



Analyse prosodique des affects sociaux dans l'interaction face à face en japonais

Dominique Fourer, Marine Guerry, Takaaki Shochi, Jean-Luc Rouas,
Jean-Julien Aucouturier, Albert Rilliard

► To cite this version:

Dominique Fourer, Marine Guerry, Takaaki Shochi, Jean-Luc Rouas, Jean-Julien Aucouturier, et al.. Analyse prosodique des affects sociaux dans l'interaction face à face en japonais. XXXèmes Journées d'études sur la parole, Jun 2014, Le Mans, France. 2014. <hal-00992083>

HAL Id: hal-00992083

<https://hal.archives-ouvertes.fr/hal-00992083>

Submitted on 16 May 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Analyse prosodique des affects sociaux dans l'interaction face à face en japonais

Dominique Fourer¹, Marine Guerry², Takaaki Shochi^{1,4}
Jean-Luc Rouas¹, Jean-Julien Aucouturier³, Albert Rilliard⁵

(1)LaBRI - CNRS UMR5800, Univ. Bordeaux 1, France

(2) Univ. Bordeaux Montaigne, France ; (3) IRCAM - CNRS UMR9912 - UPMC, Paris, France

(4) CLLE-ERSSàB UMR5263 CNRS, Bordeaux, France ; (5) LIMSI - CNRS UPR3251, Orsay, France

RÉSUMÉ

Le but de cet article est de caractériser la prosodie attitudinale en langue japonaise. Ce travail s'appuie sur des travaux décrivant 16 attitudes correspondant à des situations de communication différentes. Ces situations peuvent, ou non être conventionnalisées dans la langue japonaise. Les paramètres estimés de fréquence fondamentale, d'amplitude et de durée ont été extraits d'énoncés exprimant ces 16 attitudes en japonais. Dans cette étude, nous présentons les effets sur ces paramètres des facteurs sexe du locuteur et expression attitudinale. Nous analysons également lesquels de ces paramètres prosodiques sont les plus discriminants pour caractériser acoustiquement chaque attitude.

ABSTRACT

Prosody analysis of Japanese social affects

The aim of this paper is to examine cues for prosodic characterization of attitudes in Japanese. This work is based on previous studies where 16 communicative social affects were defined. The audio signal parameters (fundamental frequency, amplitude and duration) of previously recorded Japanese attitudes, are statistically analyzed. Interesting interactions among the parameters, the gender and the expression of specific attitude (e.g. politeness) were found, and we report on which parameters most significantly characterize each attitude.

MOTS-CLÉS : prosodie, attitude, parole, japonais.

KEYWORDS: prosody, attitude, speech, Japanese.

1 Introduction

L'expression prosodique d'attitudes, telles qu'elles sont définies par (Wichmann, 2000), est un moyen utilisé par les locuteurs pour effectuer certains actes de langage intentionnels (Fónagy *et al.*, 1984). Ces performances prosodiques sont partiellement liées à la maîtrise qu'a le locuteur de la langue, à sa personnalité, à son sexe, ainsi qu'au contexte communicationnel et linguistique - pour une discussion voir (Morales et Rilliard, 2014). Chaque langue possède des formes et des variations prosodiques conventionnelles, dépendantes du contexte d'interaction. Les études

observant ces variations prosodiques reposent généralement sur des inventaires d'expressions conventionnalisées de la langue étudiée (Fujisaki et Hirose, 1993; Morlec *et al.*, 2001; de Moraes, 2008; Gu *et al.*, 2011). La plupart des études citées utilisent des corpus de laboratoire pour lesquels des locuteurs élicitent des variations prosodiques correspondant à un label donné. Ces labels sont interprétés par le locuteur dans son cadre culturel propre. De ce fait, la comparaison des performances entre deux langues (Shochi *et al.*, 2009; Rilliard *et al.*, 2009) se heurte aux critiques formulées par Wierzbicka sur certaines comparaisons interculturelles d'expressions émotionnelles (Wierzbicka, 1985, 1992, 2005). Les objets étudiés ne sont pas nécessairement comparables car actualisant des concepts différents. Si ce problème peut être contourné expérimentalement pour des tests de perception (Rilliard *et al.*, 2014b), les corpus enregistrés de cette manière ne permettent pas de contourner ce problème pour tout ce qui concerne les aspects de production. Dans ce but, le corpus PADE (Rilliard *et al.*, 2013, 2014a) définit des contextes d'interaction amenant les locuteurs à exprimer des variations prosodiques spécifiques à ces contextes. Les contextes restent identiques quelles que soient la culture et la langue étudiées, dans le but d'observer les variations prosodiques qui caractérisent les variations expressives correspondantes. Cet article s'attache à l'analyse acoustique des variations prosodiques produites par 19 locuteurs de langue maternelle japonaises ayant été enregistrés dans le cadre de cette étude. Une des finalités de cette analyse consiste à pouvoir synthétiser des variations prosodiques pertinentes pour exprimer ces variations attitudinales. Nous avons décidé d'utiliser uniquement les descripteurs prosodiques suivants : la fréquence fondamentale (F_0), l'amplitude *Root Mean Square* (RMS) et la durée des mores. En utilisant ces caractéristiques pour une analyse statistique basée sur une analyse de variance à mesures répétées (RM-ANOVA), nous sommes capables de déterminer comment chaque paramètre contribue à la différenciation des attitudes. Cet article est organisé comme suit : le cadre de travail utilisé pour l'expression des attitudes par les locuteurs est décrit dans la section 2. Le protocole d'enregistrement et les méthodes d'extraction des caractéristiques sont respectivement décrits dans les sections 3 et 4. Enfin, les résultats des analyses statistiques sont présentés dans la section 5 et discutés dans la section 6.

2 Contextes sociaux des expressions attitudinales

De manière similaire au protocole expérimental proposé par (Gu *et al.*, 2011), chaque sujet interagit avec un interlocuteur durant un dialogue aboutissant sur une phrase cible prononcée dans un contexte induisant une variation attitudinale. Afin d'immerger les sujets dans un contexte donné, un scénario a été préparé pour chaque attitude permettant d'aboutir sur la production des phrases cibles. 16 contextes correspondant à des expressions attitudinales conventionnalisées dans différentes langues (mais pas nécessairement dans chacune) ont été préparés : Admiration (ADMI), Arrogance (ARRO), Autorité (AUTH), Mépris (CONT), Doute (DOUB), Ironie (IRON), Irritation (IRRI), Déclaration (DECL), Question (QUES), Évidence (OBVI), Politesse (POLI), Séduction (SEDU), Sincérité (SINC), Surprise (SURP), Incertitude (UNCE), Marcher sur les œufs (WOEG). La définition de ces 16 attitudes, la description des concepts correspondants ainsi que la procédure d'enregistrement sont détaillées dans (Rilliard *et al.*, 2013). A titre d'exemple, voici un extrait de la situation créée pour l'attitude « Marcher sur des œufs » :

- Votre chef (locuteur A) vous (locuteur B) a chargé de préparer une salle pour une importante conférence. Votre chef est un homme nerveux qui ne supporte pas que les choses ne soient pas faites minutieusement et il se met facilement en colère si les choses ne sont pas à son goût.

Votre chef inspecte la salle où va commencer la conférence, lorsqu'il trouve dans une poubelle une vieille banane à moitié mangée. Il est furieux.

Il existe des contextes qui ne correspondent à aucune expression conventionnelle dans certaines langues étudiées (pour l'instant l'anglais des États-Unis, le japonais et le français hexagonal). C'est typiquement le cas de l'expression désignée par « marcher sur des œufs » qui correspond approximativement au concept japonais de *kyoshuku*, décrit par (Sadanobu, 2004) comme « correspondant à un mélange entre une honte douloureuse et de l'embarras venant du fait que le locuteur est conscient du fardeau qu'il impose à son interlocuteur de par sa requête » (p.34, traduction des auteurs). Dans ce travail, nous nous concentrons sur les résultats du japonais réalisés par des locuteurs natifs et sur l'une des phrases cibles utilisées pour l'enregistrement des attitudes : « Banana ».

3 Procédure d'enregistrement

Les performances audio-visuelles de 19 locuteurs natifs du japonais (11 femmes, 8 hommes) parlant le dialecte de Tokyo ont été enregistrées. La plupart des locuteurs sont des étudiants, qui ont été payés pour leur participation. Les enregistrements ont eu lieu dans une chambre sourde à l'Université Waseda au Japon. Le son a été enregistré à l'aide d'un microphone omnidirectionnel *Earthworks QTC1*, placé à un mètre de la bouche du locuteur (cette distance a été choisie pour limiter les perturbations d'intensité liées aux mouvements du locuteur). Le niveau du microphone a été calibré avant chaque session d'enregistrement à l'aide d'un calibre acoustique *Bruel & Kjaer*. Ainsi, le niveau de pression acoustique peut être corrigé après enregistrement à un niveau comparable entre tous les locuteurs. L'énoncé de la phrase cible « Banana » produite par chaque locuteur pour chaque attitude (total de 304 stimuli) a été manuellement isolée en fichiers individuels. Ceux-ci ont été enregistrés au format WAV en utilisant une quantification sur 16 bits avec une fréquence d'échantillonnage de 16 kHz. Chaque stimulus a été découpé afin de retirer les silences de début et de fin phrase, puis manuellement segmenté au niveau des phonèmes à l'aide du logiciel PRAAT (Boersma et Weenink, 2001).

4 Estimation des paramètres prosodiques

Chaque stimulus est caractérisé par sa fréquence fondamentale (F_0), son amplitude instantanée ainsi que la durée de chaque phonème. La F_0 mesurée en Hertz est estimée pour chaque stimulus toutes les 10 ms par l'algorithme SWIPE (Camacho et Harris, 2008). Pour l'analyse statistique (cf. Section 5), seule la valeur de F_0 correspondant à la partie voisée de chaque phonème est considérée. Par la suite, nous calculons la moyenne des F_0 pour chaque voyelle (3 valeurs par stimulus). Enfin, nous appliquons l'algorithme de *MultiDimensional Scaling algorithm* (MDS) (Bronstein *et al.*, 2006) sur chaque vecteur de paramètre F_0 normalisé en divisant chaque élément par la F_0 moyenne du sujet considéré (cf. Section 5.4). Cette normalisation permet de comparer uniquement les modulations de F_0 autour de la fréquence moyenne, entre tous les sujets du corpus. Ainsi, l'algorithme MDS permet une interprétation géométrique de la distance entre les attitudes basée sur la corrélation calculée entre les vecteurs de F_0 . Le paramètre d'amplitude est décrit par l'amplitude RMS calculée pour une fenêtre de signal de 20 ms utilisant un recouvrement de 50%. Pour ce calcul, la valeur de chaque échantillon de signal est un nombre réel normalisé sur

[-1;+1]. Pour l'analyse statistique, nous considérons uniquement les valeurs correspondant aux parties voisées. Nous calculons ensuite le RMS moyen associé à chaque voyelle (3 par stimulus). La durée de chaque syllabe est mesurée en millisecondes sur la base de la segmentation manuelle.

5 Résultats de l'analyse statistique

Les différences prosodiques sont mesurées en fonction des attitudes et du sexe des locuteurs, en considérant séparément la F_0 , le RMS et la durée syllabique. Chaque stimulus est représenté par une série de 3 mesures de chaque paramètre. Pour chaque paramètre, nous réalisons un test RM-ANOVA, avec les facteurs attitude (16) \times position de la more (3) et le facteur inter-sujet du sexe (2). Pour la suite de cette section, $F(a,b)$ correspond à la statistique calculée lors du test. Nous utilisons la notation M pour désigner la valeur moyenne du paramètre considéré.

5.1 La fréquence fondamentale

Le test RM-ANOVA appliqué sur la F_0 révèle une forte interaction attitude \times position moraïque : $F(30,210)=8.12$, $p<0.05$. Cela signifie que le profil de F_0 sur la phrase change de manière significative en fonction de l'attitude. D'après la Figure 2(a), l'interaction porte principalement sur les deux premières mores. Ainsi, la plupart des attitudes sont associées à une F_0 plus basse sur la seconde voyelle (forme V) excepté pour les attitudes AUTH, DECL, QUES et (dans une moindre mesure) SINC et SEDU qui montrent des contours en forme de V inversé. Afin de mieux caractériser ce constat, nous avons calculé (en %) le rapport de la F_0 normalisée entre la première et la seconde voyelle pour chaque stimulus (cf. Figure 2(d)). Un test RM-ANOVA entre l'attitude (16) et le sexe (2) révèle une interaction significative ($F(15,225)=13.6$, $p<0.05$). Ainsi, AUTH(+26%), DECL (+17%), QUES (+34%) ont une pente de F_0 plus importante que pour les autres attitudes ($p<0.05$, après correction Bonferroni). SINC (+7%) montre également une augmentation de la F_0 plus importante que pour CONT (-14%) ($p<0.05$, après correction Bonferroni). De plus, la différence de F_0 entre les attitudes est en interaction avec le sexe du sujet ($F(30,210)=1.95$, $p<0.05$), ce qui indique qu'un profil de F_0 différent est utilisé par les hommes et par les femmes pour la même attitude. On constate également que cette interaction de F_0 n'est pas caractérisée par la première voyelle (attitude \times sexe : $F(15,240)=1.36$, $p=0.16$), la seconde voyelle (attitude \times sexe : $F(15,240)=1.83$, $p=0.03$) ou la troisième voyelle individuellement (attitude \times sexe : $F(15,240)=0.38$, $p=0.98$), mais par les différences de F_0 entre les voyelles. Comme le montre la Figure 2(d), la différence entre les sexes est particulièrement importante en ce qui concerne le rapport entre la seconde et la première voyelle pour l'ensemble des attitudes (attitude \times sexe : $F(15,225)=2.58$, $p<0.05$) mais aussi individuellement pour les attitudes ADMI (hommes : +5%, femmes : -11%, $p=0.03$), AUTH (hommes=+10%, femmes : +37%, $p<0.05$) et QUES (hommes=+49%, femmes=+23%, $p<0.05$; toutes : Fisher LSD posthocs).

5.2 L'amplitude

Le test RM-ANOVA sur les valeurs du RMS révèle un rôle important du sexe ($F(1,17)=24.5$, $p<0.05$) qui indique que les différences d'amplitude observées chez les locutrices sont plus

importantes que chez les locuteurs. On note aussi un effet important de la position moraique ($F(2,34)=88.51, p<0.05$) : les amplitudes des premières ($M=.00071$) et des secondes ($M=.00066$) mores sont plus importantes que pour la troisième ($M=.00034$). L'attitude joue également un rôle non négligeable ($F(15,255)=1.99, p<0.05$) : certaines attitudes sont réalisées avec une amplitude plus importante que d'autres pour toute syllabe et pour les deux sexes. Les attitudes ADMI ($M=.00063$), POLI ($M=.00060$), SINC ($M=.00060$) et WOEG ($M=.00059$) ont des amplitudes plus fortes que AUTH ($M=.00054$), CONT ($M=.00052$), IRRI ($M=.00052$), QUES ($M=.00054$), SEDU ($M=.00053$), SURP ($M=.00056$) et UNCE ($M=.00052$) (test de Fisher LSD sans correction Bonferroni, $p<.05$). On note également une interaction de la position moraique \times sexe ($F(2,34)=3.51, p=0.04$) telle que la seconde voyelle est en général d'amplitude plus faible pour les hommes que pour les femmes. L'interaction attitude \times sexe n'est pas significative ($F(15,255)=0.88, p=0.58$). Une interaction importante attitude \times position de la more ($F(30,510)=6.57, p<.05$) montre que les variations intonatives changent de position dans la phrase, selon l'attitude (cf. Figure 2(b)). De plus, comme pour la F_0 , les différences sont plus importantes lorsque nous observons le rapport RMS entre la seconde et la première more bien que la seconde more soit presque toujours d'amplitude plus faible que la première pour toutes les attitudes excepté AUTH, DECL and QUES. Contrairement à la F_0 , cet effet n'entre pas en interaction avec le sexe ($F(15,255)=0.33, p=0.99$ - attitude \times sexe en utilisant le test RM-ANOVA sur le rapport du RMS de la seconde sur première voyelle). Des formes similaires sont utilisées par les hommes et les femmes pour produire la même attitude (cf. Figure 2(e)).

5.3 La durée

Le test RM-ANOVA appliqué sur la durée des mores révèle un effet significatif du sexe ($F(1,17)=4.71, p<0.05$), pour lequel les femmes ($M=137$ ms) réalisent des mores plus longues que celle des hommes ($M=117$ ms). On note une influence de la position de la more sur sa durée ($F(2,34)=95.6, p<0.001$) : la dernière more ($M=188$ ms) est presque deux fois plus longue que la première ($M=99$ ms) ou la seconde more ($M=98$ ms). La relation de la durée avec l'attitude ($F(15,255)=13.0, p<.05$), montre que certaines attitudes sont produites avec un allongement plus important que pour d'autres, quels que soient la more ou le sexe. Les attitudes AUTH ($M=162$ ms), CONT ($M=166$ ms), DECL ($M=171$ ms), DOUB ($M=155$ ms), QUES ($M=147$ ms) et SURP ($M=165$ ms) montrent des durées plus longues que les autres ($p<.05$, avec correction Bonferroni). On note un effet significatif de l'attitude sur la durée de la première more ($F(15,255)=11.09, p<.05$), de la seconde more ($F(15,255)=9.48, p<.05$) et de la troisième more ($F(15,255)=13.08, p<.05$). On ne note aucune interaction position des mores \times sexe ($F(2,34)=1.57, p=0.22$), ni attitude \times sexe ($F(15,255)=1.17, p=0.29$). Cependant, comme mentionné précédemment, il y a une interaction significative attitude \times position moraique sur la durée ($F(30,510)=12.07, p<.05$) qui montre que pour différentes attitudes, les variations de durées ne sont pas identiques sur les mores successives (cf. Figure 2(c)). Comme pour la F_0 (mais pas pour le RMS), on observe une interaction entre position moraique \times sexe ($F(30,510)=2.43, p<.05$). Cette différence entre les sexes s'observe principalement sur la dernière more ($F(15,255)=1.89, p<0.05$) (cf. Figure 2(f)). Pour les autres mores, cette interaction est moins importante (première more ($F(15,255)=1.68, p=0.05$; deuxième more : $F(15,255)=1.04, p=0.4$).

5.4 Algorithme MDS

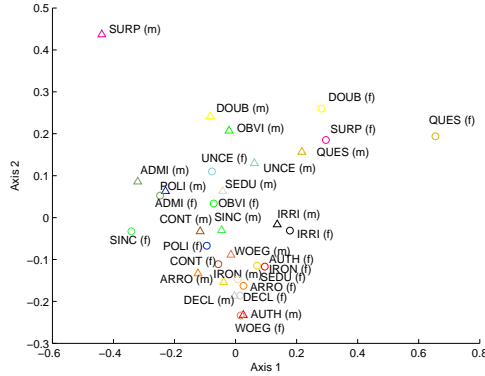


FIGURE 1 – Résultat de l'algorithme de MDS appliqué sur une fonction de corrélation entre les profils de F_0 (cercles : femmes, triangles : hommes).

Le résultat de l'algorithme MDS (Bronstein *et al.*, 2006) appliqué sur les profils F_0 moyens de chaque attitude en dissociant hommes et femmes est présenté sur la Figure 1. La distance entre chaque point est associée à la fonction de corrélation (Hermes, 1998) calculée entre les profils F_0 normalisés de chaque sujet. Pour les calculs, les profils F_0 des parties voisées ont été normalisés en durée (par rééchantillonnage) afin d'obtenir 100 mesures de F_0 associées à la durée moyenne de la phrase sur tout le corpus ($M=531.25$ ms).

6 Discussion

Cette étude sur la langue japonaise nous a permis d'explorer les caractéristiques acoustiques d'un ensemble d'attitudes non nécessairement spécifique de cette langue, dans le but d'identifier les similarités et les différences entre leurs profils prosodiques, basés sur les paramètres de F_0 , de RMS et de durée syllabique. Les résultats de l'analyse statistique montrent des interactions entre les attitudes et le sexe pour la F_0 . Cependant, nous avons vu que l'expressivité des femmes n'est pas seulement caractérisée par une F_0 (normalisée) plus élevée mais aussi par une amplitude plus élevée et des durées moraïques plus longues que celles des hommes. Cette observation est valable même si ces différences constituent plus un style lié au sexe que des variations de stratégie liées aux différentes attitudes. Nous avons également observé que les expressions de politesse (ADMI, POLI, SINC, WOEG) ont une amplitude plus élevée que les autres, y compris les expressions d'impolitesse (AUTH, CONT, IRRI). Cette observation diffère des résultats observés pour les mêmes situations chez des locuteurs d'anglais des États-Unis (Rilliard *et al.*, 2013), pour lesquels les attitudes véhiculant un trait sémantique d'imposition (AUTH, ARRO, CONT), sont réalisées avec une intensité plus élevée que les attitudes de politesse (POLI, SINC, WOEG). Ce constat ne vaut cependant pas pour l'attitude d'ADMI pour laquelle les deux groupes linguistiques montrent des stratégies comparables. Les résultats montrent d'importantes variations de durée moraïque parmi les attitudes, principalement entre les deux premières voyelles et la troisième

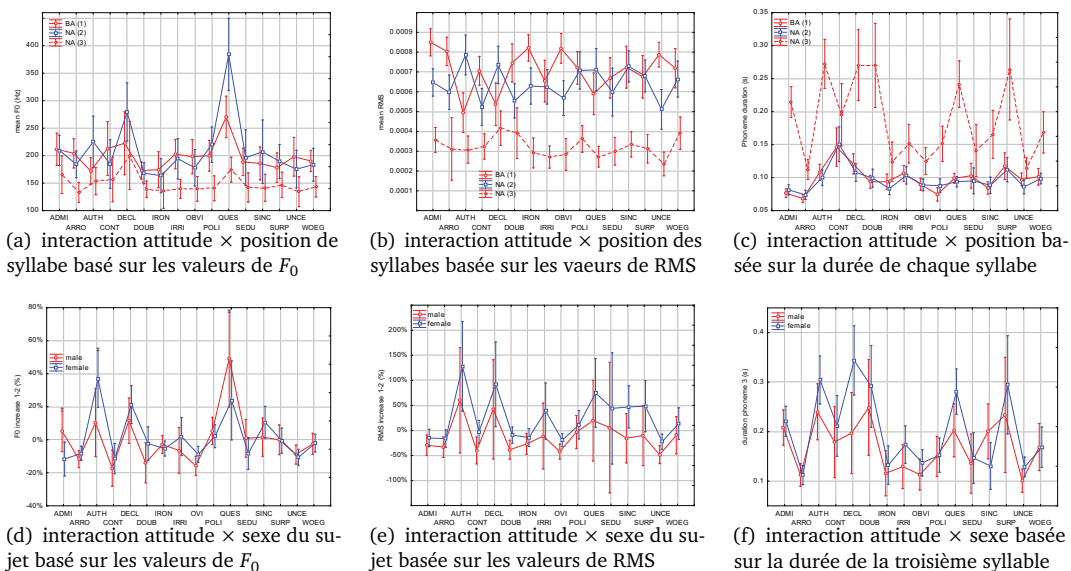


FIGURE 2 – 2(a) et 2(d) : Profils de F_0 en fonction de l'attitude et du sexe du sujet. 2(a) : F_0 moyenne (en Hz) pour chaque voyelle de « banana », pour les 16 attitudes. 2(d) : différences de la F_0 moyenne entre la première et la seconde voyelle (en % de croissance). 2(b) et 2(e) : Profils RMS entre les attitudes et le sexe. 2(b) : RMS moyen pour les 3 voyelles de « banana » pour les 16 attitudes. 2(e) : différence du RMS moyen entre la seconde et la première voyelle (en % de croissance). 2(c) : durée de chacune des 3 syllables de « banana » pour les 16 attitudes. 2(f) : durée de la troisième more pour les 16 attitudes.

voyelle (qui sont approximativement deux fois plus longue que les deux premières). Cela permet de déduire que les affects sociaux peuvent être à l'origine d'ajustement de durée basé sur la structure rythmique du japonais. Cela est en accord avec une étude antérieure de (Maekawa et Kagomiya, 2000) qui montre que la durée de la troisième voyelle est un facteur important permettant de caractériser l'expression de certaines attitudes. Le japonais standard est décrit comme une langue à accents toniques caractérisée par la montée de la F_0 au début de chaque phrase et marquée par une chute importante de la F_0 après une voyelle accentuée (Fujisaki et Sudo, 1971). La phrase « Banana » possède en japonais un accent lexical localisé sur la première more. Ainsi, nous nous attendons à observer une montée de la F_0 sur la première voyelle suivie d'une chute. Cependant, les résultats montrent que les contours de F_0 varient en fonction de l'attitude. Notre hypothèse est que l'instant où le pic de F_0 est atteint pour la première ou la seconde voyelle est lié au débit de parole ou à l'intensité. D'après nos données, il semble que certaines attitudes (AUTH, QUES et DECL) pour lesquelles le pic de F_0 arrive tardivement (sur la deuxième voyelle) ont des débits de parole plus lents. Cette position temporelle concorde avec un pic d'amplitude sur la deuxième voyelle, contrairement à la plupart des attitudes qui réalisent le pic ainsi que la chute de F_0 à l'intérieur de la première more (forme de V inversé).

L'analyse MDS basée sur la corrélation des profils de F_0 des attitudes nous permet d'identifier 3 catégories : les expressions de déférence, d'imposition et les expressions dubitatives (Brandt, 2008; Rilliard *et al.*, 2009). La première catégorie des expressions de déférence contient les

expressions ADMI, SINC, POLI (et SEDU uniquement pour les hommes) ainsi que deux attitudes (OBVI, UNCE) qui sont plus complexes à interpréter. La seconde catégorie d'imposition contient les attitudes AUTH, ARRO, IRON, IRRI ainsi que 2 autres (DECL, WOEG) qui ne correspondent pas sémantiquement aux autres expressions du groupe. Il est intéressant de noter que WOEG, qui correspond au concept de politesse de *Kyoshuku* se trouve dans la région des expressions d'imposition. On notera, d'après (Sadanobu, 2004; Rilliard *et al.*, 2014b,a), que cette expression de *kyoshuku* véhicule un trait sémantique d'imposition, certainement véhiculé par l'expression prosodique. Cela explique aussi les résultats d'études antérieures sur cette expression (Shochi *et al.*, 2006; Rilliard *et al.*, 2009), pour laquelle des incompréhensions sont observées chez des auditeurs ne parlant pas japonais. Il faut aussi noter que l'expression de WOEG se distingue de l'impolitesse par ses caractéristiques de qualité vocale spécifiques (Rilliard *et al.*, 2009). Malgré des profils de F_0 proches entre hommes et femmes pour l'attitude SEDU, d'après l'analyse RM-ANOVA, (cf. Figure 2(d)), l'analyse de MDS dissocie les performances de ces deux groupes (Figure 1) : les expressions des femmes se trouvent prosodiquement plus proches de la catégorie d'imposition, alors que les performances masculines sont groupées avec les expressions de déférences. On pourra apprécier les potentielles interprétations de ce résultat, que nous n'aurons pas la place de développer. Nous noterons tout de même que la qualité de voix devrait ici aussi jouer un rôle majeur (présence d'un souffle important chez les femmes). Dans nos prochains travaux, nous effectuerons une étude contrastive similaire entre différentes langues.

Remerciements

Merci à Donna Erickson pour ses conseils avisés ainsi qu'à Mariko Kondo et Sylvain Detey pour l'enregistrement du corpus à l'université de Waseda. Cette recherche est financée par le projet ANR PADE et le projet bordelais PEPS IDEX/CNRS Mavoix.

Références

- BOERSMA, P et WEENINK, D. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5(9/10):341–345.
- BRANDT, P A. (2008). Thinking and language. a view from cognitive semio-linguistics. *In Proc. Speech Prosody*, pages 649–654.
- BRONSTEIN, A. M., BRONSTEIN, M. M. et KIMMEL, R. (2006). Generalized multidimensional scaling : a framework for isometry-invariant partial surface matching. *Proc. Natl. Acad. Sci. U.S.A.*, 103(5):1168–1172.
- CAMACHO, A. et HARRIS, J. G. (2008). A sawtooth waveform inspired pitch estimator for speech and music. *Journal of the Acoustical Society of America*, 124:1638–1652.
- de MORAES, J. A. (2008). The pitch accents in Brazilian Portuguese : analysis by synthesis. *In Proc. Speech Prosody*, pages 389–397.
- FÓNAGY, I., BÉRARD, E. et FONAGY, J. (1984). Clichés mélodiques. *Folia Linguistica*, 17:153–185.
- FUJISAKI, H. et HIROSE, K. (1993). Analysis and perception of intonation expressing paralinguistic information in spoken Japanese. *In Proc. ESCA Workshop on Prosody*.

- FUJISAKI, H. et SUDO, H. (1971). A model for the generation of fundamental frequency contours of Japanese word accent. *Journal of acoustical society of Japan*, 27(9):445–453.
- GU, W., ZHANG, T. et FUJISAKI, H. (2011). Prosodic analysis and perception of Mandarin utterances conveying attitudes. In *Proc. Interspeech*, pages 1069–1072.
- HERMES, D. J. (1998). Measuring the perceptual similarity of pitch contours. *Journal of Speech, Language and Hearing Research*, 41(1):73.
- MAEKAWA, K. et KAGOMIYA, T. (2000). Influence of paralinguistic information on segmental articulation. In *Proc. 6th International Conference on Spoken Language Processing*, pages 349–352.
- MORAES, J. A. et RILLIARD, A. (2014). *Spoken Corpora and Linguistic Studies*, chapitre Illocution, Attitudes and Prosody. John Benjamins, Amsterdam. (à paraître).
- MORLEC, Y., BAILLY, G. et AUBERGÉ, V. (2001). Generating prosodic attitudes in French : Data, model and evaluation. *Speech Communication*, 33(4):357–371.
- RILLIARD, A., de MORAES, J., ERICKSON, D. et SHOCHI, T. (2014a). Social affect production and perception across languages and cultures – the role of prosody. *Leitura*, 52. (à paraître).
- RILLIARD, A., ERICKSON, D., de MORAES, J. et SHOCHI, T. (2014b). *Linguistic approaches to emotions in context*, chapitre Cross-Cultural Perception of some Japanese Expressions of Politeness and Impoliteness, pages 251–276. John Benjamins, Amsterdam.
- RILLIARD, A., ERICKSON, D., SHOCHI, T. et MORAES, J. A. D. (2013). Social face to face communication - American English attitudinal prosody. In *Proc. Interspeech*.
- RILLIARD, A., SHOCHI, T., MARTIN, J.-C., ERICKSON, D. et AUBERGÉ, V. (2009). Multimodal indices to Japanese and French prosodically expressed social affects. *Language and speech*, 52(2–3):223–243.
- SADANOBU, T. (2004). A natural history of Japanese pressed voice. *Journal of the Phonetic Society of Japan*, 8(1):29–44.
- SHOCHI, T., AUBERGÉ, V. et RILLIARD, A. (2006). How prosodic attitudes can be false friends : Japanese vs. French social affects. In *Proc. Speech Prosody*, pages 692–696.
- SHOCHI, T., RILLIARD, A., AUBERGÉ, V. et ERICKSON, D. (2009). *The role of prosody in affective speech*, volume Linguistic Insights 97, chapitre Intercultural perception of English, French and Japanese social affective prosody, pages 31–59. Peter Lang.
- WICHMANN, A. (2000). The attitudinal effects of prosody, and how they relate to emotion. In *Proc. ISCA Workshop on Speech and Emotion*, pages 143–148.
- WIERZBICKA, A. (1985). A semantic metalanguage for a cross-cultural comparison of speech acts and speech genres. *Language in Society*, 14(4):491–513.
- WIERZBICKA, A. (1992). Defining emotion concepts. *Cognitive Science*, 16:539–581.
- WIERZBICKA, A. (2005). Empirical universals of language as a basis for the study of other human universals and as a tool for exploring cross-cultural differences. *Ethos*, 33(2):256–291.