrases re-synthétisées ont pu être identifiées grâce à quelques syllabes seulement, dans la mesure où Oq augmente ou diminue de manière significative sur une courte durée.

Nous avons tenté de vérifier l'effet des deux variables acoustiques - Fo et l'intensité - sur nos valeurs de Oq calculées en fonction de chaque émotion. 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 (voir figures 8 à 13). On remarque que l'effet est le plus significatif pour les deux émotions 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 8, 9, 10, 11, 12 et 13

Conclusion

Dans cette étude, nous avons montré que la contribution perceptive de la qualité de la voix varie en fonction de chaque émotion. Elle est très importante pour la perception de la colère et nettement plus négligeable pour la surprise. Les évaluations de nos re-synthèses dans un test de perception ont également suggéré que la joie et la tristesse sont plus facilement identifiées grâce à une combinaison des facteurs prosodiques et de la qualité de la voix. Les variabilités inter-locuteurs que nous avons observées laissent également préjuger de l'existence de différentes stratégies pour la simulation de la tristesse, la joie et la colère.

Les quotients d'ouverture calculés à partir de courbes électroglottographiques, enregistrées simultanément avec nos stimuli, se sont révélés être significativement différents pour la joie, la colère, la tristesse et la neutralité et ont corroboré nos résultats obtenus lors de l'analyse perceptive.

Nous avons ainsi pu mettre en évidence des relations entre les variations de Fo, d'intensité et les valeurs de quotients d'ouverture, notamment pour les émotions à forte activation. Ces analyses devront cependant être approfondies, car d'autres facteurs physiologiques, tels que la pression sous-glottique, semblent avoir un rôle primordial sur ces variations (Vaissière, 2004).

Références

1. Carlson, R., Granström, B. & Nord, L. (1992). Experiments with emotive speech - acted utterances and synthesized replicas, *Eurospeech*, 92.
2. Chung, S.-J. (2000), *L'expression et la perception de l'émotion extraite de la parole spontanée : évidences du coréen et de l'anglais*, Thèse de doctorat, Université de la Sorbonne Nouvelle, Paris III, France.
3. 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.
4. Doval, B. & d'Alessandro, C. (1997). Spectral correlates of glottal waveform models, an analytic study, *Proceedings of the ICASSP*, april 21-24, Munich, Germany.

5. Fant, G. (1960). Acoustic Theory of Speech Production, *Mouton, The Hague, The Netherlands*.
6. Fónagy, I. (1971a), Double coding in speech, *Semiotica* 3, 189-222.
7. Fónagy, I. (1983a), *La vive voix: Essais de psycho-phonétique*, Bibliothèque scientifique Payot, Paris.
8. Helmholtz von, H.L.F. (1863), *Die Lehre von den Tonempfindungen*, F. Wigneg & Son, Braunschweig.
9. 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.
10. Johnstone. T & Scherer. K (1999). The effects of emotions on voice quality, *Proceedings of ICPhS*, San Francisco, 1-7 August 1999. U.S.A.
11. Klatt, D & Klatt, L. (1989), Analysis, synthesis, and perception of voice quality variations among female and male talkers, *J. Acoust. Soc. Am.* 87, (2).
12. Ladd, D. R., Silverman, K. E. A., Tolkmitt, F., Bergmann, G., & Scherer, K. R. (1985), Evidence for the independent function of intonation contour type, voice quality, and Fo range in signaling speaker affect, *J. Acoust. Soc. Am.* 78, 435-444.
13. Laver, J. (1980), *The phonetic description of voice quality*, Cambridge University Press.
14. Laver J. (1991), *The gift of speech*, Edinburgh University Press., Edinburgh, UK.
15. Laver J. (1994), *Principles of phonetics*, Oxford University Press., Oxford, UK.
16. Léon, P. R. (1971), *Essais de Phonostylistique*, Didier, Paris, Montréal, Bruxelles.
17. 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.
18. 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.
19. Osgood, C. E., Suci, G. J. & Tannenbaum, P. H. (1957), *The measurement of meaning*, University of Illinois Press.
20. Scherer, K. R. (1986), Vocal Affect Expression: A Review and a Model for Future Research, *Psychological Bulletin* 99(2), 143-165.
21. 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.
22. Scherer, K. R., Johnstone, T., & Sangsue, J. (1998), L'état émotionnel du locuteur: facteur négligé mais non négligeable pour la technologie de la parole, *Actes des XXIIème Journées d'Etudes sur la Parole*, Matigny, Suisse, 249-257.

Annexes

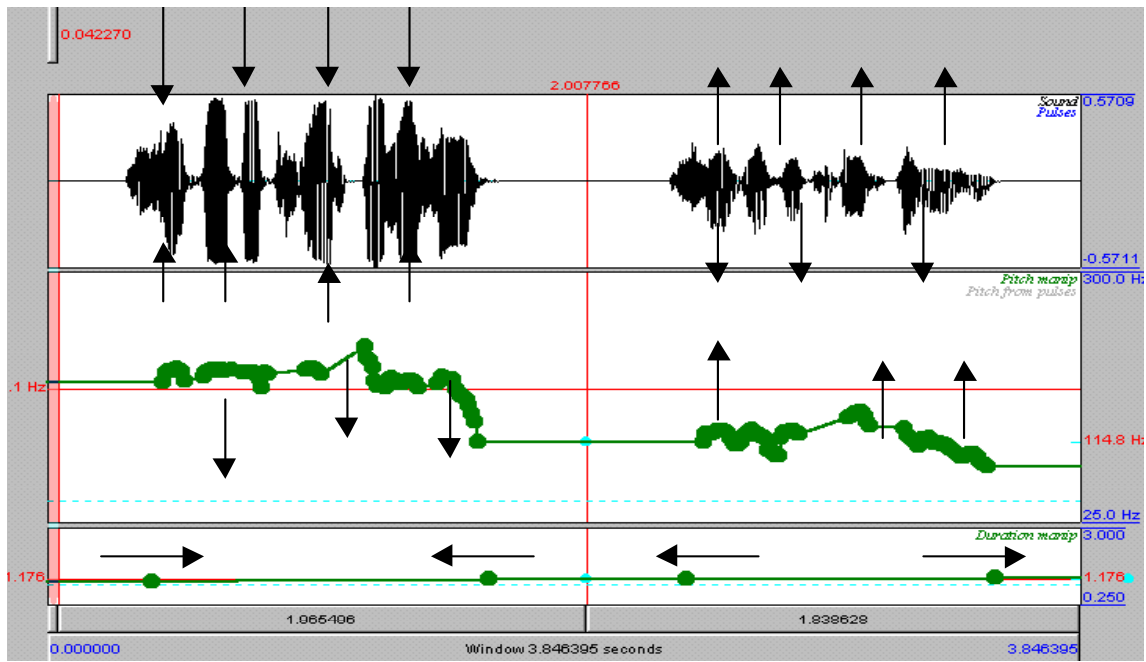
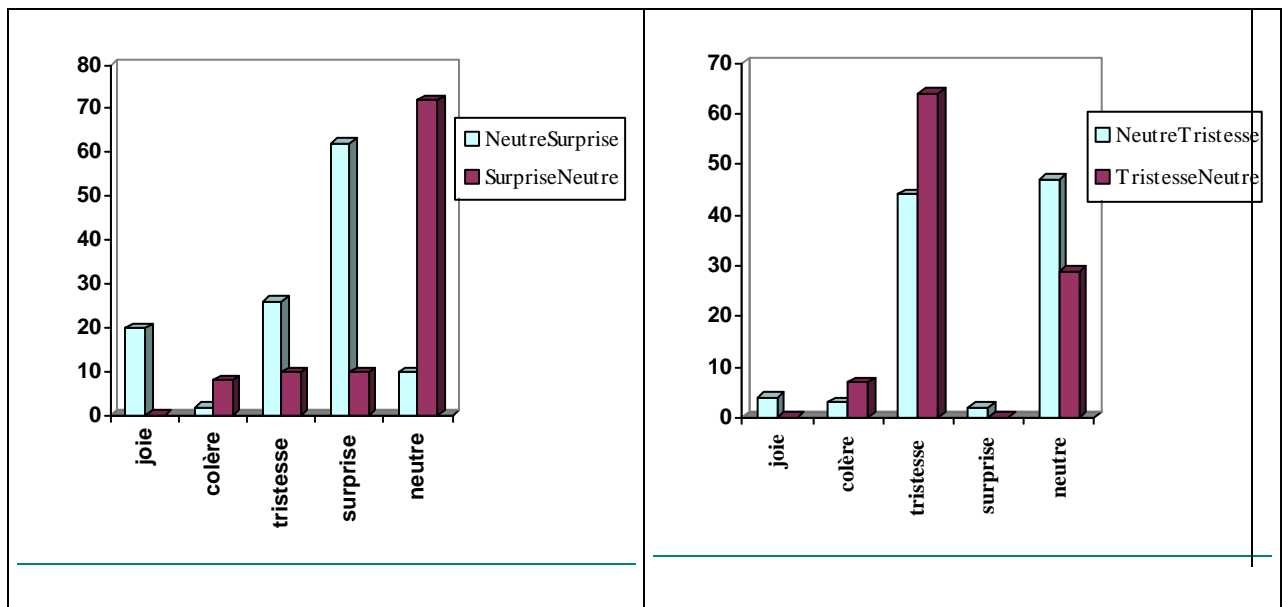
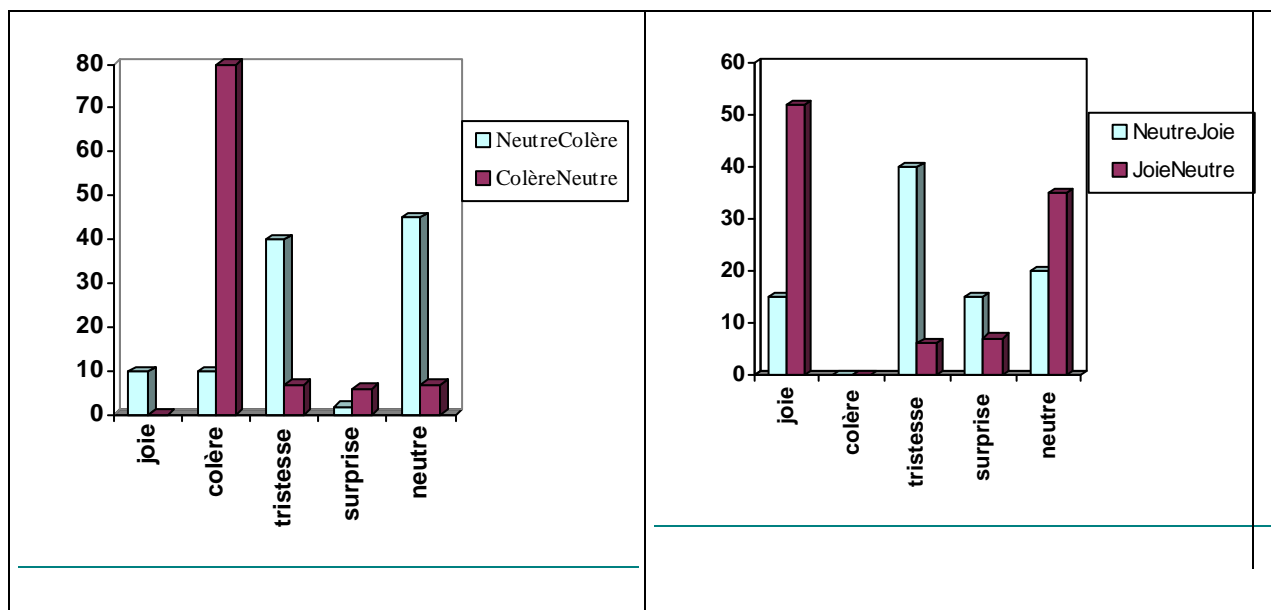


Figure 1 : La phrase “ je vais détruire ta maison ” prononcée avec colère simulée (à gauche) et de manière neutre (à droite).

Les procédés de re-synthèse sur l'intensité, la fréquence fondamentale et la durée, sont exprimés par des flèches.

A gauche la modification 2 et à droite la modification 1.





Figures 2, 3, 4 et 5 : Identification des versions re-synthétisées en pourcentages. Comparaison de la modification 1 (qualité de voix neutre représentée par la colonne de gauche) et de la modification 2 (qualité de voix émotionnelle représentée par la colonne de droite).

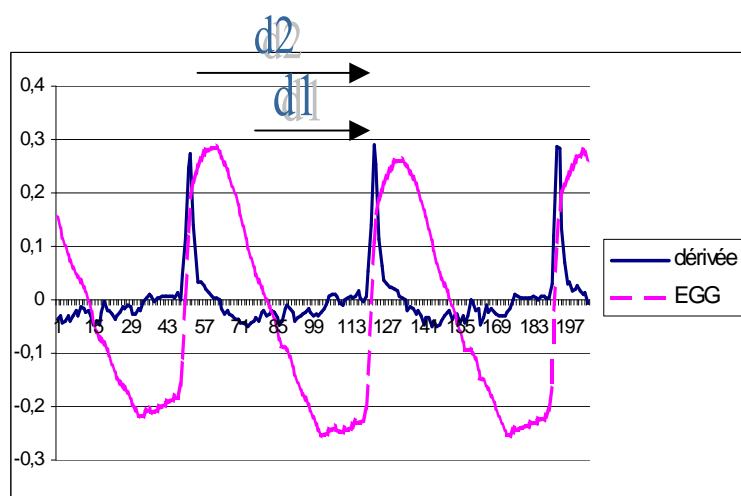


Figure 6 : Graphe de la courbe EGG et de sa dérivée (les échelles ont été modifiées pour une meilleure visibilité). O_q est calculé selon la formule :

$$O_q = d1/d2$$

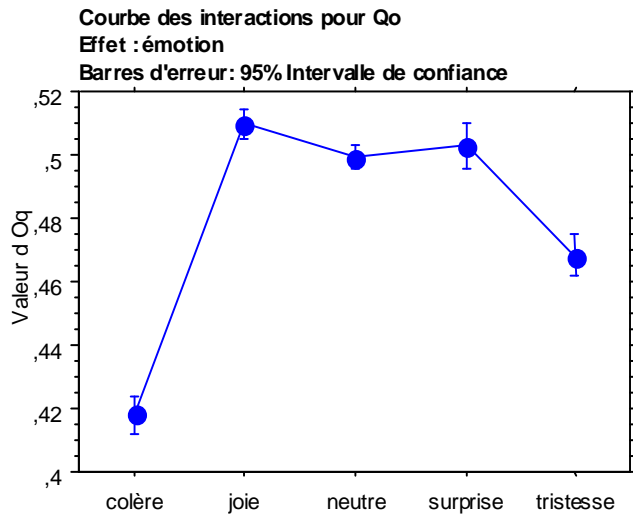
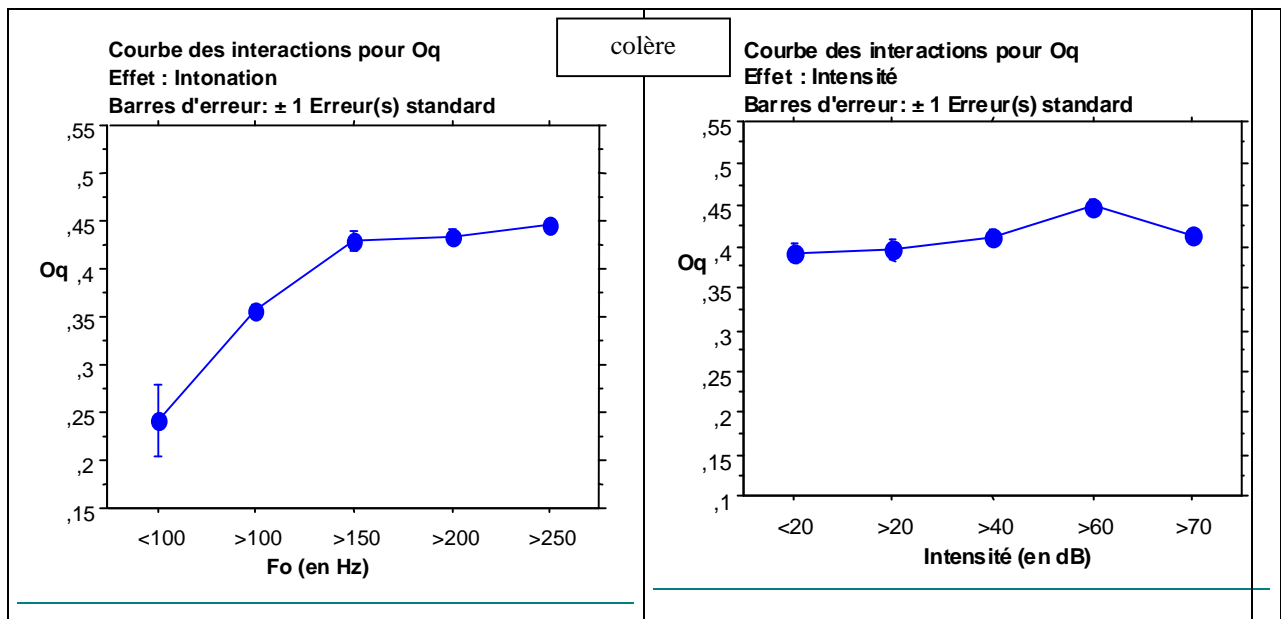
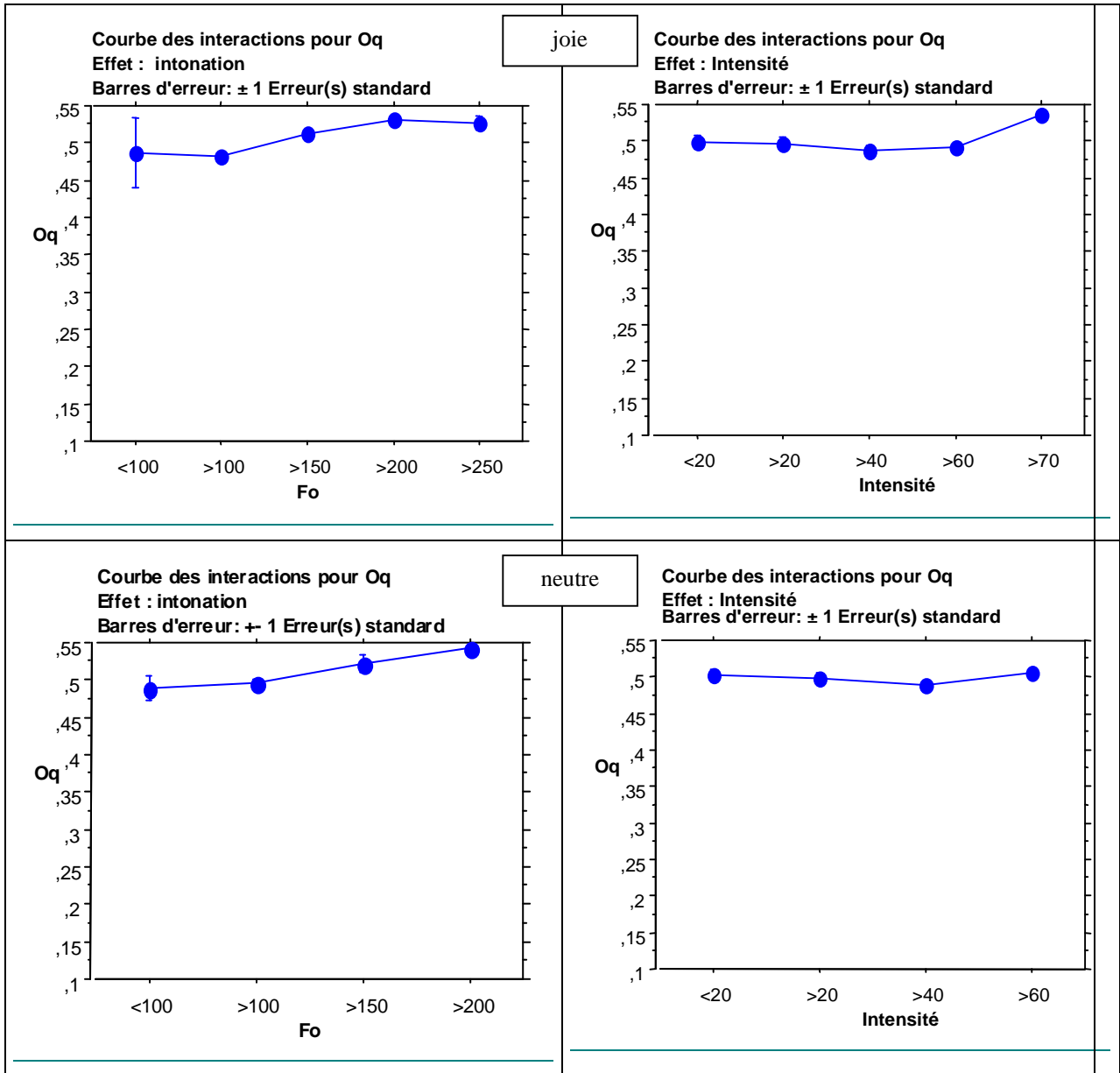


Figure 7: Valeurs moyennes du quotient d'ouverture Oq pour chaque émotion.





Figures 8, 9, 10, 11, 12 et 13 : Effet de la Fo (à gauche) et de l'intensité relative (à droite) sur Oq calculé en fonction de l'émotion (colère, joie et neutre de haut en bas).