Vous avez dit *proéminence* ?

Michel MOREL[(1)], Anne LACHERET-DUJOUR[(1)(2)], Chantal LYCHE[(3)], François POIRÉ[(4)]

[(1)] CRISCO, Université de Caen, esplanade de la Paix, 14000 CAEN
[(2)] Institut Universitaire de France, PARIS
[(3)] Université d'Oslo, NORVÈGE
[(4)] The University of Western Ontario, CANADA
morel@crisco.unicaen.fr, lacheret@crisco.unicaen.fr, chantal@ilos.uio.no, fpoire@uwo.ca

ABSTRACT

Analysing prosody requires the correct identification of prominence peaks. This paper examines the results of an experiment where 7 linguists submitted to a prominence identification task largely failed to agree. We show that prominence detection is proportionate to F0 variation, but not to length, that there exists considerable variation between judges, the best of whom barely attains a 50 % score of correct answers. We conclude that coding prosody in a large corpus will require the use of dedicated software to supplement the work done by individual coders.

1. INTRODUCTION

Le travail ici décrit, centré sur l'analyse raisonnée d'une tâche de perception de proéminences par 7 juges phonéticiens dans un extrait de conversation libre, s'inscrit au sein du projet *Phonologie du Français contemporain* (PFC) amorcé en 2002 dont les objectifs initiaux sont présentés dans Durand & Lyche [2]. Dans ce cadre, s'est mis en place un volet prosodique dont une des composantes concerne l'étude de l'interaction des niveaux segmental et suprasegmental (en particulier schwa-prosodie). Cette dernière a progressivement donné lieu à un protocole de codage stabilisé autour de deux modes de codage : standard vs. étendu (Lacheret, Lyche & Morel [4][5]). Comme tout type de transcription prosodique, ce protocole repose en premier lieu sur l'identification de proéminences syllabiques sur des bases auditives. Mais qu'entend-on exactement par ce terme ? Peut-il émerger un consensus lorsqu'on demande à des analystes d'horizons divers d'indiquer les syllabes proéminentes dans un extrait sonore, *i.e.* de coder les événements prosodiques perceptivement remarquables ? La notion même de *syllabe* n'est-elle pas trop restrictive lorsqu'on sait que la proéminence syllabique peut se propager sur les syllabes environnantes (Astesano & al. [1]) ? Enfin, que peut-on dire des points d'ancrage psycho-acoustiques et linguistiques qui sous-tendent la démarche ? En d'autres termes, quelle tâche effectuent réellement les codeurs ? Malgré les meilleures intentions du monde, ces derniers peuvent-ils faire totalement abstraction du niveau symbolique qui conditionne en partie leurs attentes et sous-tend l'émergence des proéminences (structure syntaxique, contraintes sémantiques et informationnelles) ?

En pratique, cette communication fait suite à l'enquête préliminaire de F. Poiré [8], initiateur et fédérateur du travail (mise en service du corpus, définition de la tâche, dépouillement des données). Après avoir rappelé la tâche, sa mise en oeuvre et les premiers résultats dont nous rend compte l'auteur, notre objectif est de mettre en corrélation ces observations avec les configurations acoustiques des données (F0 et durée). Ceci nous permettra de : i) statuer sur la robustesse relative de nos deux paramètres acoustiques pour le repérage des proéminences, ii) proposer des mesures pour évaluer la performance des juges et iii) déterminer le profil des juges qui peut être établi grâce aux deux notions complémentaires de *seuil* et *performance*.

2. LA TACHE ET LES PREMIERES OBSERVATIONS ASSOCIEES

L'objet de cette section est de rappeler les modalités de la tâche envisagée et les premières observations perceptives auxquelles elle a donné lieu.

2.1. La tâche

Un extrait d'environ 3 minutes de parole spontanée produite par un locuteur belge a été choisi pour réaliser la tâche d'identification des proéminences, en position finale et non finale (syllabe proéminente codée '1', syllabe non proéminente codée '0'). Cette tâche a été demandée à 7 sujets, tous phonéticiens et/ou phonologues chevronnés, spécialistes de prosodie. 165 syllabes au total ont été analysées par F. Poiré, soit parce que elles ont été codées '1' par au moins un des 7 juges (ce qui inclut les syllabes emphatiques), soit parce qu'elles sont en position accentuable (syllabes terminales de mots pleins [3]).

2.2. Premiers résultats

Les premières observations formulées par Poiré [8] à l'issue de son analyse s'articulent autour de 3 points majeurs. (i) La perception des proéminences non terminales ou portées par des clitiques reste marginale et pour cause : les sites en question ne correspondent pas aux positions attendues des syllabes accentuables. (ii) Pour l'ensemble des codeurs, la variation dans le pourcentage des syllabes

reconnues comme proéminentes (code '1') est telle (19 % à 49 %) qu'il semble raisonnable de conclure que les sujets ne partagent pas la même définition du concept. (iii) En revanche, l'accord inter-juges est plus sensible pour les syllabes non proéminentes.

3. ANALYSE ACOUSTIQUE DES RESULTATS ET DISCUSSION

3.1. Outils d'analyse

Un relevé de mesure (F0, durée) a été effectué sous PRAAT sur les 165 syllabes analysées ainsi que les syllabes environnantes, utilisées comme référence. La figure 1 présente le traitement d'un extrait de la séquence ci-dessous :

(...) 750 kilos là il te faut un permis B (...)

transcrite en sampa, où l'analyse porte sur la syllabe */lo/* .

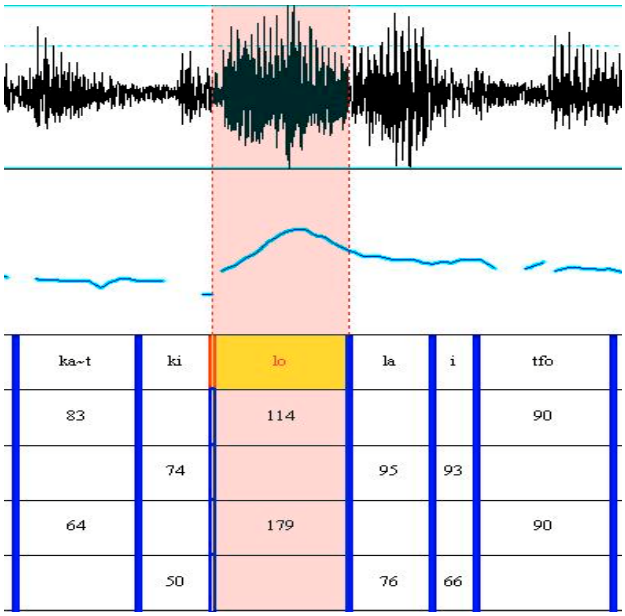

Figure 1 : mesure des proéminences sous PRAAT, illustration pour la syllabe */lo/*. Abscisse : temps. Fenêtre du haut : visualisation du signal sonore. En dessous : courbe F0 correspondante. Fenêtre du bas : 5 tires de traitement remplies manuellement :

- tire 1 : codage phonétique
- tire 2 : 114 Hz = valeur maximale de F0
- tire 3 : 74 et 95 Hz = valeurs environnantes prises comme références
- tire 4 : 179 ms = valeur maximale de durée
- tire 5 : 50 et 76 ms = valeurs environnantes prises comme références.

Sur ces bases, pour chacune de nos 165 syllabes, deux valeurs, *F0* et *durée* (en italique), ont été calculées relativement aux valeurs de référence. *F0* est exprimée en demi-tons (12 unités = rapport 2) selon la formule :

$$F0 = \frac{Log(\frac{F0max}{F0base})}{Log(2)} \times 12$$

F0max étant la valeur maximale de F0 dans la syllabe et *F0base* la moyenne des F0 des syllabes précédente et suivante. Le procédé est le même pour la durée, à ceci près que *durée* est exprimée en pourcentage de la valeur de base. La valeur 100 correspond ainsi à une durée syllabique égale à la moyenne de celle qui précède et de celle qui suit, donc à un allongement nul. Dans notre exemple (syllabe /lo/ de *kilos*), les valeurs calculées sont les suivantes :

F0 = 5,2 demi-tons *durée* = 284 %

3.2. Exploitation des données

Nous avons créé la variable *accord*, qui est la somme des valeurs 0 ou 1 attribuées par les juges. Ainsi, <*accord* = 0> correspond à une syllabe non perçue comme proéminente par l'ensemble des juges alors qu'elle est considérée comme accentuable et <*accord* = 7> à une syllabe perçue proéminente par les 7 juges. Nous avons ensuite mis en correspondance pour chaque syllabe la valeur de *accord* avec les valeurs calculées *F0* et *durée*, afin de comparer perception et mesure (figures 2 et 3).

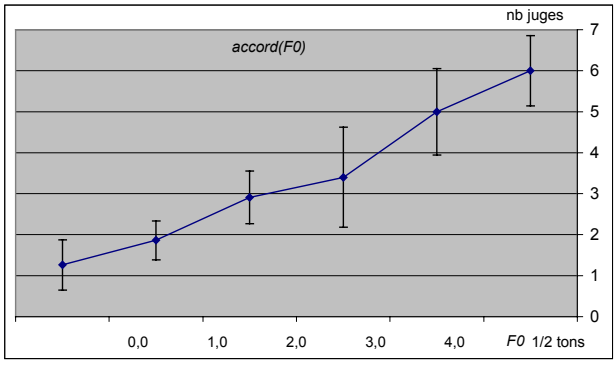

Figure 2 : *F0* en abscisse, *accord* en ordonnée. Chaque point constitue la moyenne des valeurs de *accord* dans la plage considérée de *F0*. Les barres d'erreur représentent les intervalles de confiance à 5 %.

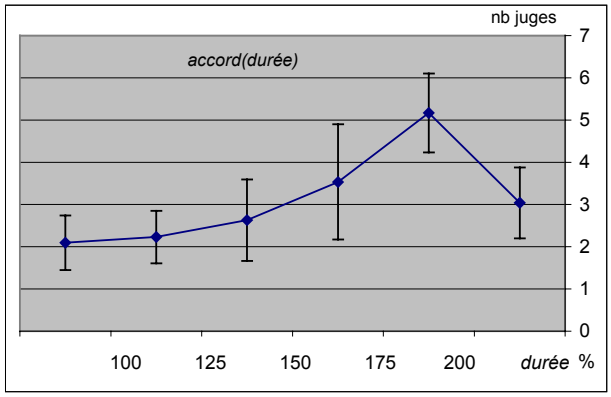

Figure 3 : *durée* en abscisse, *accord* en ordonnée. Chaque point constitue la moyenne des valeurs de *accord* dans la plage considérée de *durée*. Les barres d'erreur représentent les intervalles de confiance à 5 %.

Première remarque : si l'accord des juges est significativement proportionnel aux variations de F0, il n'en va pas de même pour la durée. La figure 2 montre que l'accord des juges est bien corrélé à *F0* dans les valeurs fortes, un peu moins dans les valeurs faibles où il reste de l'ordre de 1 à 2, mais avec toujours le même sens de progression. On peut donc conclure, toute chose égale par ailleurs, que *F0* constitue un corrélat acoustique fiable de la perception des proéminences. La figure 3 en revanche met en évidence une faible corrélation entre *accord* et *durée*, avec une dispersion importante (grands intervalles de confiance). Seules les valeurs de *durée* de 170 à 200 % semblent convaincre plus de la moitié des juges. Mais surtout, les extrémités de ce graphique montrent deux phénomènes qui limitent encore la fiabilité de la durée comme corrélat acoustique de la perception des proéminences. D'une part, des syllabes de faible durée peuvent être perçues comme proéminentes (variation de F0 marquée, allongement nul). D'autre part, les augmentations fortes de durée (supérieures à 200 %) sont souvent associées à des hésitations et de fait perçues comme telles et étiquetées '0'.

Ce dernier point renvoie au concept même de proéminence : si une proéminence est une figure qui se détache sur un fond, alors une durée longue devrait être considérée comme telle. Or la plupart des juges ne considèrent pas l'hésitation comme une proéminence. On peut émettre l'hypothèse qu'ils considèrent – consciemment ou non – sa fonction de gestion du tour de parole comme non pertinente dans le cadre d'une étude sur la prosodie. Quoi qu'il en soit, la question du rôle de la durée dans la notion de proéminence pourrait faire l'objet d'une étude spécifique. Ici, les juges ont appliqué leur propre conception de la proéminence, avec comme résultat une faible corrélation entre durées et proéminence perçues.

La première partie de cette étude confirme la légitimité de sélectionner la F0 comme corrélat acoustique robuste de la proéminence (invariant à la perception de l'ensemble des sujets). Reste à statuer sur la performance individuelle de chaque juge. Autrement dit, qu'en est-il de la variation inter-juges dans la tâche de détection des proéminences ?

Pour chaque juge, nous avons constitué deux sous-ensembles de syllabes : codées '0' vs. '1'. Nous avons rangé les *F0* correspondantes par ordre croissant dans chaque sous-ensemble. La figure 4 présente les résultats obtenus par le juge 3 (courbes croissantes noires) choisi au hasard pour illustrer notre propos. Chez ce dernier, 108 syllabes sont codées '0' et 57 '1'. La référence (courbe croissante grise) correspond à toutes les valeurs de *F0* rangées par ordre croissant. Elle représente donc le résultat virtuel d'un juge *parfait*, qui aurait classé non proéminentes les 108 syllabes où *F0* est la plus basse et proéminentes les 57 syllabes où *F0* est la plus haute. Nous avons appelé *seuil* la valeur de *F0* correspondant à la frontière entre ces deux groupes de syllabes (1,4 demi-tons dans l'exemple choisi figure 4). Bien évidemment, chaque juge possède son propre seuil de perception des proéminences. Celui-ci varie même dans d'assez fortes proportions d'un juge à l'autre. Mais attention : ce seuil ne préjuge en rien de la performance de chacun. (*infra*, table 1, figure 5). En pratique, l'évaluation des performances d'un juge consiste à vérifier s'il a bien perçu comme proéminentes les syllabes dont *F0* est supérieure à son seuil personnel – 1,4 demi-tons pour le juge 3, par exemple – et comme non proéminentes celles dont *F0* est inférieure à ce seuil, autrement dit à tester la cohérence du juge dans la tâche réalisée. Cela revient donc à comparer les résultats de chaque juge à ceux d'un juge *parfait* travaillant avec un seuil de perception identique. Ainsi, pour notre juge 3, la courbe (noire) présente une forte discontinuité à la frontière des syllabes codées '0' et '1'. Ce qui signifie que des valeurs de *F0* hautes n'ont pas été associées à des proéminences, alors que des valeurs basses l'ont été. Quant aux 6 autres juges, ils présentent des caractéristiques similaires, sans pour autant bien sûr que les différences avec le juge *parfait* concernent les mêmes syllabes.

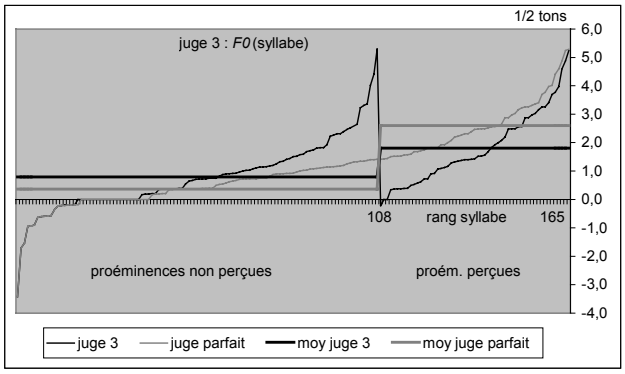

Figure 4 : *F0* en ordonnée, rangée par ordre croissant respectivement sur les syllabes perçues non proéminentes et les syllabes perçues proéminentes.

A ce stade le taux de bonnes réponses ne suffisait pas pour conclure ; il fallait en effet hiérarchiser les erreurs en accordant plus d'importance aux dysfonctionnements manifestes (ex. fortes valeurs de *F0* non perçues comme indices de proéminences et inversement). Pour ce faire, nous avons calculé, pour nos syllabes étiquetées '0' et '1', les moyennes respectives des F0 de chaque juge et du juge *parfait* (plateaux noirs pour notre juge 3 et plateaux gris pour le juge parfait). La différence entre les deux moyennes est plus forte pour le juge *parfait* que pour le juge 3 (respectivement 2,2 et 1,0 demi-tons). En effet, chaque fois que l'on classe une syllabe dans la mauvaise catégorie, les deux moyennes se rapprochent, et ceci d'autant plus que la *F0* de cette syllabe est éloignée du seuil de perception, donc que l'erreur est importante. Nous avons appelé *séparation* la différence entre ces deux moyennes et l'avons utilisée comme indice de performance individuelle des juges. La valeur maximale (MAX) de *séparation* correspond au juge *parfait* ; sa valeur minimale (MIN) correspond à un juge qui classerait aléatoirement les syllabes comme proéminentes ou non ; auquel cas la moyenne des *F0* serait la même dans les deux catégories – non perçues et perçues –, *séparation* serait alors nulle. Concrètement, pour chacun de nos 7 juges, la valeur de *séparation* est comprise entre MIN et MAX.

Table 1 : seuil de perception des proéminences, indice de séparation et performance de chaque juge.

	seuil (1/2 tons)	séparation (demi-tons)	performance (%)
Juge *parfait*	0,9 à 2,3	2,0 à 2,6	100
Juge 1	1,7	1,36	56
Juge 2	0,9	1,05	51
Juge 3	1,4	1,02	46
Juge 4	2,3	1,16	44
Juge 5	1,2	0,91	43
Juge 6	0,9	0,87	42
Juge 7	1,4	0,91	41

La table 1 montre que (i) les seuils de perception des proéminences sont très variables d'un juge à l'autre, (ii) seuls deux juges dépassent une performance de 50 %, ce qui confirme le caractère empirique et en partie aléatoire du jugement individuel perceptif.

Enfin, une représentation graphique des résultats de la table 1, illustrée par la figure 5, confirme l'indépendance entre seuil de perception et taux de performance.

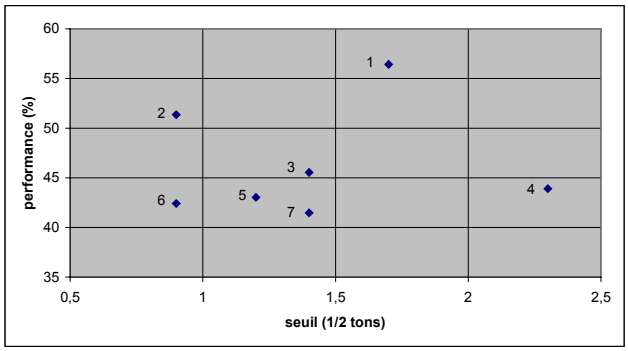

Figure 5 : seuil de détection des proéminences et performance des juges numérotés de 1 à 7.

4. CONCLUSION

Les résultats obtenus à partir de la tâche considérée illustrent plusieurs points. (i) Y compris chez les experts, la notion de proéminence semble loin d'être consensuelle. En conséquence, pour les tâches de codage à venir dans PFC, il paraît nécessaire d'en préciser la portée plus finement, bref, de mieux circonscrire la consigne afin d'aboutir à des étiquetages plus homogènes (par exemple s'entendre sur le traitement des hésitations, également fixer une hiérarchie dans la sélection des proéminences a priori perçues, *i.e.* déterminer un seuil en deçà duquel la syllabe ne peut pas être étiquetée proéminente – voir aussi Martin [6]). (ii) Même dans ces contextes, il est dangereux de se contenter d'un sous-bassement purement auditif pour reconstruire la structure prosodique ; des logiciels d'aide à la détection s'avèrent fondamentalement nécessaires. Au delà de l'utilisation systématique et maintenant classique de l'affichage de F0, voire de la durée, une représentation de la mélodie mesurée, facile à mettre en œuvre, à lire et à interpréter s'avère fort utile (voir par exemple le logiciel Prosogramme développé par Mertens [7]). Une telle approche permettra notamment de contrôler la cohérence des codages obtenus en sachant qu'ils peuvent venir d'horizons fort divers (codeur ± néophyte en phonétique, système dialectal du codeur et du locuteur ± différents), et ainsi de se donner des moyens pour se défaire de la vision pessimiste selon laquelle la détection des proéminences s'apparente plus à un art qu'à une pratique rigoureuse (Martin [6]).

BIBLIOGRAPHIE

[1] C. Astesano, M. Morel, A.L. Coquillon, R. Espesser, M. Besson et A. Lacheret-Dujour. Marquage acoustique du focus contrastif non codé syntaxiquement en français. *25èmes Journées d'Étude sur la Parole*, Fès, Maroc, 19-22 avril 2004.

[2] J. Durand et C. Lyche. Le projet 'Phonologie du Français Contemporain' (PFC) et sa méthodologie. In *Corpus et variation en phonologie du français*. E. Delais-Roussarie et J. Durand (éd.) Toulouse. PUM, pages 213-276, 2003.

[3] A. Lacheret et F. Beaugendre. *La prosodie du français*. Editions du CNRS, Paris, 1999.

[4] A. Lacheret, C. Lyche et M. Morel. Codage prosodique : lien prosodie - schwa/liaison. *Bulletin PFC* n° 3, pages 89-98, 2003.

[5] A. Lacheret, C. Lyche et M. Morel. Pour une transcription prosodique normalisée au sein du projet PFC (Phonologie du français contemporain) : champ d'action et perspectives. *25èmes Journées d'Etude sur la Parole*, Fès, Maroc, 19-22 avril 2004.

[6] P. Martin. La transcription des proéminences accentuelles : mission impossible ? *Bulletin PFC* n° 6, ERSS, UMR 5610, CNRS et Université de Toulouse-Le Mirail, pages 81-87, 2006.

[7] P. Mertens. Un outil pour la transcription de la prosodie dans les corpus oraux. *Traitement Automatique des langues* n° 45 (2), pages 109-130, 2004.

[8] F. Poiré. La perception des proéminences et le codage prosodique. *Bulletin PFC* n° 6, ERSS, UMR 5610, CNRS et Université de Toulouse-Le Mirail, pages 69-79, 2006.