HAL
open science

Analyse par synthèse des contours intonatifs en anglais britannique: étude de la relation entre l'annotation des fromes et fonctions prosodiques

Saandia Ali

To cite this version:

Saandia Ali. Analyse par synthèse des contours intonatifs en anglais britannique: étude de la relation entre l'annotation des fromes et fonctions prosodiques. Rencontres Jeunes Chercheurs en parole RJCP 2007, Jul 2007, Paris, France. hal-00364097

HAL Id: hal-00364097

https://hal.science/hal-00364097v1

Submitted on 25 Feb 2009

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 par synthèse des contours intonatifs en anglais britannique : Etude de la relation entre annotation des formes et fonctions prosodiques

Saandia Ali

CNRS, Laboratoire Parole et Langage, Université de Provence
29 avenue Robert Schuman, 13100 AIX EN PROVENCE
Courriel : saandia.ali@lpl.univ-aix.fr

ABSTRACT

This paper presents a general model for the relation between representations of form and function for speech prosody on a multi-lingual basis. It outlines a procedure for analysing prosody by synthesis - generating formal representations from a minimal representation of prosodic functions and comparing the output with the observed data. This then allows the functional representation to be enriched and to test whether it provides a closer fit to the data. The procedure is specifically applied to the intonation patterns of British English. Five successively more complex models of mappings are presented and applied to fifteen continuous passages from the Eurom1 corpus. The quality of fit of the models is finally measured by linear correlation with hand corrected modelled fundamental frequency curves, intonation unit by intonation unit.

1. INTRODUCTION

L'objectif principal de ce travail de recherche est d'étudier la relation entre la représentation des formes et fonctions prosodiques en anglais britannique [Arm26], c'est-à-dire que l'on s'intéresse d'une part aux fonctions attribuées à la prosodie et à la meilleure façon de représenter toutes ses formes significatives. Les théoriciens de l'anglais [Crutt86] s'accordent à dire que l'intonation tient une place très importante dans la communication parlée en tant que marqueur de l'attitude du locuteur, signal de l'alternance des tours de parole etc.. Elle contribue largement à la production et à l'interprétation des énoncés. Il est toutefois difficile de répertorier la totalité des fonctions remplies par la prosodie et la multiplicité des termes employés par les différents chercheurs montre qu'il est encore délicat de définir ce qui est actuellement véhiculé par l'intonation et surtout par l'intonation uniquement par opposition aux autres canaux que sont le lexique, la grammaire, la sémantique, la pragmatique, les phénomènes paralinguistiques…Au niveau formel, de même il existe multiples modèles des formes prosodiques de l'anglais qui s'opposent par les présupposés théoriques par lesquels ils sont contraints. On distinguera ainsi, par exemple, l'école britannique de l'intonation [O'Con61] de la théorie auto segmentale américaine [Sil92]. On présente ici une procédure d'analyse par la synthèse pour tenter d'apporter des réponses à ces questions.

2. NIVEAUX DE REPRÉSENTATIONS

On se propose de caractériser la relation entre la représentation des formes et des fonctions prosodiques en utilisant l'analyse par la synthèse [Hir05]. Il s'agit ainsi de générer des représentations formelles à partir d'une représentation minimale des fonctions prosodiques et de comparer le résultat obtenu avec les données observées. Cela permet ensuite d'enrichir la représentation fonctionnelle et de voir si cela permet une meilleure reproduction des données.

Cinq modèles successifs et de plus en plus complexes du « mapping » entre formes et fonctions sont présentés et mis en œuvre sur 15 passages du corpus Eurom1. On évalue enfin la qualité de la prédiction des modèles en calculant la corrélation linéaire avec la courbe de F0 modélisée et corrigée manuellement pour chaque unité intonative. On présuppose pour cette analyse, trois niveaux intermédiaires de représentation entre le niveau du signal acoustique et le niveau fonctionnel [Hir00] : la représentation phonétique, la représentation phonologique de surface et la représentation phonologique sous-jacente. La représentation phonétique est fournie par l'algorithme Momel qui permet d'obtenir une représentation automatique discrète de la courbe brute de la fréquence fondamentale sous forme d'une séquence de points cibles. Les points cibles ainsi obtenus sont convertis automatiquement en symboles tonals selon l'alphabet INTSINT [Hir98] (représentation phonologique de surface). Enfin l'annotation fonctionnelle IF [Hir77] représente le niveau le plus abstrait. On considère deux niveaux d'analyse : l'unité intonative caractérisant les types de frontières (terminale/ non terminale) et l'unité tonale pour annoter les niveaux de proéminence (inaccentué, accentué, nucléaire, emphatique).

3. MODÈLES DE SYNTHÈSE DE L'INTONATION

Il s'agit d'une approche incrémentale comparant à chaque niveau d'approximation le résultat de la prédiction et les patrons intonatifs observés dans les données. A ce niveau d'expérimentation, l'annotation fonctionnelle est faite manuellement. On utilise ici 15 passages du corpus Eurom1 (chacun contenant 5 phrases) lus par trois locuteurs. Les règles de dérivations de l'annotation fonctionnelle vers l'annotation formelle par l'alphabet INTSINT sont exécutées dans un script Praat. Les symboles INTSINT sont ensuite convertis en points cibles que l'on interpole quadratiquement afin d'obtenir un

« pitch-tier » utilisable dans Psola pour la resynthèse. Le segment ainsi resynthétisé peut faire l'objet d'une évaluation subjective [Hir05].

Dans le reste de cette section, les différents modèles successifs sont présentés.

3.1. Modèle nul

Le modèle « nul » s'appuie sur l'hypothèse que l'intonation n'est dérivée d'aucune information fonctionnelle. La prédiction des tons se limite à une montée initiale de M (mid) à T (top) suivie d'un abaissement progressif de F0 tout au long du passage pour atteindre le ton B. La figure 1 montre la courbe du modèle superposée à la courbe dérivée des points cibles de Momel pour un passage.

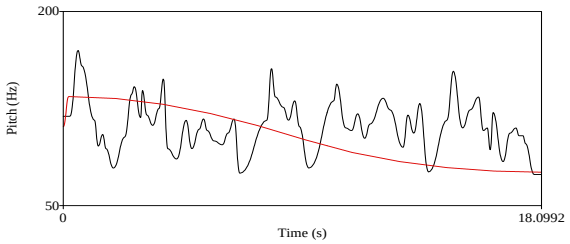

Figure 1: Courbe de F0 dérivée de l'annotation INTSINT [MH…..B] superposée à la courbe quadratique interpolée crée par Momel pour le passage Fao30072 du corpus Eurom1.

3.2. Modèle IU

De toute évidence le contour intonatif du passage illustré dans la figure 1, ne consiste pas en un seul contour continu mais plutôt en une séquence de contours distincts. Ces contours sont représentés, dans une première approximation, par des chutes. Le texte est donc annoté grâce à des symboles de frontière ([ ]) le divisant en Unité intonatives. Chaque unité intonative est alors modélisée en une séquence de tons [M-T…..B] où M et B sont alignés avec les frontières gauches et droites de la séquence. T est aligné de façon fixe à 200ms du début de l'unité.

La représentation fonctionnelle a été effectuée sous Praat dans un textgrid où chaque intervalle correspondait à une unité intonative permettant de générer les contours les plus proches des points cibles manuellement corrigés.

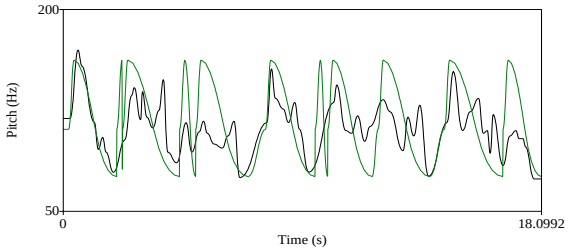

Figure 2 : La même courbe que dans la figure 1 superposée au résultat de l'approximation du modèle IU

3.3. Modèle terminal

Dans le modèle IU, toutes les unités intonatives sont représentées par un contour descendant alors que beaucoup d'entre elles ont un contour globalement descendant et une montée finale. Fonctionnellement, cela permet de distinguer le caractère final ou terminal du caractère non terminal d'un énoncé (cf : Tune 1 et tune 2 [Arm26] ou terminal +/- [Hir77]). Les séquences INTSINT générées étaient alors [M-H….B-B] ou [M-H….B-H].

3.4. Modèle accent

Dans le modèle accent, on introduit la notion de proéminence. Bien qu'en anglais il existe un accent lexical, tous les accents lexicaux ne portent pas forcément une proéminence supérieure au niveau de la phrase (accent de phrase ou stressed vs accented). Ces accents de phrase peuvent aussi parfois marquer des syllabes non accentuées lexicalement (bien que ce ne soit pas le cas dans ce corpus). Les accents de phrase sont représentés dans Praat sur un second intervalle (Praat interval tier) nommée Tonal Unit (unité tonale) suivant une tradition remontant à [Jass62].

Les étiquettes INTSINT sont distribuées comme indiqué ci-dessous : le ton M est aligné avec le début de chaque IU et soit H soit B avec la fin de l'IU. Le ton H est aligné avec le début du premier TU de chaque IU qui est annoté ['] et l'étiquette D est alignée avec les TU suivants. Un ton B est aligné après le dernier H ou D du dernier TU étiqueté ['].

3. 4. 1. Nucleus

Adoptant l'annotation fonctionnelle de [Hir77], on introduit ensuite la notion de nucleus ou centre intonatif. La structure intonative que l'on décrit alors est familière dans la tradition britannique de l'intonation ([O'con61], [Crutt86] où (X) représente un élément optionnel) :

(prehead) (head body) nucleus (tail)
She's going to the cinema after shool.
Les tons associés à cette structure sont alors :
(prehead) (head(body…….)) nucleus(tail)

[	'	'	*	'	|/+
M	H	D	D/H	-B	B/H

Ou de manière plus compacte :
[M (ʻT (ʻD)) *H/D ʻ-BB|/H+

où le choix entre H/D dépend du fait qu'il y a ou non une tête ou head précédent ce ton. Le choix entre B/H dépend de l'étiquette terminale ou non terminale de l'IU.

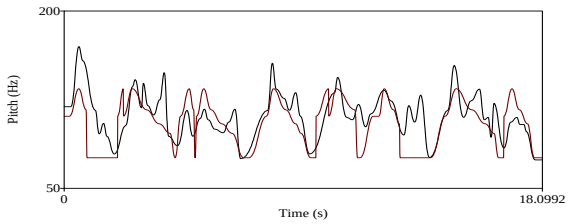

Figure 3 : La courbe de F0 générée par momel superposée au résultat du modèle accent.

3.5. Modèle emphatique

3.5.1. Le nucleus emphatique

La dernière fonction considérée est l'emphase notée [!*] Elle permet d'abord de distinguer un nucléus ou centre intonatif marqué d'un nucléus non marqué (emphatique ou non emphatique).

Cela conduit à décrire les quatre contours définis par [Arm26] et le système d'annotation fonctionnelle IF [Hir77] et [Hir98]).La caractéristique de base d'un nucléus emphatique est la réalisation de son point de départ à une valeur de F0 supérieure à la hauteur des accents précédents.

3.5.2. Tonique de tête emphatique

Le nucléus emphatique terminal ou non terminal peut-être pré signalé au niveau de la tête de l'énoncé [Hir98]. On peut alors noter le caractère emphatique au niveau du TU pour le nucleus ou au niveau de l'IU pour une tête emphatique.

On utilise deux séquences de tons INTSINT : soit une suite de chutes (T+L, H+L H+L….) pour un IU non terminal, soit une séquence ascendante (BUUU..) pour un IU terminal.

3. 5. 2. Cas particuliers

Certains phénomènes intonatifs peuvent être annotés avec une utilisation non courante des frontières intonatives. [Hir98] Cela signifie qu'une unité intonative pourrait n'avoir qu'une frontière initiale [mais aucune frontière finale (cf. : the « interrupted glide down » de [Kin58]) ou le cas contraire comme dans l'intonation des parenthèses, les commentaires ou « afterthoughts », le « divided fall-rise ». la dernière modélisation contient les catégories fonctionnelles suivantes :

{ [, +, |, +, [|, [+, [!, [!+, [!|}

Tonal unit: { 0, ', *, !* }

Cela permet de définir les contours intonatifs suivants :

(i) *no nucleus initial*:

[M ('T ('D_1))

(ii) *neutral terminal:*

[M ('T ('D_1)) *H/D-B B|

(iii) *neutral non-terminal:*

[M ('T ('D_1)) *H/D-B H+

(iv) *emphatic terminal*

[M ('T ('D_1)) !*T -B B|

(v) *emphatic non-terminal*

[M ('T ('D_1)) !*T-B H+]

(v) *marked emphatic terminal*

[!T ('B ('U_1)) !*T-B B|

(vi) *marked emphatic non-terminal*

[!M ('TL ('HL_1)) !*T-B H+

(vii) *no nucleus terminal*

('B ('B)) B|

(viii) *no nucleus non-terminal*

('B ('B)) H+

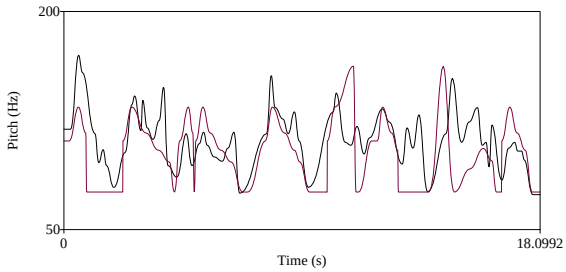

Figure 4 : La courbe issue du modèle emphatique superposée à la courbe issue de Momel..

Ci dessous un exemple de l'annotation fonctionnelle complète de l'un des passages.

[Mu'nicipal *Fire 'Service 'speaking| # [We're 'trying to lo'cate an e'mergency *caller+ [who 'rang *off+ # [wi'thout 'giving any 'personal *details| # [He ap'peared to 'be on the 'local *network| # [He con'nected on our !*line 'number+ # ['seven six *two+ 'five 'eight *four| # [!We'd ap'preciate im'mediate at'tempts to *trace him| [be'cause he 'sounded *desperate|

Figure 5 : Passage fao30072 avec l'annotation fonctionnelle utilisée dans l'analyse.

La première unité intonative de ce passage sous la forme d'un texgrid est montrée dans la figure 6.

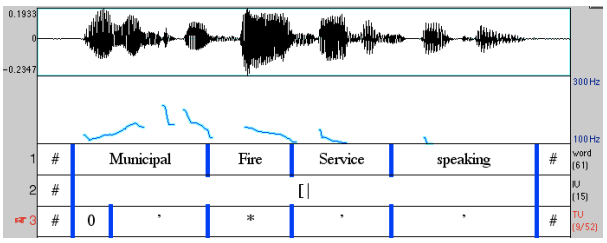

Figure 6 : Un extrait du textgrid correspondant au premier IU du passage ci-dessus.

4. Evaluation

Les cinq modèles décrits précédemment ont été mis en application grâce à un script de Praat sur les 15 passages lus par le locuteur fa du corpus.

La courbe brute de la fréquence fondamentale des enregistrements a été modélisée avec l'algorithme Momel. Les paramètres spécifiques à chaque locuteur (tessiture du locuteur : « key and range ») ont été estimés en utilisant

l'algorithme INTSINT. Ici on garde les mêmes valeurs pour les 15 passages : key= 111, range= 1.1 octaves). On a ensuite calculé pour chaque IU, le coefficient de corrélation linéaire entre la courbe de Fa générée par Momel et la courbe issue des différents modèles.

.

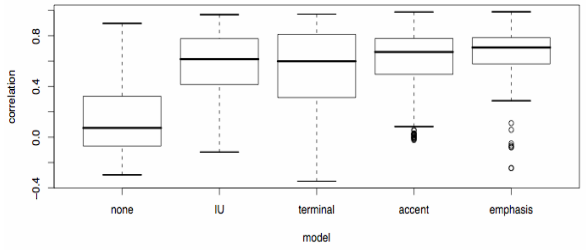

Figure 7 : coefficients de corrélation pour chaque IU entre le résultat des modèles et les points cibles de Momel

5. Conclusion

Le meilleur coefficient de corrélation obtenu dans cette expérience n'est pas particulièrement élevé (r =0.66) (bien que certaines valeurs individuelles soient beaucoup plus élevées supérieures à 0.9 parfois). Cependant, à ce stade de notre recherche, il n'y a pas eu de tentative d'optimisation des paramètres du modèle. Il s'agit ici de présenter une méthode de recherche et d'analyse. La prochaine étape devrait conduire à une annotation fonctionnelle totalement automatique d'une part et à une représentation optimisée en terme de forme d'autre part.

5. Références

[Arm26] Armstrong, L.E and I. Ward 1926. *A Handbook of English Intonation*. Second edition. Heffer. Cambridge

[Boe06] Boersma, P & Weenink, D. 2006. Praat Doing phonetics by computer. [computer program]. Version 4.5.06 http://www.praat.org/.

[Chan95] Chan, D, Fourcin, A, Gibbon, D, Granstrom, B, Huckvale, M, Kokkinakis, G, Kvale, K, Lamel, L, Lindberg, B, Moreno, A, Mouropoulos, Senia, F, Trancoso, I., Veld, C & Zeiliger, J. 1995. EUROM- A Spoken Language Resource for the EU, in *Proceedings of Eurospeech'95*. (Madrid, Spain, September, 1995). (1), 867-870

[Crutt86] Cruttenden, A 1986. *Intonation*. Cambridge University Press

[Hir05] Hirose, K, Hirst, D.J. & Sagisaka, Y. (eds.) 2005. *Quantitative prosody modeling for natural speech description and generation* (=*Speech Communication*, 46 (3-4)

[Hir77] Hirst, D.J. 1977. *Intonative Features. A Syntactic Approach to English Intonation*. (Mouton La Haye).

[Hir98] Hirst, D.J. 1998. Intonation in British English. in Hirst & Di Cristo (eds) 1998. 56-77

[Hir05] Hirst, D.J. 2005. Form and function in the representation of speech prosody. In Hirose et al. (eds) 2005, 334-347.

[Hir98] Hirst, D.J, Aubergé, V, Rilliard, A. 1998. Comparison of subjective evaluation and an objective evaluation metric for prosody in text-to-speech synthesis. *Proceedings ESCA/COCOSDA workshop on Speech Synthesis*, Jenolan, Caves, Australia. November 1998, 1-4

[Hir05] Hirst, D.J, Auran, C. 2005. Analysis by synthesis of speech prosody: the ProZed environment. *Proceedings of Interspeech/Eurospeech 05. 9th European Conference on Speech Communication and Technology*, September 2005, Lisbon. 3225-3228

[Hir98] Hirst, D.J. & Di Cristo, A. (eds) 1998. *Intonation Systems. A survey of Twenty Languages.* (Cambridge, Cambridge University Press).

[Hir00] Hirst, D.J, Di Cristo, A. & Espesser, R. 2000. Levels of representation and levels of analysis for intonation. in M. Horne (ed) *Prosody : Theory and Experiment*. Kluwer Academic Publishers, Dordrecht. 51-87

[Jass62] Jassem, W. 1962. *Intonation of Conversational English. (Educated southern English).* Travaux de la société des sciences et des letters de Wroclaw seria A. Nr 45

[Kin58] Kingdon, R. 1958. *The Groundwork of English Intonation*. London, Longman

[O'Con61 O'Connor, J.D. and Arnold, G.F. 1961. *Intonation of Colloquial English*. London: Longman (2nd edition, 1973).

[Sil92] Silverman, K, Beckman, M, Pitrelli, J, Ostendorf, M, Wightman, C, Price, P, Pierrehumbert, J., Hirschberg, J. 1992. ToBI : a Standard for Labelling English Prosody. *Proceedings ICSLP92*, 2, 867- 870, Banff, Canada