

Influence des caractéristiques des contours mélodiques sur la durée des mots en anglais britannique contemporain

Caroline Bouzon

Laboratoire Parole et Langage UMR 6057
29 avenue R. Schuman – 13621 Aix-en-Provence Cedex 1, France
Tél. : ++33 (0)4 42 95 36 24 – Fax : ++33 (0)4 42 59 50 96
Mél : caro.bouzon@wanadoo.fr

ABSTRACT

Because pitch glides greatly influence the duration of words, we test here the influence of three characteristics of the pitch glide in final position of intonation units on the duration of words in British English. These characteristics are the direction, complexity and width of the glide. We observed the influence of each individual parameter before a minor and major boundary, and then the interaction between them. The effect of the parameters is different before a minor or a major boundary, hence the need to maintain this distinction in a model of duration. The effect also varies depending on the type of interaction, the influence of some factors being increased or decreased according to the other parameter it interacts with.

1. INTRODUCTION

La variation de la durée, quelle que soit l'unité utilisée, est un phénomène bien connu, tout particulièrement en anglais britannique. Une problématique actuelle qui en découle est de savoir comment le langage fonctionne en temps réel et comment se fait l'organisation temporelle de la parole. De nombreuses recherches portent sur la prédiction de l'organisation temporelle de la parole. Klatt [Kla87] a développé un modèle fondé sur onze règles de prédiction de la durée segmentale. Klatt applique des pourcentages d'allongement et de compression à des durées phonémiques inhérentes, en fonction de ces onze règles prédéfinies. Campbell [Cam92] élabore un modèle de prédiction de la durée au niveau supérieur de la syllabe. Le modèle permet tout d'abord de prédire la durée des syllabes, puis de déterminer la durée des segments de cette syllabe, en fonction de règles d'ajustement et d'élasticité. Le synthétiseur de parole ProSynth [Loc97][Ogd99] permet également une prédiction de la durée syllabique, basée sur le chevauchement des syllabes entre elles et sur la compression des syllabes en fonction de leur complexité et de leur position dans le pied. Ces exemples de modèle ont tous pour but d'atteindre une meilleure prédiction de l'organisation temporelle de l'anglais, afin d'améliorer la qualité et le naturel de la parole synthétique. Notons que l'objet d'étude peut varier d'approche en approche : la syllabe chez [Cam92] et [Loc97], le segment chez [Kla87] et [Möb00]. Cependant, la parole synthétique manque encore de naturel, Zellner-Keller [Zel96] remarque que tant que nous ne saurons pas mieux comment se fait l'organisation temporelle de la parole, la

parole synthétique manquera de naturel. L'organisation temporelle a en effet longtemps été reléguée au second plan par de nombreux chercheurs au profit de la mélodie.

2. TRAVAUX PRELIMINAIRES

Cette analyse découle de [Bou01] qui porte sur l'analyse prédictive de l'influence de certains facteurs prosodiques sur la durée des mots en anglais britannique, basée sur le corpus MARSEC [MAR]. Ce corpus possède une annotation prosodique et est aligné en mots, il contient au total plus de six heures de parole, le tout réparti en onze types de discours. Quelques synthèses informelles de textes anglais à partir de la durée mesurée des mots du corpus, ont montré que l'utilisation d'unités prosodiques plus grandes que la syllabe apporte un gain de naturel à la parole synthétique. Suite à ces résultats et compte tenu de notre corpus, nous nous proposons de travailler au niveau du mot. Nous observons ainsi l'allongement de la durée du mot de manière globale, sans tenir compte des unités inférieures telles que la syllabe ou le phonème. En suivant la classification en traits distinctifs de Hirst [Hir77], nous avons testé l'influence de plusieurs facteurs prosodiques sur la durée des mots [Bou01]. Le facteur le plus influent est la position finale des mots dans l'énoncé. Les autres facteurs importants sont l'influence du contour mélodique, l'accent lexical, le nombre de syllabes dans le mot, la compression des mots en position initiale et l'allongement des mots précédant un autre mot.

Nous explorons ici l'effet du contour mélodique (« pitch glide »), et par-là même, des paramètres de largeur, direction et complexité du contour mélodique. Nous entendons par *largeur* l'étendue du contour : une chute peut être large ou étroite selon que la variation de la voix est importante ou non. La *direction* distingue les contours ascendants (montée) et descendants (chute). La *complexité* définit un changement de direction du mouvement de la voix à l'intérieur d'un contour. Dans la terminologie de [OCO61], on a par exemple la chute-montée (« fall-rise ») : la première partie du contour est descendante, la seconde ascendante.

Les données sur lesquelles nous travaillons sont les mêmes que pour [Bou01]. Nous avons prédit la durée de chaque mot en fonction d'un tableau de durées moyennes des phonèmes anglais pris dans [Cam92]. Nous avons ensuite comparé les durées prédites de chaque mot à leur durée dans le corpus. Nous avons ainsi obtenu l'erreur absolue en ms de chaque mot, à laquelle nous ferons référence par le terme 'allongement'.

L'originalité de cette étude est que nous cherchons à prédire l'organisation temporelle au niveau du *mot*. Nous nous limitons dans cette analyse aux mots en position finale, à savoir chacun des derniers mots absolus d'unité mineure ou majeure. Nous nous appuyons sur la distinction entre les frontières mineures et majeures d'unité intonative définie par l'annotation du corpus MARSEC. Une analyse de variance révèle que l'effet du contour mélodique sur l'allongement des mots en position finale d'unité intonative est très significatif ($F(1,11508)=176.964$, $p<.0001$). Partant de ce résultat, nous avons testé l'influence des caractéristiques de largeur, complexité et direction des contours mélodiques définies plus haut et dans [Bou01].

3. INFLUENCE DES PARAMETRES

Pour chaque analyse, nous avons trié les données pour qu'il ne reste que les mots *avec* un contour mélodique. Ceci permet de mieux distinguer la différence entre les mots sans aucun contour mélodique et les mots avec un contour mélodique simple, étroit ou statique. Nous utiliserons les abréviations D-mineure (frontière mineure à droite) pour position finale d'unité intonative mineure, et D-majeure (frontière majeure à droite) pour position finale d'unité majeure.

1.1 Largeur

L'influence de la largeur du contour mélodique est très significative sur les mots en position finale, que la frontière soit mineure ou majeure ($F(1,7216)=10.959$, $p<0.001$). L'allongement provoqué par la largeur en D-mineure et D-majeure est faible, bien que toujours statistiquement significatif. Le rapport d'allongement est : *étroit*<*large*.

1.2 Complexité

L'effet de la complexité du contour mélodique est très significatif ($F(1,7216)=21.243$, $p<.0001$), tout comme la différence entre un contour simple et un contour complexe ($p<0.02$). L'effet en D-majeure est très marqué ($F(1,3261)=27.089$, $p<.0001$), avec un allongement moyen d'environ 54ms. En revanche, l'effet de la complexité n'est pas significatif sur les mots en D-mineure ($F(1,4754)=0.802$, $p>0.37$), l'allongement absolu moyen ne représentant qu'environ 3ms. Il ressort donc que la complexité du contour mélodique n'influence la durée des mots qu'en position finale absolue (unité majeure).

1.3 Direction

L'effet de la direction sur l'allongement des mots en fin d'unité n'est pas significatif ($F(2,7214)=0.055$, $p>0.94$), et seule la différence entre une montée et une chute est significative ($p<.0001$). L'ordre d'influence de ces différentes directions n'est pas régulier et change selon le contexte. En D-mineure, on obtient le rapport

d'allongement *statique*<*montée*<*chute*. La chute entraîne l'allongement le plus important, suivi de la montée, puis le contour statique. En revanche, en D-majeure, on obtient : *statique*<*chute*<*montée*, le rapport entre montée et chute est donc inversé. Il semblerait que la direction n'influence pas la durée des mots de manière identique s'il s'agit d'une frontière mineure ou majeure. Il faut cependant remarquer que les différences d'allongement ne sont pas très élevées. En D-mineure, seule la différence entre une montée et une chute est significative ($p<.0001$), on a donc bien *montée*<*chute*. En D-majeure en revanche, la différence entre une montée et une chute n'est pas significative, ce qui signifie que l'on ne peut pas tenir compte de cette inversion du schéma.

4. INTERACTION DE PARAMETRES

Jusqu'ici, nous avons observé l'effet des trois paramètres en fonction du type de frontière finale qui suit. Nous allons ici nous intéresser à l'interaction entre les paramètres de contour, d'abord en D-mineure, puis en D-majeure. Les analyses de variance ne nous permettent pas de grouper les trois paramètres dans la même analyse, nous présenterons donc les interactions par paires. Nous remarquons que, dans les interactions impliquant la direction, nous avons enlevé des données les mots avec un contour statique puisqu'ils ne peuvent pas être larges et/ou complexes.

2.1 Position finale d'unité mineure

Interaction entre largeur et direction L'interaction entre ces deux paramètres n'est pas significative ($F(1,4750)=2.893$, $p>0.08$). Les deux schémas respectifs que nous avons précédemment, *étroit*<*large* et *montée*<*chute*, ne sont pas modifiés (figure 1). L'effet de la largeur est significatif avec les montées, mais pas avec les chutes ($F(1,1890)=0.982$, $p>0.32$), probablement parce qu'une montée large finale revêt un aspect plus emphatique qu'une chute large finale.

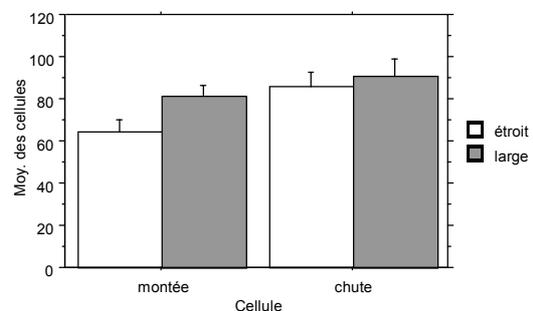


Figure 1: interaction entre *largeur* et *direction*

Interaction entre complexité et largeur Cette interaction n'est pas significative ($F(1,4750)=0.117$,

$p>0.73$). L'effet du paramètre *largeur* est significatif (nous conservons le rapport *étroit<large*), mais pas celui de *complexité* (figure 2), comme nous l'avons vu dans 1.1 et 1.2.

Il est intéressant de noter que la différence d'allongement entre les contours larges (simples) et les contours larges *et* complexes n'est pas significative ($F(1,2364)=1.638, p>0.20$). Ceci confirme le rôle secondaire de la complexité du contour. Il faut cependant remarquer que cette analyse porte sur tous les mots du corpus (pouvant contenir jusqu'à neuf syllabes). Les effets sont donc généralisés : un contour complexe peut par exemple porter sur une syllabe, comme sur neuf. Il serait intéressant de refaire ces analyses en triant au préalable les mots en fonction du nombre de syllabes qu'ils contiennent.

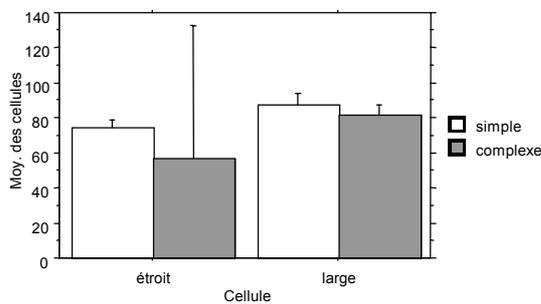


Figure 2: interaction entre *complexité* et *largeur*

Interaction entre complexité et direction Cette interaction n'est pas significative ($F(1,4750)=0.259, p>0.61$).

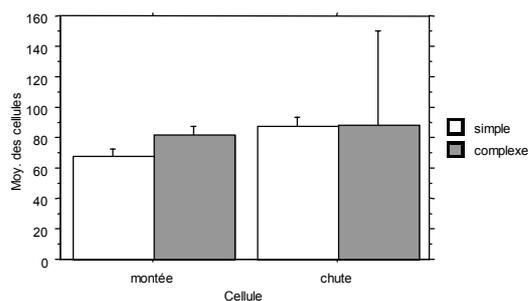


Figure 3: interaction *complexité* et *direction*

Nous conservons le schéma relatif au paramètre de direction *montée<chute* (figure 3), et en accord avec 1.2, la différence entre *simple* et *complexe* n'est pas significative. Quand les deux paramètres interagissent, on a bien *montée simple<montée complexe*, mais la différence entre *chute simple* et *chute complexe* n'est pas significative ($F(1,1890)=2.686^E-4, p>0,99$). De plus, l'écart type montre que nous ne pouvons pas tenir compte des chutes complexes.

Nous allons maintenant voir si ces interactions se retrouvent en position finale d'unité majeure.

2.2 Position finale d'unité majeure

Interaction largeur et direction Cette interaction est significative ($F(2,2458)=9.130, p<0.003$). Le rapport de largeur est toujours *étroit<large*, mais le rapport de direction n'est pas aussi régulier. Lorsque le contour est étroit, on a : *montée<chute*, mais quand il est large, on a : *chute<montée*. Ceci signifie qu'en position finale d'unité majeure, une montée large est plus allongée qu'une chute large, mais qu'une chute étroite est, elle, plus allongée qu'une montée étroite (figure 4). On peut émettre l'hypothèse qu'une montée large dans une telle position est emphatique, ou du moins, qu'elle met en relief le mot sur lequel elle porte au travers de l'allongement.

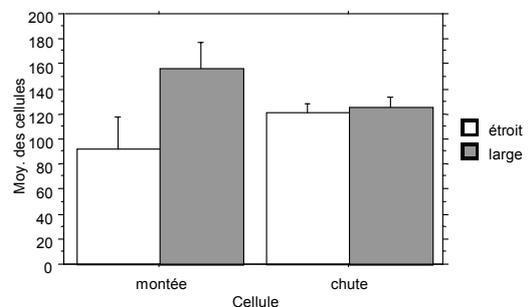


Figure 4: interaction entre *largeur* et *direction*

Il est intéressant de remarquer qu'il existe peu de différence d'allongement entre une chute simple et une chute large en position finale. Comme en D-mineure, la largeur du contour influence les montées, mais pas les chutes ($F(1,2282)=0.652, p>0.42$).

Interaction entre complexité et largeur

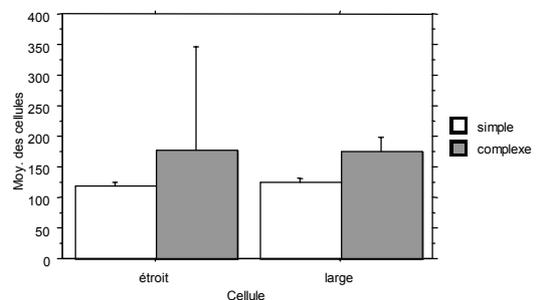


Figure 5: interaction *complexité* et *largeur*

L'interaction entre les deux paramètres est ici encore non significative ($F(1,2458)=0.016, p>0.90$). Les rapports ne sont pas modifiés : *étroit<large* et *simple<complexe*. L'effet de la complexité est très marqué lorsqu'ils

interagissent; en revanche, l'effet de la largeur n'est pas significatif (figure 5). Ces résultats sont en accord avec 1.2, mais pas avec 1.1. Une fois de plus, l'écart type pour les mots étroits complexes est tel que cet allongement ne peut pas être pris en compte.

Interaction entre complexité et direction Cette interaction n'est pas significative ($F(1,2458)=0.705$, $p>0.40$). Contrairement à D-mineure (mais en accord avec 1.2), l'effet de la complexité est très significatif. Lorsque les deux paramètres interagissent (figure 6), on a toujours *simple*<*complexe* quelle que soit la direction du contour, mais par contre la différence entre une montée complexe et une chute complexe est non significative ($F(1,97)=0.023$, $p>0.87$), les deux étant très marquées. Ceci s'explique par le fait que dans les deux cas, les deux mêmes sortes de contours (montée et chutes) sont impliquées.

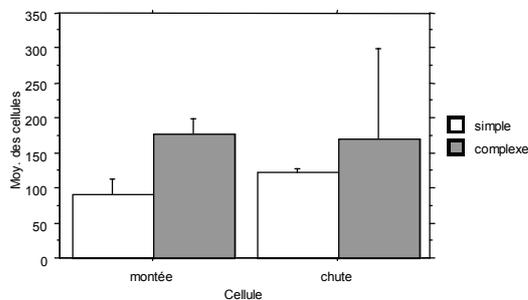


Figure 6: interaction complexité et direction

5. CONCLUSION

Nous avons observé les effets et les interactions des paramètres de direction, complexité et largeur sur l'allongement des mots anglais en position finale. Ces effets ne sont pas réguliers lorsqu'il y a interaction entre paramètres. D'une manière générale, les chutes sont plus allongées que les montées (certainement en partie pour des raisons physiologiques de relâchement), sauf si ces contours sont en contexte plus ou moins emphatique. Globalement, les effets des divers facteurs varient en contexte D-mineure ou D-majeure. Dans [Bou01], la complexité n'était pas ressortie comme importante dans les résultats, tout comme la différence entre D-mineure et D-majeure. Nous avons ici observé que la complexité jouait un rôle important en D-majeure, mais pas en D-mineure, son degré d'influence variant selon les interactions. Comme nous l'avons observé dans [Bou02] et comme nous le remarquons ici, le parallèle entre D-mineure et D-majeure révèle qu'il existe une différence fondamentale entre les deux frontières : les effets des divers facteurs sont amplifiés en D-majeure, et l'ordre de grandeur des allongements est presque doublé en D-majeure. Il est donc important de conserver la distinction entre frontière finale mineure et majeure dans un modèle

de prédiction de la durée, car le locuteur distingue bien deux niveaux d'unité dans son discours.

Nous avons utilisé des analyses de variance, mais elles nous restreignent à seulement deux paramètres par analyse. Or il serait intéressant de voir l'effet cumulé de ces trois paramètres dans la même analyse. Ceci serait réalisable avec, par exemple, une analyse de type CART ou avec un modèle additif. De plus, ces analyses sont purement empiriques dans le sens où nous nous sommes contentés de *tester* l'influence de ces facteurs sur l'allongement des mots. Il faudrait maintenant quantifier ces effets, puis vérifier les prédictions par synthèse et/ou tests perceptifs, dans le but d'élaborer un modèle de prédiction. Nous pensons que travailler sur une telle unité peut être bénéfique pour la prédiction, en gardant toutefois en tête les désavantages que cette approche représente par rapport au travail sur des unités telles que le pied, la syllabe ou le phonème.

BIBLIOGRAPHIE

- [Bou01] Bouzon C. (2001), "Influence des facteurs prosodiques sur la durée des mots en anglais britannique contemporain", Mémoire de DEA, Aix-en-Provence.
- [Bou02] Bouzon C. & Hirst D. (2002), "The Influence of Prosodic Factors on the Duration of Words in British English". In *Proceedings of the Speech Prosody 2002 conference*, Bel B. & Marlien I. (eds) Aix-en-Provence.
- [Cam92] Campbell N. (1992), Multi-Level Timing in Speech. PhD thesis, University of Sussex.
- [Hir77] Hirst D.J. (1977), Intonative features. The Hague: Mouton Publishers.
- [Kla87] Klatt D. (1987), "Review of Text-to-speech Conversion for English". *JASA* 82, pp.737-793.
- [Loc97] Local J. & Ogden R. (1997), "A Model of Timing for Nonsegmental Phonological structure", In *Progress in Speech Synthesis*, Van Santen J.P., Sproat R.W, Olive J.P. & Hirschberg J. (eds). Springer, pp.109-121.
- [MAR] MARSEC
http://midwich.reading.co.uk/research/speech_lab/marsec/marsec.html
- [Möb00] Möbius B. & Van Santen J. (2000), "Phonetically Motivated Modeling of Prosody". *Prosody 2000*, pp. 161-166.
- [OC61] O'Connor J.D. & Arnold G.F. (1961), "Intonation of Colloquial English", London : Longman.
- [Ogd99] Ogden R. et al. (1999), "Temporal Interpretation in ProSynth, a Prosodic Speech Synthesis System". *XIVth International Congress of Phonetic Sciences*, pp.1059-1062.
- [Zel96] Zellner-Keller B. (1996), "Structures temporelles et structures prosodiques en français lu". *Revue française de linguistique appliquée*, Vol. 1, pp. 7-23.