LA VALEUR ABSOLUE DU GRADIENT DE FO : DEFINITION, LOCALISATION ET DISTRIBUTION EN LECTURE DE TEXTE FRANCAIS SOUS TROIS CONSIGNES DIFFERENTES

G. Caelen-Haumont

Laboratoire de la Communication Parlée
UA n° 368 INPG / ENSERG, Grenoble , France.

ABSTRACT

This paper aims at putting forward a new pitch parameter, which is the absolute value of the Fo gradient in the lexical word. Three sets of reading instructions which become more and more exacting with regard to discourse intelligibility, do not upset this parameter prevalence, but nevertheless exert a significant modulation on the distribution of the three parameters analysed in this study.

1. INTRODUCTION

Généralement les études qui portent sur l'analyse du pitch (ou Fo) s'appuient sur les valeurs moyennes calculées sur l'ensemble de la voyelle, sur la partie stable centrale [5], ou aux deux-tiers [4]. Parfois encore 3 références sont prises, aux frontières et au centre de la voyelle [3].

Dans l'étude que nous avons menée sur les relations de coïncidence numériques entre 6 modèles prédictifs (2 syntaxiques, 3 sémantiques, 1 pragmatique) et les paramètres prosodiques [2], nous proposons outre les paramètres de l'énergie et de la durée, et les paramètres mélodiques "classiques" du maximum de Fo (ou FoM) et Fo moyen (ou Fom), un nouveau paramètre mélodique qui s'est révélé très efficace, à savoir la valeur absolue du gradient de Fo (ou |ΔFo|). Ce nouveau paramètre est dans cette étude relatif aux unités lexicales.

L'expérimentation porte sur un texte[1] de 30 mots lexicaux composé de 3 phrases et de 11 "groupes minimaux". Nous rappelons [1] que ces groupes minimaux sont définis syntaxiquement et prosodiquement comme les groupes syntaxiques de plus bas niveau, immédiatement supérieurs à la structure superficielle, éventuellement associés avec le groupe syntaxique suivant de manière à former une structure de 5 syllabes, nécessaire et suffisante pour l'autonomie prosodique du groupe. Trois consignes de lecture ont été présentées à 12 locuteurs (1° lecture naturelle et intelligible 2° lecture très intelligible 3° lecture très très intelligible pour un ordinateur). Avant l'enregistrement, les mots les plus spécialisés ont été explicités, si besoin était, car l'expérimentation ne portait pas sur la compréhension du texte, mais sur la communication de cette compréhension, autrement dit, sur le "faire-comprendre".

Les 36 enregistrements organisés en base de données, ont été segmentés et étiquetés par un expert-phonéticien.

[1] Le texte est le suivant : "D'éminents biologistes et d'éminents zoologistes américains ont créé pour des vers géants un nouveau phylum dans l'actuelle classification des nombreuses espèces vivantes. Ces longs vers prospèrent sur le plancher marin des zones sous-marines profondes. Des sources thermales chaudes y maintiennent une température moyenne élevée."

2. CHOIX DES PARAMETRES MELODIQUES

L'étude dans son ensemble analyse 14 paramètres mélodiques qui se subdivisent en 3 types (valeur absolue du gradient de Fo, maximum de Fo et Fo moyen), en trois localisations (ensemble du mot, syllabe finale, "contour"), et deux contextes de référence (le texte et la phrase). Ces contextes de référence définissent en fait deux espaces de réduction des valeurs numériques, réduction qui concerne aussi bien les modèles que les paramètres mélodiques. Compte-tenu de la petite part de connaissance que véhicule chaque modèle pris isolément, et de toutes les sources de variabilité tant linguistiques qu'extra-linguistiques, un espace à quatre niveaux nous a semblé correspondre à un juste compromis.

L'ensemble de ces combinaisons aboutit à 18 paramètres, mais le nombre a été réduit à 14 par suppression de Fo maximum et Fo moyen dans le contour et dans les deux contextes de référence.

En outre de manière à rendre égales les conditions de sélection de ces divers paramètres en vue d'une comparaison inter-locuteurs, les frontières des items ont été localisés à l'écran en prenant soin de ne pas relever aux bornes de ceux-ci, — qui sont souvent le lieu des valeurs numériques extrêmes—, les unités phonétiques réputées non voisées, de même que les /ə/ qui leur sont postérieurs, le voisement du premier ou l'existence du second relevant de la variabilité locuteurs.

Dans cette communication nécessairement réduite par rapport à l'autre [2], nous n'envisagerons que les 3 types de paramètres (|ΔFol, FoM, Fom) en neutralisant contextes de référence et localisations.

3. CRITERES DE REALISATION DES PARAMETRES. CONTEXTE DES PHRASES

Le paramètre ΔFo est certainement le paramètre le plus délicat à réaliser pour le locuteur dans la mesure où il exige de positionner au sein des pentes mélodiques croissantes et décroissantes le temps de quelques ms deux cibles qui sont à la fois par rapport au mot lexical des extrema absolus (et relativement inverses), et par rapport à la chaîne mélodique de la phrase et du texte, des extrema relatifs. Lorsque pour une raison ou une autre, l'effort est trop grand, les locuteurs positionnent une des deux cibles, en l'occurrence le maximum de Fo, en une position clé du mot lexical. Lorsque ces conditions sont encore trop difficiles, il suffit alors d'ajuster au besoin par approximations successives grâce au feed-back, les valeurs mélodiques moyennes du registre voulu pendant l'énonciation des unités phonétiques voisées du mot, soit un temps considérablement plus long. Ces 3 paramètres |ΔFol, FoM et Fom semblent en fait se comporter comme les avatars progressivement détériorés d'un même processus.

En ce qui concerne les phrases du texte, on remarque que la phrase 1 a la propriété à la fois d'être la plus longue (de l'ordre de deux fois) et de détenir les mots les plus spécialisés. La phrase 2 possède le lexique d'accès le plus facile. La phrase 3 est la plus courte mais présente une information inattendue.

4. METHODOLOGIE D'ANALYSE

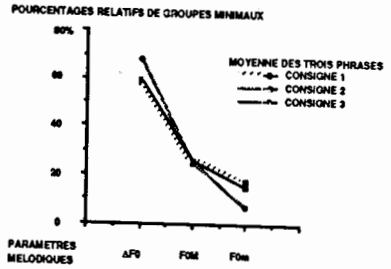
Les combinaisons modèle/paramètre offrant le plus grand nombre de coïncdences ayant été retenues, le principe d'analyse consiste à suivre d'aussi près que possible l'évolution de la distribution des modèles et des paramètres qui leur sont liés. Le groupe minimal est l'élément de base, nécessaire et suffisant, pour être la cible d'un changement de stratégie, mais dans la majeure partie des cas, il se combine à d'autres pour former des macro-structures signifiantes qui fournissent précisément le support à l'expression de la stratégie mise en oeuvre par le locuteur.

Dans ces conditions, la méthode de travail consiste à sélectionner pour le premier groupe minimal, la meilleure combinaison modèle prédictif / paramètre mélodique, —meilleure au sens numérique—, et ensuite à trouver pour le ou les groupes suivants, le meilleur compromis entre ces meilleurs taux de coïncidence et les principes de cohésion et de cohérence du système qui poussent à conserver le cadre conceptuel et mélodique, c'est-à-dire le meilleur compromis entre la dynamique et l'économie du système.

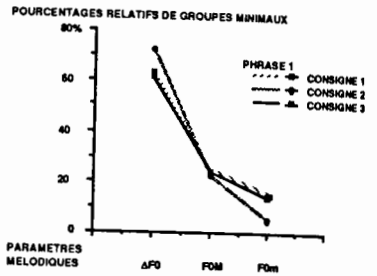
4. RESULTATS

L'étude présente se limite à l'analyse générale de la distribution des paramètres mélodiques tous locuteurs confondus, en fonction des consignes de lecture et des phrases. Nous comparerons les paramètres sous l'angle de leurs pouvoirs explicatifs.

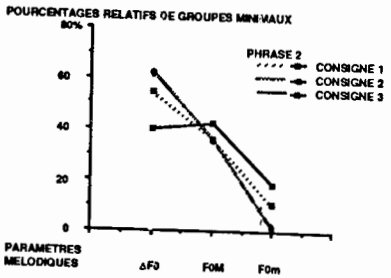
Le graphique 1 ci-dessous présente les pourcentages moyens relatifs des trois paramètres, toutes phrases confondues, en fonction des 3 consignes. Alors que les débits moyens ralentissent très sensiblement, nous constatons que les effectifs des consignes 1 et 3 restent voisins. Les débits de parole (plus pauses) varient en effet tous locuteurs confondus, respectivement de la consigne 1 à la 3, de 2.23 à 1.82 puis à 1.05 mots / seconde. Il ressort des pourcentages que le paramètre |ΔFol correspond en consignes 1 et 3, à 58 ou 59% des observations totales des 3 paramètres, alors que Fom ne compte que 15 à 17% de celles-ci. FoM quant à lui, reste très stable puisqu'il recueille 25 et 26% effectifs, score invariant en consigne 2



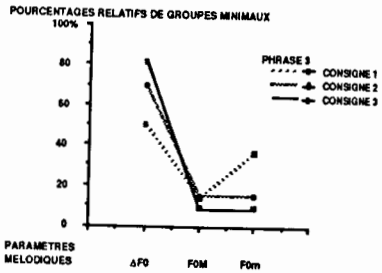
Graphique n° 1



Graphique n° 2



Graphique n° 3



Graphique n° 4

Graphique n° 1 : Pourcentages moyens relatifs des groupes minimaux tous locuteurs confondus, toutes phrases confondues, en fonction des trois types de paramètres mélodiques, Valeur absolue du gradient de Fo (ou |ΔFo|), Maximum de Fo (ou F0M) et Fo moyen (ou F0m).

Graphique n° 2 à 4 : Pourcentages moyens relatifs des groupes minimaux tous locuteurs confondus en phrase 1 (graphique 2), en phrase 2 (graphique 3), en phrase 3 (graphique 4) , en fonction des 3 consignes de l'énoncé et des trois types de paramètres mélodiques, Valeur absolue du gradient de Fo (ou |ΔFo|), Maximum de Fo (ou F0M) et Fo moyen (ou F0m).

(25%). Par rapport aux effectifs des autres consignes, la consigne 2 opère donc une augmentation de ceux de |ΔFo| (68%) au dépens exclusif de Fom (7%). L'évolution de la distribution des effectifs des trois paramètres montre donc dans l'ensemble et indépendamment de l'évolution de la distribution des modèles linguistiques, que l'effet des consignes de lecture, s'il joue considérablement sur le débit de parole, ne modifie pas du tout au tout le choix des paramètres. La consigne 2 toutefois représente une réponse (version paramètres mélodiques) à la première exigence d'augmentation de l'intelligibilité en favorisant au prix de la difficulté de réalisation, le paramètre le plus délicat à mettre en oeuvre, |ΔFo|. La consigne 3 apporte une autre solution qui ramenant les effectifs des paramètres à ceux de la consigne 1, privilégie, indépendamment du domaine conceptuel des modèles linguistiques étudié par ailleurs [2], le paramètre durée de réalisation des unités phonétiques et des pauses.

Les graphiques 2 à 4 montrent l'effet des consignes sur chacune des phrases. Le graphique 2, très proche du graphique 1, maximalise les effectifs de |ΔFo| avec respectivement 63 à 61% des effectifs totaux pour les consignes 1 et 3, et 72% pour la consigne 2.

Le graphique 3 montre que la distribution

des effectifs en consigne 1 est globalement intermédiaire des effectifs des consignes 2 et 3. De manière générale les effectifs de |ΔFo| sont, quelle que soit la consigne, toujours inférieurs (la fourchette est comprise entre 40 et 62%) à ceux de la phrase 1. Il est intéressant de constater que cette dégradation se fait au profit du paramètre qui ne requiert pas le moins d'attention puisque Fo maximum connait une nette progression au dépens également de Fo moyen. La consigne 3, la plus contraignante de toutes, révèle une perturbation assez remarquable puisque, exceptionnellement, FoM recueille les pourcentages les plus grands (42%) au détriment de |ΔFo| (40%), évolution négative qui alimente aussi Fom (18%).

Le graphique 4 qui illustre la phrase 3 est caractéristique. Dans les conditions de lecture naturelle et intelligible (consigne 1), les effectifs de |ΔFo| sont toujours prédominants, mais depuis la phrase 1, ils ne cessent de décroître, ce qui peut facilement s'expliquer par les caractéristiques des phrases elles-mêmes ou encore l'effet fatigue (respectivement 63% -> 54% -> 50%). Lorsque les consignes de lecture exigent plus d'intelligibilité, le contrôle de l'utilisation des paramètres se révèle plus ferme. Il s'opère alors un retournement de tendance et les effectifs de |ΔFo| augmentent sensiblement (50% -> 70%) au détriment exclusif de Fom (36% -> 15%). La consigne 3 prolonge l'effet en augmentant encore les effectifs de |ΔFo| (70% -> 81%) au dépens de Fom mais aussi de FoM et dans les mêmes proportions (15% -> 9%).

Si l'on analyse ces résultats en prenant en compte d'une part la succession des phrases et leurs caractéristiques propres, il apparaît que 1° le contenu de l'information de la phrase 1 est communiqué avec une grande attention en utilisant dans la proportion globale de 2 fois sur 3, le paramètre de |ΔFo| 2° dans les conditions de lecture naturelle et intelligible, l'effet d'un contexte moins difficile (et sans doute aussi de la fatigue) se fait progressivement sentir de la phrase 1 à la phrase 3, détériorant progressivement les performances de |ΔFo| 3° une consigne de lecture "moyennement" contraignante (lecture très intelligible) a pour effet de potentialiser en moyenne les ressources des locuteurs et d'augmenter très sensiblement la proportion de |ΔFo| dans les phrases 2 et 3 4° une consigne encore plus stricte (lecture très très intelligible pour un ordinateur) a inversement l'effet de radicaliser les comportements de la consigne 1, en accusant fortement la détérioration attestée en phrase 2, mais inversement le relâchement substantiel de la tension en cette phrase 2 a pour effet de recréer les conditions favorables à une attention plus soutenue en phrase 3, ce qui est effectivement réalisé comme le montre la sélection massive de |ΔFo| (81%).

5. CONCLUSION

Cette communication a révélé l'efficacité d'un nouveau paramètre mélodique, qui se définit dans le cadre du mot lexical, et qui est la valeur absolue du gradient de Fo. Ce paramètre est sélectionné par les locuteurs dans les proportions globales de 2 fois sur 3, alors que le maximum de Fo représente aussi les deux-tiers des effectifs restants. Des consignes de lecture plus contraignantes ne remettent pas généralement en cause sa suprématie mais modulent de manière significative la distribution des effectifs de ces 3 paramètres. Ce paramètre possède la propriété d'exprimer de la manière la plus adéquate, la relation entre d'une part l'organisation cognitive des informations (approximées par les modèles) et d'autre part l'organisation mélodique de la chaîne parlée, mais mobilisant de ce fait fortement les facultés d'attention des locuteurs, il est nécessairement relayé par d'autres paramètres moins exigeants, mais aussi moins expressifs.

REFERENCES

[1] CAELEN-HAUMONT, G. (1989), "Une représentation syntaxique adaptée à la prosodie", *J. d'Acoustique*, 2, 137-146.

[2] CAELEN-HAUMONT, G. (1991), "*Stratégies des locuteurs et consignes de lecture d'un texte: analyse des intéractions entre modèles syntaxiques, sémantiques, pragmatique et paramètres prosodiques*", Thèse d'Etat, Aix-en-Provence.

[3] EMERARD, F., BENOIT, C. (1987) "De la production à l'extraction, l'état d'un chantier", 16èmes JEP, SFA-CNRS, Hamammet, Tunisie 224-226.

[4] ROSSI, M. (1971), "Le seuil de perception des glissandos", *Phonetica, 23*, 129-161.

[5] VAISSIERE, J. (1989) "On Automatic Extraction of Prosodic Information for Automatic Speech Recognition System", EUROSPEECH, Vol. 1, Paris, 202-205.