



HAL
open science

Caractérisation des variations individuelles du contour de fréquence du fondamental observées dans des phrases lues en anglais

Jacqueline Vaissière

► **To cite this version:**

Jacqueline Vaissière. Caractérisation des variations individuelles du contour de fréquence du fondamental observées dans des phrases lues en anglais. 20ème Journées d'Étude sur la Parole, 1994, Trégastel, France. pp.87-90. halshs-00703581

HAL Id: halshs-00703581

<https://shs.hal.science/halshs-00703581>

Submitted on 24 Jun 2012

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.

CARACTÉRISATION DES VARIATIONS INDIVIDUELLES DU CONTOUR DE FREQUENCE DU FONDAMENTAL OBSERVÉES DANS DES PHRASES LUES EN ANGLAIS.

JACQUELINE VAISSIÈRE
Institut de Phonétique, CNRS-UA 1027
75005 Paris, France

et
ATR Interpreting Telephony Research Laboratories
Kyoto, Japon.

Résumé:

Le but de cette présentation est de décrire une méthode de caractérisation des courbes du fondamental (Fo) à trois étapes, permettant de capturer les différences individuelles. Premièrement, la phrase est segmentée sur la base du passage de Fo par la ligne de base (BL) en "unités de sens", nommés HP (par analogie avec "hat-pattern"). Une "unité de sens" est composée des mouvements suivants: [montée + (plateau P) + descente + (Montée de continuation, Rc)]. Deuxièmement, la distance acoustique entre HPs successifs est estimée selon les critères suivants: (i) présence ou absence d'une pause, précédée ou non, sur la dernière syllabe avant la pause, par Rc ou une valeur extra-basse, (ii) interruption de la tendance à décliner régulièrement: soit un rehaussement ("resetting") ou un nivellement, (iii) un pic de Fo plus élevé après la frontière et/ou un abaissement du pic de Fo avant la frontière. Troisièmement, le degré de jointure entre les mots à l'intérieur d'un HP est estimé selon l'importance de la rupture de la continuité du Fo sur P: présence d'un Rc, ou pic de Fo sur l'une des syllabes accentuées. La division de la phrase en HPs est sensiblement la même pour les locuteurs. Les différences de degrés de frontières marqués entre HPs successifs, et les sous-divisions des HPs, et surtout les procédés les plus fréquemment employés permettent de comparer les productions des locuteurs. L'étude du Fo de 1000 phrases nous a permis de redécouvrir la validité du principe du HP pour la parole lue, et de mettre en évidence des "habitudes prosodiques" (séquences particulières de mouvements de Fo) propres à chacun des locuteurs.

Mots-clés: *prosodie, anglais, différences entre locuteurs, modélisation*

INTRODUCTION

Les variations de fréquence du fondamental (Fo), durée et intensité, sont contrôlées en partie par des règles d'origine linguistique et qui sont différentes dans les langues à tons, les langues à accent de mot contrastif, les langues où la position de l'accent lexical est fixe par rapport aux frontières de mot, ou dépend de la structure morphémique, ou de la catégorie grammaticales des mots, etc. De plus, les marques prosodiques d'intonation dépendent de la coexistence éventuelle de marqueurs morphologiques ou de

transformations syntaxiques (pronom interrogatif ou de l'inversion du sujet pour marquer une question, par exemple).

Les variations dépendent également de phénomènes plus généraux, relevant des particularités du système de perception ou de production (Collier, 1984; Vaissière, soumis par publication), indépendants de la langue en question. Le cantonnement à des explications relevant de la linguistique ou à la phonologie particulière d'une langue peut aboutir à un obscurcissement des problèmes (Voir Garde, p. 5).

Plus on dispose de données sur diverses langues, et plus on est frappé par les similarités prosodiques entre les langues (Vaissière, 1982): on peut donner pour exemples l'utilisation des pauses pour la segmentation, l'adaptation faite à d'autres langues de la célèbre formule de Dennis Klatt pour la génération de la durée en anglais, et la tendance de Fo à décliner tout au long des phrases, dans les langues à tons (Meyers, 1974) ou autres (Cohen et t'Hart, 1967), Fo plus élevée après les consonnes sourdes, etc. Selon notre hypothèse, des propriétés acoustiques, sans doute incontrôlées au départ, d'origine physiologique, sont devenues des réalités psychologiques: la tendance à la déclinaison devient une caractéristique nécessaire pour une synthèse correcte des phrases déclaratives (Abramson, 1978 sur le Thaï), et elle est prise en compte par l'auditeur pour la comparaison de la hauteur des pics (Pierrehumbert, 1979, sur l'anglais). Ces tendances, dont on peut faire l'hypothèse qu'elles sont les mêmes pour toutes les langues, peuvent être phonologisées ou non dans une langue donnée..

Malgré l'abondance des recherches, il manque toujours un modèle de notation des courbes de Fo permettant de mettre en valeur les différences entre les langues. Le problème est qu'il n'existe pas de modèle unique pour une langue donnée, et plus une langue est étudiée, et plus les modèles abondent et se contredisent, du moins, en partie. Notre approche de description est très fortement marquée par nos expériences sur l'utilisation de la prosodie en reconnaissance, où on se trouve confronté par les différences entre locuteurs. Nous ne croyons pas au "modèle miracle", satisfaisant à la fois les besoins de la synthèse (codage des effets combinés du contexte, de la syntaxe et du rythme), ceux de la reconnaissance (informations à extraire automatiquement, Vaissière, 1985), des linguistes (données permettant de renforcer ou d'infirmer telle ou telle hypothèse), de l'orthophoniste et du professeur de langue (notation des erreurs prosodiques), etc. D'autres

méthodes sont actuellement développées: le système TOBI pour l'anglais, basé à la fois sur l'inspection visuelle de la courbe de Fo et l'audition des phrases, actuellement en cours de développement et qui représente ce sur quoi un groupe d'experts reconnus en prosodie a réussi à s'entendre, le système automatique de stylisation des courbes de Fo développé à Aix, (Hirst et Di Cristo (1993), etc. Il reste à tester leur capacité à décrire les différences entre les locuteurs.

Le français, avec ses divers degrés de continuation et de finalité marqués par Fo, et d'allongements finals (Delattre, 1962) illustre comment une langue peut jouer des paramètres (Fo, durée) pour marquer différents degrés de frontière. En anglais, les paramètres prosodiques sont fortement influencés par la réalisation des accents de mots, mais de nombreux articles ont montré que la différence de degré entre accents de mots, la réalisation d'allophones initiaux, et l'allongement final, contribuaient aussi à marquer des frontières de degrés différente entre les mots. L'aspect démarcatif de Fo est-il plus difficile à mettre en évidence en anglais qu'en français? La structure prosodique est-elle la même en anglais qu'en français?

1. CORPUS, LOCUTEURS ET MATÉRIEL

Nous avons utilisé une partie du corpus sur l'anglais disponible sur le réseau informatique de la compagnie ATR, à Kyoto, 200 phrases prononcées par 5 locuteurs (4 britanniques et 1 américain). L'affichage du spectrogramme, de la segmentation, du Fo et l'écoute d'une liste de phrases préalablement définie est possible en un temps très rapide (utilisation du logiciel Xwaves), et ces facilités ont rendu possible l'étude d'un grand nombre de phrases, dans d'excellentes conditions.

2. DESCRIPTION

Rappel de la notion de hat-pattern

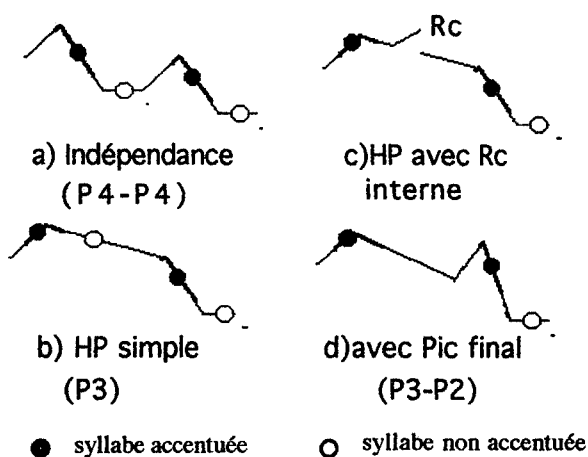


Figure 1a et b: Principe du groupement par hat-pattern: la séquence [montée+descente + montée+descente] correspondant à deux mots (en a) est

réduite en la séquence [montée + descente] (en b), correspondant aux 2 mots regroupés.

Figure 1c et 1d: Décomposition d'un hat pattern: par une montée de continuation (c) ou un pic sur le plateau (d). La ressemblance avec les patterns français a été indiquée.

Notre description est basée sur le principe bien connu du HP (expression reprise de l'école hollandaise de prosodie, Cohen et t'Hart, et appliquée à l'anglais par Maeda, Maeda, 1975). Le principe est fort simple. La figure 1a schématise la réalisation de deux accents successifs (schématisation inspirée de Delattre, 1962). La figure 1b illustre le principe de regroupement par HP: le premier mot tend à recevoir un mouvement montant (ton haut ou parfois montant sur la syllabe accentuée), et le second une descente de Fo. Les syllabes inaccentuées centrales tendent à rester sur le registre Haut, et la forme générale de la courbe de Fo est celle d'un chapeau (procédé 1). Une autre réalisation consiste à réaliser les deux mouvements opposés dans le premier mot, et à désaccentuer le second (procédé 2).. Le nombre de HPs dans une phrase correspond grosso modo à sa division en groupes de sens. Ces deux principes de regroupement nous semblent relever de principes éthologiques (voir Vaissière, en préparation) et seraient en conséquent des tendances universelles (à démontrer).

La figure 4 illustre des différences entre locuteurs. le loc 1 (L1) a regroupé "the candidate felt" et "trapezoidal badge", L1 par le procédé 1, et les deux autres locuteurs ont utilisé le procédé 2. Quand l'accent lexical du premier mot est situé après les deux premières syllabes du mot ("trapezoidal"), la montée tend à correspondre à la première syllabe du mot (pour les variations entre locuteurs sur ce point, voir Vaissière, 1985- la même tendance a été observée en Espagnol, Vaissière 1983). (Seuls 3 locuteurs sont illustrés sur la figure 4 par manque de place).

2.A. Division "manuelle" en HP.

La figure 4 illustre le résultat d'une décomposition visuelle en HP par passage par BL (même phrase prononcée par trois locuteurs, L1, L2 et L3, ([indique la retour à la BL).

Par exemple, pour L1, les regroupements pourraient se décrire de la façon suivante:

"The candidate felt [BL+Rc] that a trapezoidal badge [BL+Rc] would have more impact [BL+Rc] than [BL] the usual [BL] rosette" (6 HPs)

2.B. Division automatique en HP

L'automatisme est un fait hautement désirable pour l'étiquetage de grandes bases de données. Comme la pente générale de BL peut changer d'une phrase à l'autre pour un même locuteur (en fonction de sa durée, de son rang dans la liste, de la fatigue du locuteur, etc.) et même à l'intérieur de la phrase, on a cherché un critère de détection automatique des passages par BL pouvant s'appliquer sur la base d'une seule phrase. Cette recherche s'est avérée vaine.

La solution qui nous a paru la meilleure, mais qui nécessite quand même une vérification manuelle des

résultats, est la suivante et s'opère en deux temps. Premièrement, à partir du Fo de la première syllabe, on essaie de tracer une droite, la plus longue possible, qui suive les mouvements de Fo de la suite sans les croiser (figure 2a). Un croisement indique que la BL est atteinte (par exemple, l'intersection de la courbe de Fo avec les droites 3, 5, 6 et 7 sur notre schématisation), et on recommence à partir de ce dernier point, qui devient le premier point du tronçon du HP suivant. Le nombre de passage par BL permet de compter les HPs. Dans un deuxième temps, en partant de chaque passage par BL, on trace une droite vers la gauche pour atteindre les valeurs plus basses du début du HP, sans jamais croiser la courbe de Fo (prise en compte du rehaussement). Les tronçons de lignes pleines (étape 1) et les tronçons de lignes pointillées (étape 2), à la figure 4, illustrent schématiquement ce qu'aurait pu donner la segmentation automatique avec un tel algorithme.

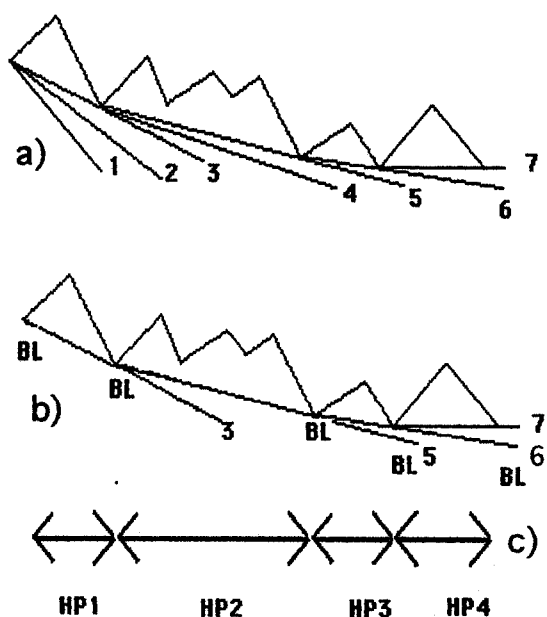


Figure 2: Première étape de la détermination automatique des passages par la ligne de base BL

La position exacte de la frontière entre deux HP tient compte de la limite des phonèmes. Si le passage par BL se fait entre deux voyelles, ou à la fin d'une voyelle ("felt", "badge", "impact", "visual", "rosette", pour L2, Fig. 4), on coupe à l'endroit exact du passage par BL; sinon, on coupe à la fin de la voyelle ("felt", "badges", "impact" pour L1, "candidate", "badge", "impact" pour L3), et ce "retard" est dû à la présence d'un Rc (règle identique en français).

2.C. Marquage de degrés de frontière entre HPs successifs

La phrase anglaise peut recevoir soit une structure que nous interprétons comme parallèle, où les coupures entre les HPs semblent être d'égal importance (figure

3a), soit une structure marquée (figures 3b à 3f), comme dans notre description traditionnelle pour le français. Le degré de frontière entre les HPs peut être varié par le locuteur à l'aide d'un certain nombre de procédés classiques, que les divers articles sur la prosodie ne cessent d'illustrer (il n'est pas possible d'en faire une liste ici). Ces procédés incluent l'insertion d'une pause (fig. 3c; Grosjean et al., 1979), (généralement accompagnée d'un allongement prépausal de la syllabe finale), la "cassure" de la continuité de BL par un nivellement (fig. 3b), ou un rehaussement (en anglais, "Resetting", Maeda, 1975, Cooper et Sorensen, 1981; en japonais, c'est une des composantes principales du modèle de Fujisaki, Fujisaki & al, Sudo (1971) souvent accompagné d'un rehaussement du plateau (fig. 3d), ou encore le pic sur un HP peut être abaissé par montrer sa dépendance avec le HP précédent (Fig. 3e) ou élevé pour au contraire, montrer son indépendance (Fig. 3f). Le fondamental sur la dernière syllabe avant la frontière peut recevoir une montée de continuation (un crochet ou "hook" en anglais selon la description de Delattre, 1962) ou encore (plus rarement) une valeur de Fo plus basse (voir le choix opposé des locuteurs 2 et 3 pour le mot "candidate", Fig. 4). Tous ces procédés sont relativement indépendants: on peut avoir un "resetting" ou une montée de continuation sans l'insertion de pause, par exemple et des études de perception sont nécessaires pour comparer leur efficacité perceptive à segmenter et hiérarchiser les groupes de sens dans la phrase.

2.D Subdivision d'un hat-pattern et différences entre locuteurs.

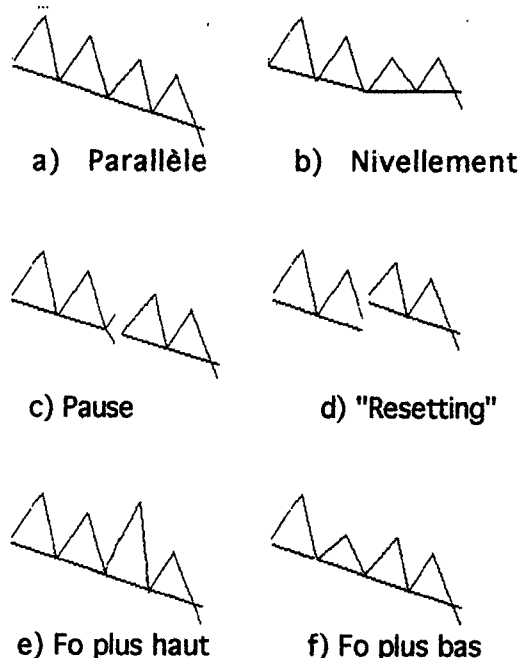


Figure 3: Quelques procédés classiques utilisés pour l'augmentation du degré de séparation entre les deuxième et troisième HPs.

Alors que la division en HPs a été sensiblement la même pour tous les locuteurs (voir Lea, 1974), les variations de Fo à l'intérieur d'un même HP ont montré de grandes différences individuelles. Par exemple, dans la suite (adjectif + nom), "trapezoidal badge", on observe une montée de Fo durant la consonne initiale sourde de "trapézoïdal" pour L1, L2 et L3. La descente est réalisée durant la syllabe accentuée du second mot "bAdge" pour L1, mais après la syllabe accentuée du premier mot pour L2 et L3.

Nous savons par expérience que ces différences ne sont pas aléatoires, elles sont la conséquence d'une ou plusieurs tendances sous-jacentes en conflit, que chaque locuteur tend à régler de façon individuelle. L'analyse visuelle des 1000 phrases montre de nombreuses régularités dans les différences entre locuteurs. On citera deux exemples:

a) la montée de Fo peut occuper toute la plage, et l'amplitude de la descente est alors inférieure à celle de la montée (Loc. L1) ou sensiblement égale (L2). Ou alors, la montée initiale n'occupe qu'une part réduite de la plage, et on observe un pic sur le plateau et une descente d'amplitude plus importante (L3).

(b) Le premier et le dernier HPs dans une phrase peuvent être plus ou moins influencés par leur position dans la phrase (pattern éventuellement "plus montant" en début de phrase pour L1).

3. DISCUSSIONS ET CONCLUSION

Bien que la ressemblance puisse être l'effet du hasard, les schémas typiques des trois locuteurs anglais rappellent ceux des locuteurs français (Vaissière, 1975), P3-P4, P3-P2, et P2 pour L1, et L2 et L3, respectivement.

Ces répétitions de patterns identiques d'une phrase à l'autre chez un même locuteur contribuent très certainement à l'impression du style particulier à chaque locuteur. Le degré différent des perturbations de ces schémas en début et en fin de phrase est peut-être dû à un poids différent à accorder au constituant HP dans sa combinaison avec les constituants de rang supérieur, GS et inférieur (Vaissière, 1992).

Le passage de Fo par la BL est donc une base essentielle dans notre description. Ce passage nous semble plus important que les détails de la courbe de Fo. Le retour à la BL semble avoir une signification particulière du point de vue physiologique, une sorte de remise à zéro, simulant la fin d'un effort physiologique. Les données sur la perception des courbes de Fo semblent aller dans cette direction: une chute de Fo ne donnera l'expression de finalité que si le Fo atteint la BL (Collier, 1984), mais ces résultats demandent à être confirmés sur des locuteurs de plusieurs langues.

Cette description n'est qu'une description de plus à ajouter à la liste déjà longue des descriptions offertes pour l'anglais. Elle a pour avantage d'être inspirée essentiellement par les données elles-mêmes. Une première inspection de la base de quelques données disponibles sur le Japonais à ATR (avec l'aide appréciée de nos collègues et d'Hélène Valbret) et à travers la

littérature (Fujisaki et Sudo, 1971; Kubozono, 1993) nous laisse à penser que les principes décrits ci-dessus pour le regroupement des HP sont également valables pour le japonais (cf. les exemples illustrés dans le livre de Kubozono). Nous pensons que des progrès en prosodie passe désormais par une accumulation de données sur de très nombreuses langues (voir, par exemple, Touati, 1987, pour une comparaison suédois-français).

Cette étude est très incomplète dans le sens où elle concerne uniquement les pauses et les variations de Fo, sans considération des variations d'intensité et surtout de durée. La structure temporelle à elle seule (par un allongement de voyelle finale) peut être responsable de la perception de ruptures (Duez, 1994 pour le français). Des études statistiques sur les structures temporelles des phrases sont une tradition à ATR (Sagisaka, 1992; Campbell, 1991). Il serait intéressant de comparer nos résultats avec les "courbes de durée", calculées par Nick Campbell sur les mêmes 1000 phrases. L'étude de la corrélation entre durée et intensité, durée et Fo s'est révélée négative en Japonais (Sagisaka, 1992), alors que l'amplitude et la durée semblent très corrélées en anglais (Fry, 1955). Il faudra peut-être, dans une étude des rapports entre durée, et Fo en anglais pour le marquage des frontières, tenir compte du fait que ce ne sont pas tant les valeurs de Fo sur les voyelles accentuées ou les tons qui importent, mais le passage de Fo par BL.

REMERCIEMENTS

Je tiens beaucoup à remercier la compagnie ATR qui m'a offert un havre de recherche durant les étés de ces deux dernières années, en particulier Messieurs Sagisaka et Campbell, dont j'ai beaucoup appris.

RÉFÉRENCES

- Abramson, A.S., (1978), "Static and dynamic acoustic cues in distinctive tone", *Language and Speech*, 21, 319-325.
- Campbell, N. (in press), "Automatic detection of prosodic boundaries in speech" accepté for publication in *Speech Communication*.
- Campbell, N., (1991), "Segment duration in a syllabic frame", *J. Phon.* 19, 37-47.
- Cohen, A., and t'Hart, (1967), "On the anatomy of intonation", *Lingua* 19, 177-192.
- Collier, R., (1984), "Some physiological and perceptual constraints on tonal systems", dans *EXPLANATIONS FOR LANGUAGE UNIVERSALS*, Butterworth & al (éditeurs), Mouton, New York, Berlin, Amsterdam.
- Cooper, W.F., & Sorensen, J.M., (1981), "FUNDAMENTAL FREQUENCY IN SENTENCE PRODUCTION", New-York: Springer.