

Perception des émotions « positives » et « négatives » chez les auditeurs français et japonais à travers le contour de F0

Hideko Yamasaki

Université de Paris III, Laboratoire de Phonétique et Phonologie - UMR7081
19 rue des Bernardins, Paris, France
e-mail : hidy@tiscali.fr

ABSTRACT

The present study reports perceptual aspects of “positive” or “negative” emotions for synthetic stimuli: 5 different contours of F0 were analysed in the four last syllables of French utterances (filtered or non-filtered). The results showed that French subjects (n=60) associate “positive” emotions with F0 raising contour and the “negative” emotions with F0 “falling” contour. Japanese subjects (n=20) perceived “negative” emotions in “flat” and “falling” contours. The second analysis confirmed that the semantic effects were higher for the French subjects who integrated both F0 contour and semantic aspects in determining the emotion, and that Japanese subjects integrated mostly F0 contour.

1. INTRODUCTION

L'intégration de plusieurs indices acoustiques (F0 (registre, plage, contour), durée, intensité, débit de la parole) est l'un des éléments clefs dans l'expression des énoncés émotionnels. Les diverses études antérieures ont indiqué les tendances suivantes : un F0 élevé évoque l'expression de la joie et de la colère, au contraire, un F0 bas évoque la tristesse avec souvent une plage de F0 réduite (Fónagy[3], Frick[4], Léon[5], Scherer[8],[9]). Il a été également indiqué que le F0 et la durée sont plus marquées dans l'expression de la « joie » par rapport aux émotions « neutre », « colère », « tristesse » (Pell[6]). Il est toutefois difficile de donner un modèle définitif et figé sur le rapport entre les indices acoustiques et l'expression des différents états émotionnels. Cette difficulté vient :

- premièrement de la différence méthodologique (émotions simulées ou réelles),
- deuxièmement de la difficulté d'établir une définition absolue des différents états émotionnels (joie, colère, peur, tristesse, etc. ; l'émotion est considérée comme un phénomène continu et non catégoriel),
- troisièmement de la différence inter-individuelle dans la représentation des émotions chez chaque individu (Colletta & Tcherkassof[2]), et
- quatrièmement les problèmes d'intégration de divers paramètres au niveau perceptif.

L'étude de Chung ([1] : en coréen et en anglais) a montré que l'émotion a été mieux identifiée avec le contour de F0 en fin de phrase par rapport au début ou au milieu de phrase. Dans notre étude, le recueil du corpus dans l'extrait du discours d'un test psychologique (test de

Rorschach (Rauche[7])) avait permis d'obtenir des énoncés avec des émotions réelles. Les expériences décrites ici sur un corpus naturel et synthétisé doivent permettre de mieux appréhender la perception des émotions en rapport avec le contour de F0 et celle de l'effet sémantique (influence des mots évoquant l'émotion « positive » ou « négative »), chez les auditeurs français et les auditeurs japonais.

2. EXPERIENCE 1 : PERCEPTION DES ENONCES « NATURELS » VS « FILTRES »

2.1. Sujets

Groupe français (FR) : 60 auditeurs français (M=35ans).

Groupe japonais (JP) : 20 auditeurs japonais qui vivent en France (M=33ans).

2.2. Stimuli

Le déroulement du test psychodiagnostique de Rorschach de deux sujets non-pathologiques (françaises, 45 et 50 ans) a été enregistré. Ce discours contiendrait l'expression de différents états émotionnels en réaction à 10 taches d'encre prédéfinies. Dans ce discours (± 80 énoncés/sujet), les énoncés « positifs » et « négatifs » ont été sélectionnés en fonction du contenu latent du matériel et de la présence du lexique émotionnel (ex. « jolie robe » (positif), « persécution » (négatif)). La validité de ce choix réalisé par l'auteur est vérifiée à l'étape suivante au moyen d'un test de perception. Le concept de ces deux classifications (valences) à deux extrêmes a permis une application à des sujets de cultures différentes sans qu'il y ait une confusion dans la représentation émotionnelle. Dans notre étude, l'émotion « positive » réfère à la joie et/ou au plaisir, tandis que l'émotion « négative » réfère à l'angoisse, évoqués par le matériel du test.

Après l'extraction du corpus, les stimuli ont été repartis en deux groupes : stimuli « Naturels (N) » et stimuli « Filtrés (F) » à bande passante de 500Hz à 1000Hz avec l'atténuation de l'effet sémantique (phrases très peu compréhensibles). De cette façon, nous avons obtenu 56 stimuli « émotionnels » (14 x N« positifs », N« négatifs », F« positifs », F« négatifs »).

2.3. Procédure

Les auditeurs ont répondu individuellement dans une salle calme équipée d'un ordinateur. Le sujet entendait une série de stimuli au moyen d'un casque. La tâche expérimentale du sujet consistait à donner à chaque stimulus l'une des deux valences émotionnelles : « positive » ou « négative » (procédure du choix forcé). Le sujet devait répondre le plus rapidement possible (environ 2 secondes) en cliquant dans la case correspondante sur l'écran de l'ordinateur. Le sujet entendait d'abord la série des stimuli naturels, puis celle des stimuli filtrés.

2.4. Résultats

Nous avons calculé le taux moyen de réponses correctes (TRC), à savoir les réponses en accord avec notre choix émotionnel.

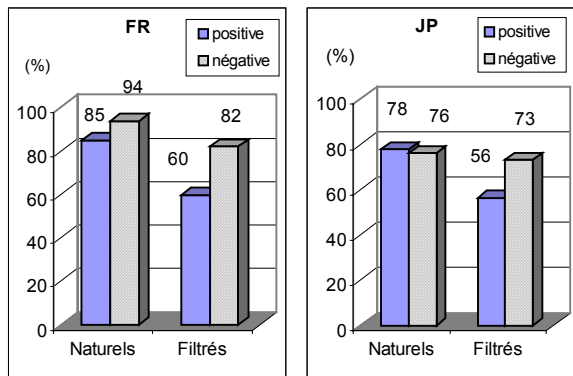


Figure 1 : Taux moyen de réponses correctes (56 stimuli : 60 sujets français vs 20 sujets japonais).

Groupe FR : les sujets français ont catégorisé les énoncés « Naturels » comme notre choix (TRC élevé : >85%). En particulier, nous avons observé un effet plafond pour les énoncés Négatifs « Naturels » (94%).

Groupe JP : les sujets japonais ont, comme les sujets français, catégorisé les énoncés « Naturels » en accord avec notre choix (TRC > 75%).

Les sujets japonais ont donné significativement moins de réponses correctes par rapport aux sujets français ($F(1,52)=27,72$; $p<.0001$). Par ailleurs, il est remarqué que, pour les deux groupes de sujets, l'aspect acoustique des énoncés « Filtrés » donne une impression globale de l'émotion négative, à savoir la baisse significative de taux de réponses correctes pour les énoncés « Filtrés » ($F(1,52)=19,21$; $p<.01$).

Les résultats de l'expérience 1 ont ainsi indiqué que la classification des énoncés avec le contenu émotionnel a été conforme à notre choix chez les sujets français et japonais. Ces résultats indiquent la fiabilité de notre choix de corpus « positif » ou « négatif ».

3. EXPERIENCE 2 : ENONCES NATURELS VS FILTRES AVEC 5 CONTOURS DIFFERENTS DU F0

Cette expérience a pour but d'observer la perception des émotions en rapport avec le contour de F0 en fin de phrase, ainsi qu'avec l'influence sémantique.

3.1. Sujets : identique à l'expérience 1.

3.2. Stimuli :

A partir de 12 énoncés (12-14 syllabes) ayant le contenu sémantique « positif » ou « négatif » (classification confirmée dans l'expérience 1), nous avons obtenu les stimuli synthétisés avec 5 contours de F0 différents en fin de phrase (4 dernières syllabes). De cette manière, nous avons obtenu 60 énoncés synthétisés « Non-Filtrés » avec l'effet sémantique, puis 60 énoncés synthétisés et « Filtrés » avec atténuation de l'effet sémantique (procédure du filtrage identique à l'expérience 1).

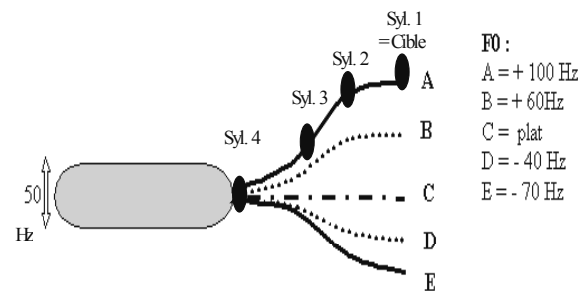


Figure 2 : Schéma de la synthèse du F0 à cinq niveaux.

3.3. Procédure : identique à l'expérience 1.

3.4. Résultats

1. Taux de réponses « positives » : résultats en fonction du taux de réponses « positives » pour les stimuli « Non-Filtrés » et « Filtrés ».

Groupe FR :

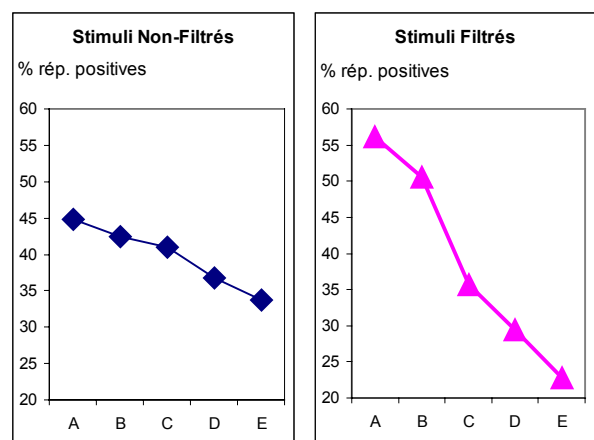


Figure 3 : Taux de réponses « positives » : contour du F0 en fin de phrase (120 stimuli, 60 sujets français).

Énoncés non-filtrés : le taux de réponses « positives » est en rapport avec le contour de F0 en fin de phrase : plus le F0 est montant en fin de phrase, plus les auditeurs français perçoivent l'émotion « positive ». Au contraire, plus le F0 est descendant, plus ils perçoivent l'émotion « négative ». Le contour nettement descendant en fin de phrase de « D (-40 Hz) » et de « E (-70 Hz) » auraient attiré les réponses « négatives » chez ces auditeurs.

Énoncés filtrés : la même tendance est observée avec un effet renforçant. Pour les énoncés filtrés (donc avec l'effet sémantique atténué), les auditeurs français s'appuient nettement plus sur le contour du F0 dans la décision de la valence émotionnelle.

L'analyse de variance indique une différence significative dans l'interaction entre l'effet de contour et l'effet de filtrage : l'effet de contour est plus fort pour les énoncés « Filtrés » que pour les énoncés « Non-Filtrés » ; ($F(4,236) > 25,56$, $p < .001$). L'analyse complémentaire indique que la pente de régression est significativement plus forte pour les énoncés « Filtrés » que pour les énoncés « Non-Filtrés » (t de student : $t(59) = 6,94$, $p < .001$). Ces résultats indiquent que les sujets français ont une tendance nette à percevoir les valences affectives des énoncés en fonction du contour du F0 en fin de phrase pour les énoncés « Filtrés ».

Groupe JP :

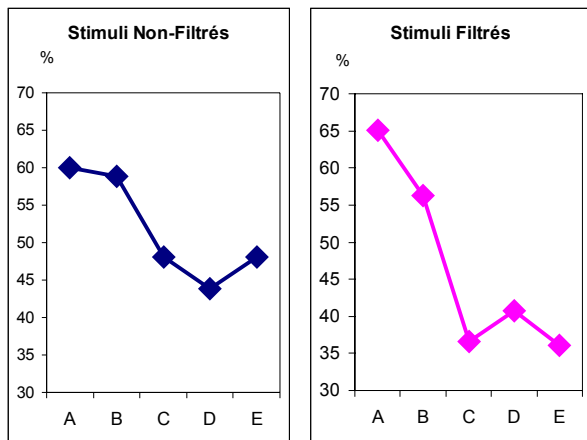


Figure 4 : Taux de réponses « positives » : contour du F0 en fin de phrase (120 stimuli, 20 sujets japonais).

Énoncés non-filtrés : la perception de l'émotion est, pour les auditeurs japonais, également liée au contour de F0. Toutefois, pour ces sujets, nous avons remarqué que le taux de réponses « positives » est surtout lié au contour montant. Les auditeurs japonais ont lié le contour plat (la monotonie) et le contour descendant à l'émotion négative.

Énoncés filtrés : la même tendance est observée avec un effet renforçant. La perception de l'émotion positive est nettement plus liée au contour montant, et les réponses « positives » diminuent nettement pour le contour plat (C) et les contours descendants (D,E).

L'analyse de variance indique une différence marginale dans l'interaction entre l'effet de contour et l'effet de

filtrage : l'effet de contour a tendance à plus influencer sur les énoncés « Filtrés » ; ($F(4,76) > 2,42$, $p < .055$). L'analyse complémentaire indique que la différence de la pente de régression entre les énoncés « Filtrés » et les énoncés « Non-Filtrés » est également marginale (t de student : $t(19) = 1,89$; $p < .076$).

Ces résultats indiquent que les sujets français montrent, dans la décision de la valence émotionnelle, une tendance nette à s'appuyer sur le contour du F0 lorsque les énoncés sont « Filtrés ». Pour les sujets japonais, la même tendance existe, mais avec un effet moindre.

2. Effet sémantique et contour du F0

Dans cette analyse, nous avons étudié le nombre de réponses « positives » dans les énoncés synthétisés « Non-Filtrés » ou « Filtrés ».

Groupe FR :

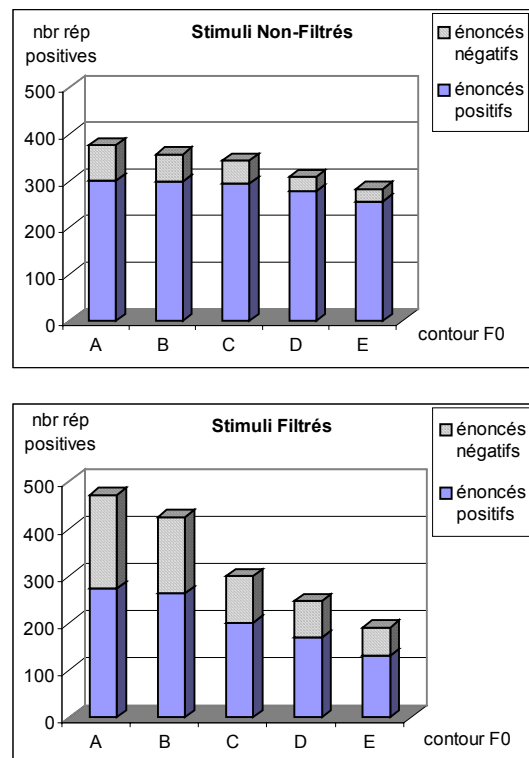


Figure 5 : Taux de réponses « positives » : « non-filtrés » vs « filtrés » (120 stimuli, 60 sujets français).

Pour les énoncés « Non-Filtrés », nous avons observé très peu de réponses « positives » lorsque leur contenu était négatif. De plus, le nombre de réponses « positives » diminue légèrement du contour de F0 montant au descendant. Pour les énoncés « Filtrés », l'effet est plus prééminent. On observe une augmentation de réponses « positives » pour les énoncés de contenu « négatifs ». Le nombre de réponses « positives » diminue de façon nette vers le contour du F0 descendant.

Groupe JP :

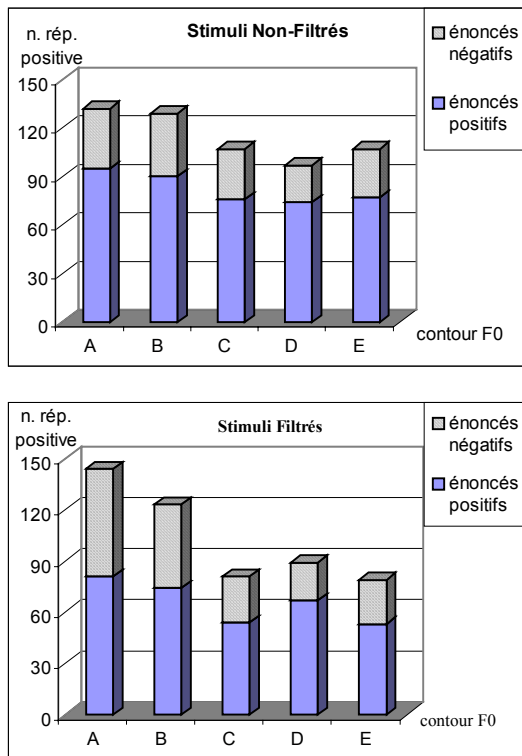


Figure 6 : Taux de réponses «positives» : énoncés « non-filtrés » vs « filtrés » (120 stimuli; 20 sujets japonais).

Chez les auditeurs japonais, le nombre de réponses « positives » n'est pas régressif en fonction des contours. Pour les stimuli « Non-Filtrés » et « Filtrés », le nombre de réponses « positives » est moindre pour le contour plat et les contours descendant (C,D,E).

Dans les deux groupes de sujets, une baisse de la fréquence globale des réponses positives dans les énoncés « Filtrés » ($X^2(1) = 200.5$, $p < .00001$ pour FR, $X^2(1) = 8.2$, $p < .005$ pour JP). Il est noté que chez les sujets français, contrairement aux sujets japonais, la diminution du nombre de réponses positives entre les énoncés « Non-filtrés » et « Filtrés » est nettement marquée. Les résultats de cette analyse indiquent que les sujets français (français langue maternelle) auraient pris en compte la sémantique dans la décision de la valence émotionnelle tout en étant parallèlement influencés par le contour du F0. En revanche, les sujets japonais (français langue étrangère) semblent s'appuyer plus sur le contour du F0, et ce, même pour les énoncés (« Non-Filtrés ») avec le contenu sémantique perceptible.

3. DISCUSSION GENERALE

Les résultats de nos expériences ont montré le lien possible entre le contour de F0 (fin de phrase) et la perception de l'émotion « positive » ou « négative » : le F0 montant en fin de phrase est lié à la perception de la valence « positive » et le F0 descendant, à la perception de la valence « négative ». Ces résultats sont en accord avec les observations antérieures qui ont décrit une

tendance telle que le F0 élevé est lié à l'impression de gaieté, et le F0 bas, à l'impression de tristesse. L'importante taille de l'échantillon de cette étude a permis d'obtenir des résultats fiables. Il est toutefois important de mentionner que le F0 n'est pas le seul indice dans l'évocation de l'émotion : notre seconde expérience a montré le rôle non négligeable de l'effet sémantique qui serait présent chez les auditeurs de la langue maternelle, et moindre chez les auditeurs de la langue étrangère.

REMERCIEMENT

L'auteur tient à remercier Professeur Jacqueline Vaissière, Monsieur Pierre Hallé et Cédric Gendrot pour leurs commentaires pertinents sur le présent travail.

BIBLIOGRAPHIE

- [1] S.J. Chung, Perception de l'émotion : les indices prosodique et la durée minimale. In *Actes de XXII^{ème} Journées d'Etudes sur la Parole*, Martigny-Suisse, 159-162, 1998.
- [2] J.M. Colletta et A. Tcherkassof, Les émotions : une problématique pluri- et interdisciplinaire, In *Les émotions : cognition, langage et développement*, Ed. Mardaga, Sprimont, 5-11, 2003.
- [3] I. Fónagy, La vive voix : Essai de psychophonétique, Ed. Payot, Paris, 1986.
- [4] R.W. Frick, Communicating emotion: the role of prosodic features. In *Psychol. Bull.*, 97 :412-429, 1985.
- [5] P.R. Léon, Précis de phonostylistique, Ed. Nathan, Paris, 1971.
- [6] M.D. Pell (2001), Influence of emotion and focus location on prosody in matched statements and questions. In *J.Acoust.Soc.Am.*, 109 (4), 1668-1680
- [7] N. Rauche de Traubenber, La pratique du Rorschach, Paris, PUF, 1970.
- [8] K.R. Scherer, Minimal cues in the vocal communication of affect : judging emotions from content-masked speech. In *J.Psycholinguistic Res.1*, 269-285, 1972.
- [9] K.R. Scherer, D.R. Ladd, K.E.A. Silverman, Vocal cues to speaker affect : Testing two models. In *J.Acoust.Soc.Am.*, 76 (5): 1346-1356, 1984.