HAL
open science

Une évaluation de l'algorithme de stylisation mélodique MOMEL

Estelle Campione, Jean Veronis

To cite this version:

Estelle Campione, Jean Veronis. Une évaluation de l'algorithme de stylisation mélodique MOMEL. Travaux interdisciplinaires du Laboratoire Parole et Langage, 2000, 19, pp.27-44. hal-00285557

HAL Id: hal-00285557

https://hal.science/hal-00285557v1

Submitted on 5 Jun 2008

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.

Une évaluation de l'algorithme de stylisation mélodique momel

Estelle Campione, Jean Véronis[1]

Résumé

Cet article présente une évaluation quantitative et qualitative de l'algorithme de stylisation mélodique momel (Hirst & Espesser, 1993), qui permet d'extraire le contour macroprosodique des énoncés de façon indépendante de la langue, et en donne une représentation sous forme de points-cibles. L'évaluation a été conduite sur un grand corpus français et italien (12 locuteurs au total, 55 minutes de parole, soit 10260 points-cibles). Nos résultats confirment l'excellente performance de l'algorithme qui, sur les deux langues confondues, ne produit que 3,90% de bruit (points erronés) et 6,09% de silence (points manquants), les résultats étant notablement meilleurs sur l'italien que sur le français. Un fort pourcentage des erreurs sont systématiques et pourraient être évitées par une modification simple de l'algorithme, qui conduirait à une notable réduction du bruit et une quasi-disparition du silence.

Abstract

AN EVALUATION OF THE MOMEL ALGORITHM FOR PITCH CONTOURS STYLIZED. This paper describes a quantitative and qualitative evaluation of the momel melodic stylisation algorithm (Hirst & Espesser, 1993). This algorithm extracts the macroprosodic contour of utterances in a language-independent way, and produces a target-point representation. The evaluation was performed on a large corpus in French and Italian (12 speakers altogether, 55 minutes of speech, i.e. 10260 target-points). Our results confirm the excellent performance of the algorithm, which produces only 3.90% of noise (erroneous points) and 6.09% of silence (missing points). Results are better on Italian than French. A large percentage of errors are systematic and could be avoided by a simple modification of the algorithm. This would result in a noticeable reduction of noise and a near suppression of silence.

Mots-clés : prosodie, stylisation, intonation, évaluation
Keywords : prosody, stylisation, intonation, evaluation.

Introduction

La fréquence fondamentale (F0) peut être considérée comme la superposition de deux composantes indépendantes (Di Cristo & Hirst, 1986) :

• la composante macroprosodique qui reflète le choix d'un patron intonatif par le locuteur

• la composante microprosodique, qui ne résulte pas d'un choix du locuteur, mais de caractéristiques des segments phonématiques individuels de l'énoncé (effet intrinsèque des voyelles et des consonnes).

[1] Coordonnées des deux auteurs : Jeune Équipe DECLIC, Université de Provence, 29, av. R. Schuman, 13621 Aix-en-Provence. e-mail : estelle.campione@up.univ-aix.fr

La stylisation de la F0 consiste à factoriser ces deux composantes pour ne conserver que le contour macroprosodique de l'énoncé. De nombreux instituts ont, depuis les années soixante, travaillé sur la stylisation automatique de la F0 (Cohen & 't Hart, 1965 ; 't Hart & Collier, 1975 ; Rossi, 1971, 1978, Di Cristo, 1985; 't Hart *et al.,* 1990 ; Taylor, 1993, 1994 ; d'Alessandro & Mertens, 1995). Toutefois, alors que la plupart des méthodes proposées utilisent une stylisation basée sur des segments de droites le plus souvent discontinus (car limités aux parties voisées), le Laboratoire Parole et Langage a développé une méthode originale, MOMEL (pour MOdélisation MELodique), qui représente la totalité de l'énoncé par une courbe lisse et continue (Hirst & Espesser, 1993), composée d'une fonction spline quadratique, c'est-à-dire d'une série d'arcs de parabole se reliant par des tangentes communes.

L'hypothèse sous-jacente à cette méthode est que le contour macroprosodique stylisé doit être pratiquement identique aux courbes brutes de la F0 d'énoncés constitués uniquement de sonorantes. L'observation de ces courbes montre qu'elles sont continues et lisses. Les fonctions splines quadratiques sont les fonctions continues les plus simples ne présentant pas d'angles vifs (leur dérivée première est continue). A l'inverse, une modélisation par segments de droite, bien que pouvant être continue si les parties non voisées sont interpolées, présente une suite de points de rupture consécutifs. De plus, une courbe spline quadratique peut être représentée par une suite de points correspondant aux seuls changements significatifs (passages par zéro de la tangente). Ces points significatifs ou points-cibles peuvent donc constituer une représentation phonétique de la F0.

L'algorithme MOMEL a été utilisé sur diverses langues au Laboratoire Parole et Langage et ailleurs (Hirst & Espesser, 1993 ; Véronis *et al.,* 1994; Aasa & Stangert, 1996 ; Hirst *et al.,* 2000), et ses performances sont satisfaisantes, bien qu'un certain nombre d'erreurs soient généralement constatées dans la stylisation des énoncés, qui doivent être corrigées manuellement. Nous proposons dans cet article une évaluation quantitative et qualitative de l'algorithme sur un grand corpus français et italien (12 locuteurs au total, 55 minutes de parole). Nos résultats confirment l'excellente performance de l'algorithme, et indiquent qu'un fort pourcentage des erreurs sont systématiques et pourraient être évitées par une modification simple de l'algorithme, ce qui confirme les observations préliminaires d'Astesano *et al.* (1997).

1. L'algorithme MOMEL

L'algorithme MOMEL procède en quatre étapes. Nous reprenons la description de Hirst & Espesser (1993). Après élimination de quelques valeurs aberrantes (étape 1), la partie centrale de l'algorithme (étape 2) utilise une technique de « régression modale asymétrique ». Cette étape est basée sur l'hypothèse que le seul effet de la composante microprosodique de la F0

est d'abaisser localement les valeurs de la courbe macroprosodique sous-jacente. La technique de régression modale est appliquée dans une fenêtre d'analyse glissante (moving window) qui calcule une cible optimale pour les valeurs de la F0 de cette fenêtre. L'étape suivante de l'algorithme (étape 3) partitionne les cibles candidates. La dernière étape (étape 4) réduit toutes les cibles de chaque partition en une seule cible.

• Étape 1 : toutes les valeurs supérieures à un seuil donné par rapport aux valeurs adjacentes (5%) sont neutralisées. Cette étape permet d'éliminer quelques valeurs aberrantes ou douteuses aux transitions entre voisées et non voisées.
• Étape 2 : cette étape se décompose en trois sous-étapes appliquées à chaque instant *x* :
a) Dans une fenêtre d'analyse d'une longueur A (300 ms) centrée sur *x,* les valeurs F0 ne faisant pas partie d'un intervalle défini par deux seuils, *Hzmin* et *Hzmax,* sont neutralisées. Les seuils sont choisis de la façon suivante :
• *Hzmin* = 50 Hz ;
• *Hzmax* = moyenne des 5% des valeurs les plus hautes, multipliée par 1.3.
b) Une régression quadratique est calculée sur toutes les valeurs non-neutralisées, et toutes les valeurs à plus d'une distance *D* (5%) au-dessous de la F0 estimée par la régression sont neutralisées. Cette étape est réitérée jusqu'à ce que plus aucune valeur ne soit neutralisée.
c) Un point-cible local <*t, h*> est calculé à partir de l'équation de régression. Ce point correspond à l'extremum de la parabole correspondante (Figure 1). Si *t* est inférieur à *Hzmin* ou supérieur à *Hzmax,* alors le point-cible est ignoré et traité comme valeur manquante.

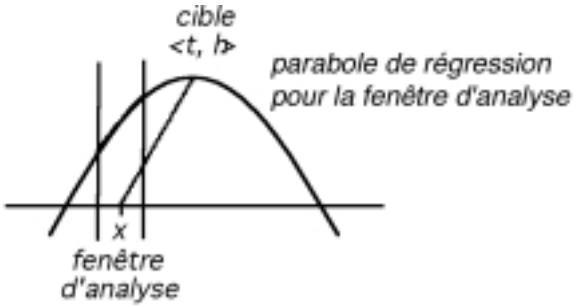

Figure 1.
Calcul d'un point-cible local

• Étape 3 : la partition des points candidats s'effectue dans une fenêtre de réduction glissante *R* (200 ms). Cette fenêtre est divisée en deux moitiés, gauche et droite, et une frontière de la partition est insérée si la moyenne des positions points-cibles candidats dans les deux moitiés diffèrent au-delà d'un certain seuil (Figure 2).

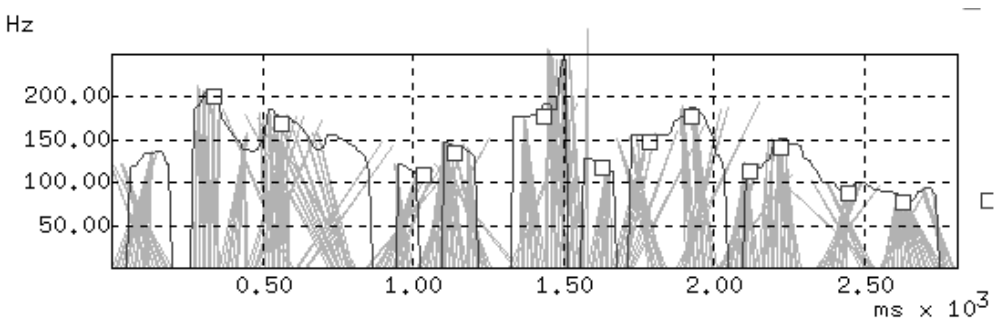

Figure 2.
Estimation des points

• Étape 4 : dans chaque segment de la partition les points candidats plus éloignés qu'une déviation standard par rapport à la moyenne du segment sont éliminés. La valeur moyenne des points restants est calculée comme l'estimation finale de *<t, h>* pour le segment (Figure 3).

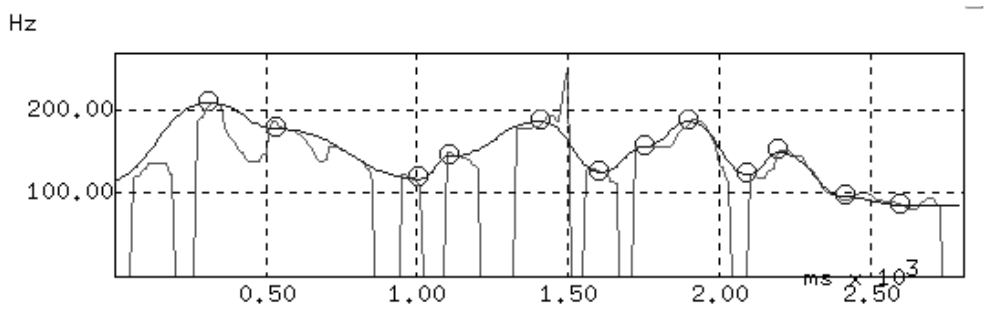

Figure 3.
Courbe de Fo et courbe spline quadratique obtenue

Les trois paramètres utilisés par l'algorithme (la fenêtre d'analyse [*A*], le seuil de distance [*D*] et la fenêtre de réduction [*R*]), ont été estimés à partir d'un petit corpus composé de deux phrases contenant toutes les occlusives et fricatives du français et prononcé par 10 sujets.

2. Description du corpus

Le corpus utilisé dans cette étude est une partie du corpus de parole multilingue EUROM1 qui a été réalisé et collecté dans le cadre du projet SAM (Multi-lingual Speech Input/Output Assessment, Methodology and Standardisation, cf. Chan *et al.,* 1995).

Nous avons centré notre étude sur les parties française et italienne du corpus et nous avons travaillé uniquement sur les passages lus par dix locuteurs (5 hommes et 5 femmes) dans chacune des langues. Ces passages regroupent cinq phrases liées sémantiquement entre elles.

2.1. Le corpus français

Le corpus français comprend quarante passages de cinq phrases, soit au total 200 phrases différentes. Ces passages sont regroupés en quatre groupes de 10, qui sont lus par dix locuteurs de façon que :

• les dix locuteurs prononcent chacun un groupe (dix passages);

• les deux premiers groupes soient lus par trois locuteurs ;

• les deux derniers groupes soient lus par deux locuteurs.

Au total, 100 passages (500 phrases) sont lus, totalisant 36,51 minutes de parole. Chaque phrase est lue en moyenne par 2,5 locuteurs et la durée par locuteur varie de 3 minutes 28 secondes à 4 minutes 37 secondes.

2.2. Le corpus italien

Le corpus italien comprend également quarante passages de cinq phrases, soit au total 200 phrases différentes. Ces passages sont la traduction des passages français, ou plutôt, les deux sont la traduction de la version anglaise. La traduction est assez libre, et constitue souvent une adaptation.

Ces passages sont organisés en groupes et lus de façon différente du français. En effet, les 40 passages sont regroupés en huit groupes de cinq, et non quatre groupes de dix, et sont lus par dix locuteurs de façon que :

• les dix locuteurs prononcent chacun trois groupes (quinze passages);

• les six premiers groupes soient lus par quatre locuteurs.

• les deux derniers groupes soient lus par trois locuteurs.

Au total, 150 passages (750 phrases) sont lus, totalisant 54,31 minutes de parole. Chaque phrase est lue en moyenne par 3,75 locuteurs et la durée par locuteur varie de 5 minutes 02 secondes à 7 minutes 11 secondes.

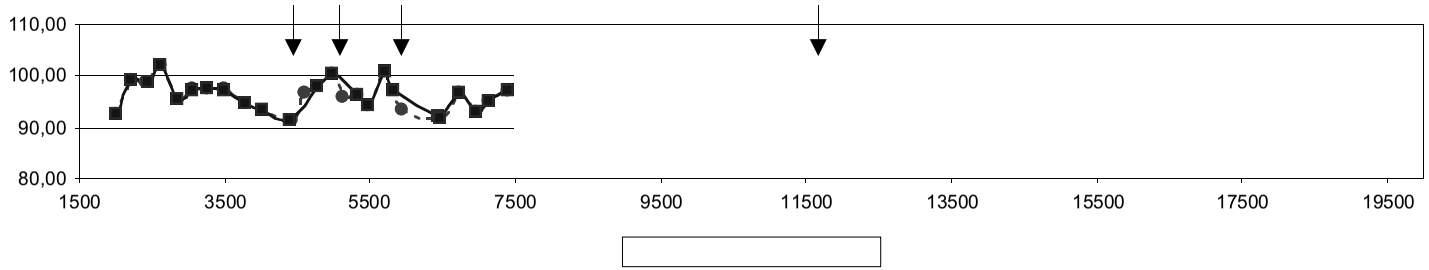

Figure 4.

Exemple de correction d'un passage (abscisses en ms, ordonnées en demi-tons)

3. Méthodologie de correction

La totalité du corpus a été stylisée automatiquement avec l'algorithme MOMEL. Cet algorithme présente quelques imperfections qui induisent des erreurs de points-cibles dans les énoncés. Nous avons corrigé les points-cibles erronés de tous les passages du corpus. Pour cela, nous avons appliqué une stratégie de correction de nature minimaliste, consistant à ne corriger que les points-cibles qui donnaient lieu à une différence audible entre la courbe originale de F0 et la resynthèse obtenue par la stylisation, jusqu'à ce que la re-synthèse soit jugée identique à l'original. La figure 4 montre un exemple de correction manuelle sur un passage du corpus (locutrice française *ja*, passage *00*).

La méthodologie consiste en une validation perceptive de la stylisation par une comparaison auditive entre la F0 originale et la courbe modélisée, à l'aide de la technique de re-synthèse PSOLA (Hamon *et al.,* 1989). Deux méthodes différentes (qualitative et quantitative) ont été employées pour analyser les résultats.

Nous avons corrigé manuellement six locuteurs dans chaque langue (22 minutes de parole pour le français et 33,2 pour l'italien, soit 10260 points-cibles vérifiés au total). Les autres locuteurs ont été corrigés avec la même stratégie par d'autres membres de l'équipe.

Toutes les erreurs de codage des points-cibles ont été notées pendant la correction, et classées en plusieurs catégories :

• *les points-cibles erronés* : MOMEL produit un certain nombre de points incorrects. Ces points peuvent être répartis en deux sous-catégories :

. *les points-cibles mal placés* : les points-cibles sont détectés par MOMEL, mais sont mal positionnés sur la courbe stylisée.

. *les points-cibles redondants* : dans sa version actuelle, MOMEL produit des points-cibles parfois très proches l'un de l'autre et donc redondants. La suppression de l'un des points ne modifie pas la re-synthèse.

• *les points-cibles manquants* : dans un certain nombre de cas, la version actuelle de MOMEL produit une courbe stylisée qui ne suit pas toujours la courbe originale de F0 de façon très fidèle. L'ajout de points-cibles permet d'améliorer l'adéquation de la courbe stylisée et de la courbe originale.

Après la phase de correction manuelle du corpus, nous avons conduit une évaluation de la partie que nous avons personnellement corrigée (six locuteurs dans chaque langue). Deux types d'évaluation ont été réalisés :

• *une évaluation quantitative,* visant à mesurer l'efficacité de l'algorithme MOMEL, et la proportion des différents types d'erreur ;
• *une évaluation qualitative,* permettant d'analyser de façon plus précise les erreurs les plus fréquemment rencontrées et leur contexte d'apparition.

4. Résultats

4.1. Analyse quantitative

4.1.1. Mesures d'évaluation

Les erreurs se répartissent comme indiqué dans le tableau 1. La colonne MOMEL donne d'une part le nombre total de points proposés par l'algorithme (sous-colonne *Nb Points*), et d'autre part le nombre de points corrects parmi ceux-ci (sous-colonne *Corrects*). La colonne *EXPERT* donne le nombre total de points dans la stylisation finale corrigée par l'expert. La colonne *ERREURS* donne la répartition des erreurs selon les différents types (Mal placés, Redondants, Manquants) ainsi que leur nombre total. Enfin, la colonne *EVALUATION* donne plusieurs mesures d'évaluation bien connues en recherche documentaire (cf. van Rijsbergen, 1979) :
• le *rappel,* représentant la proportion des points déterminés correctement par MOMEL par rapport à l'ensemble des points de la courbe stylisée finale, approuvés ou corrigés par l'expert ;
• la *précision,* représentant la proportion des points déterminés correctement par MOMEL par rapport au nombre total de points qu'il fournit.
Ces deux mesures sont complémentaires de deux autres mesures, qui sont parfois plus parlantes :
• le *silence,* représentant la proportion des points « oubliés » par MOMEL par rapport à l'ensemble des points de la courbe stylisée finale ;
• le *bruit,* représentant la proportion des points déterminés incorrectement par MOMEL par rapport au nombre total de points qu'il fournit.

Ces mesures se calculent de la façon suivante :

$$rappel = \frac{Pts_Corrects_MOMEL}{Total_Pts_Expert} \qquad précision = \frac{Pts_Corrects_MOMEL}{Total_Pts_MOMEL}$$

silence = 1- rappel bruit = 1- précision

Nous avons également utilisé une mesure qui permet de représenter de façon globale l'efficacité d'un système, en tenant compte à la fois du rappel et de la précision (F-*measure* de van Rijsbergen, 1979) :

$$efficacité = 2\frac{rappel.précision}{rappel + précision}$$

4.1.2. Étude globale

Le tableau 1 montre qu'au total, MOMEL produit 3,90% de bruit, et 6,09% de silence. Son efficacité globale est de 94,99%. Les erreurs se répartissent à peu près pour moitié en bruit (52,3%) et en silence (47,7%). Le bruit est composé de 68,5% de points mal placés et de 31,5% de points redondants.

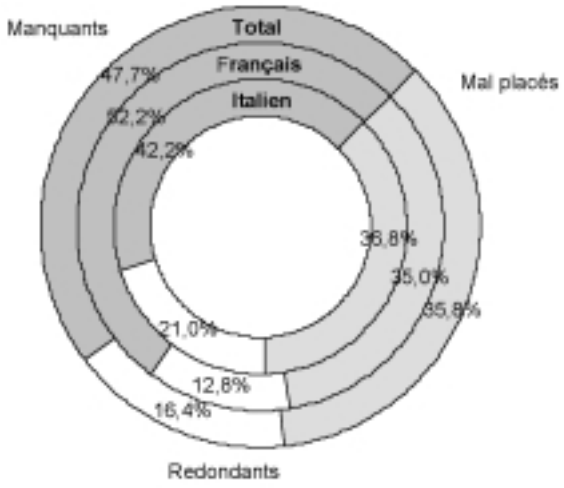

Figure 5.
Répartition des types d'erreur

LANGUE	LOC	*Nb points*	*Corrects*	*Nb points*	*Mal placés*	*Redondants*	*Manquan*
FRANÇAIS	fa	688	648	717	29	11	40
	ja	737	718	756	9	10	29
	mh	653	613	693	33	7	47
	ro	648	619	674	23	6	32
	sc	705	670	728	23	12	35
	sh	665	630	691	28	7	33
Sous-total		4096	3898	4259	145	53	216
ITALIEN	a0	1020	987	1029	18	15	24
	ag	801	781	825	14	6	30
	ba	1139	1094	1158	29	16	35
	b4	1006	971	1020	23	12	26
	b6	996	976	995	10	10	9
	b7	968	928	974	29	11	17
Sous-total		5930	5737	6001	123	70	141
Total		10026	9635	10260	268	123	357

Tableau 1.
Evaluation de la stylisation automatique

4.1.3. Étude par langue

Les résultats sont assez différents en français et en italien : on observe des résultats moins bons sur le français, en particulier au niveau du silence (8,48% au lieu de 4,40%, voir figure 5). Un test de χ^2 (tableau 2) montre que l'hypothèse nulle (absence d'effet selon la langue) peut être rejetée ($p < 0,0001$).

	p
Bruit	<0,0001
Silence	<0,0001

Tableau 2.
Test de χ^2 (effet de la langue)

Deux causes (non mutuellement exclusives) peuvent être envisagées pour expliquer cette différence : elles peuvent être dues soit à des différences entre langues, soit à une différence de stratégies de correction, qui pourrait être explicable par la moins grande maîtrise d'une des deux langues par la correctrice (qui est l'un des deux auteurs : Estelle, de français langue maternelle et italien langue seconde). Il est difficile de départager complètement ces deux hypothèses en l'absence de tests systématiques. Toutefois, l'examen de l'écart moyen entre les parties voisées de la courbe stylisée et de la courbe originale de F0 peut fournir un certain nombre d'indices (dans le cas idéal, cet écart devrait n'être constitué que par les effets microprosodiques).

D'une part, le tableau 3 montre que l'écart moyen entre la courbe stylisée fournie automatiquement par MOMEL et la courbe originale (colonne *Auto*) est plus important en français qu'en italien, ce qui confirme une différence objective entre langues.

D'autre part, le tableau concerne la totalité des passages français et italien du corpus EUROM1, incluant donc aussi bien la partie corrigée par l'un des deux auteurs (Estelle, six locuteurs dans chaque langue), que la partie corrigée par d'autres membres du laboratoire (Corinne et Fabienne, quatre locuteurs dans chaque langue) [seule la partie corrigée par Estelle a été incluse dans l'évaluation précédemment décrite]. On voit (colonne *Eval*) que la correction manuelle réduit l'écart entre courbe stylisée et courbe originale, et l'on peut constater qu'en français, la réduction est du même ordre pour les deux correctrices (voir aussi tableau 4 et figure 6). Par contre, en italien, Estelle a eu tendance à moins corriger que l'autre correctrice (Fabienne, pour qui l'italien est également seconde langue). Or, Estelle maîtrise mieux la langue, puisqu'elle la pratique depuis plusieurs années et qu'elle a séjourné régulièrement en Italie, alors que Fabienne la pratique depuis peu. On peut donc imaginer que Fabienne ait eu tendance à rechercher une conformité basée partiellement sur une écoute " musicale " des phrases (que l'on sait plus exigeante qu'une écoute linguistique).

Il semble donc que la différence de bruit et de silence observée entre le français et l'italien ne soit pas due à une différence entre stratégie de correction dans les deux langues, mais bien à une différence de performance de MOMEL. Une étude plus approfondie serait nécessaire pour étudier en détail les caractéristiques linguistiques qui pourraient être sources d'erreurs plus importantes pour l'algorithme (une différence dans le nombre des pauses pourrait partiellement expliquer le silence plus grand en français : voir § 4.2.3).

LANGUE	LOCUTEUR	CORRECTEUR	ÉCART MOYEN	
			Auto	Eval
français	*bf (fr)*	Corinne	0,053	0,047
	bo	Corinne	0,072	0,06
	fa	Estelle	0,072	0,063
	ja	Estelle	0,058	0,053
	mh	Estelle	0,065	0,053
	ro	Estelle	0,069	0,06
	sc	Estelle	0,049	0,041
	sh	Estelle	0,054	0,047
	sl	Corinne	0,059	0,053
	vi	Corinne	0,052	0,047
Ensemble du français			0,06	0,052
italien	*ao*	Estelle	0,043	0,04
	au	Fabienne	0,05	0,044
	b4	Estelle	0,056	0,05
	b6	Estelle	0,039	0,039
	b7	Estelle	0,041	0,039
	ba	Estelle	0,054	0,05
	bf (it)	Fabienne	0,056	0,047
	bk	Fabienne	0,042	0,033
	bl	Fabienne	0,044	0,038
	ag	Estelle	0,054	0,048
Ensemble de l'italien			0,048	0,043
Ensemble des deux langues			0,053	0,047

Tableau 3.
Écart moyen entre courbes stylisées et Fo originale

LANGUE	DONNÉES	CORRECTEUR		
		Corinne	Estelle	Fabienne
Français	Moyenne Auto	0,059	0,061	
	Moyenne Eval	0,052	0,053	
Italien	Moyenne Auto		0,048	0,048
	Moyenne Eval		0,044	0,041

Tableau 4.
Écart moyen entre courbes stylisées et Fo originale par correcteur

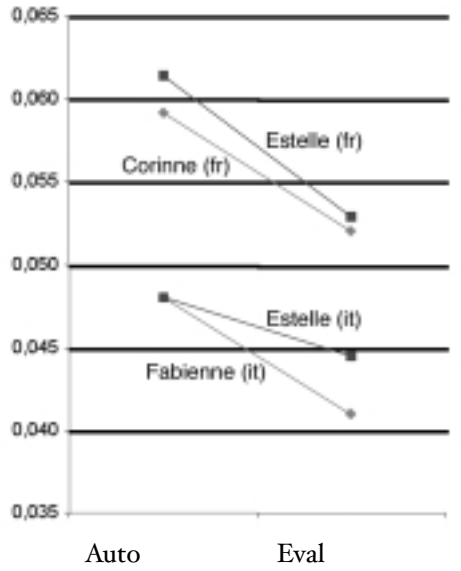

Figure 6.
Écart moyen entre courbes stylisées et Fo originale

4.1.4. Étude par locuteur

Les résultats varient quelque peu selon les locuteurs (figure 7). Certains locuteurs sont très bien stylisés (*ja* en français, *b6* en italien), d'autres le sont de façon plus médiocre (*mh* en français).

Le tableau 5 montre les résultats des tests de χ^2 : on ne peut rejeter l'hypothèse nulle (absence d'effet locuteur) que dans un cas : le silence pour les locuteurs italiens. Dans un autre cas, le bruit en français, elle pourrait être rejetée au seuil 0,05, mais pas au seuil 0,01. On ne peut donc pas globalement conclure à un effet significatif des locuteurs.

	p	
	Français	Italien
Bruit	0,0293	0,0586
Silence	0,1946	<0,0001

Tableau 5.
Test de χ^2 (effet des locuteurs)

Un test de χ^2 montre également que l'on doit rejeter l'hypothèse selon laquelle le sexe des locuteurs a une influence sur la qualité de la stylisation.

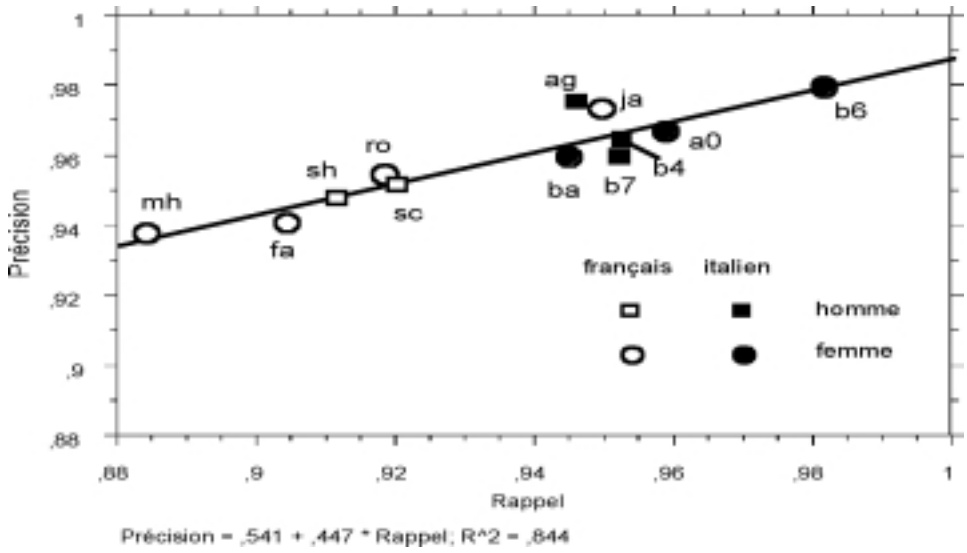

Figure 7.
Évaluation par locuteur

4.2. Analyse qualitative

Nous avons noté pour chaque catégorie d'erreurs (points mal placés, redondants et manquants) les erreurs les plus fréquemment rencontrées et leur contexte d'apparition. Il apparaît que la plupart des erreurs sont de nature systématique dans les deux dernières catégories, et qu'elles pourraient être évitées par une correction appropriée de l'algorithme MOMEL. On peut prédire une réduction du bruit à moins de 3% et une quasi disparition du silence.

4.2.1. Points mal placés

Dans cette catégorie, les erreurs interviennent quel que soit le contexte (en début, milieu ou fin de phrase ou de mot), de façon assez peu systématique. Du point de vue perceptif, ces

erreurs représentent généralement des corrections minimes (figure 8 – les points-cibles originaux sont représentés par des carrés, les points validés par des ronds).

4.2.2. Points redondants

Environ les trois quarts des erreurs de ce type sont dus à des points-cibles collés. Comme le montre la figure 9, l'élimination d'un des points-cibles ne modifie pas la courbe stylisée et n'entraîne pas la modification des points-cibles adjacents. Ce type d'erreurs pourrait être facilement corrigé en modifiant l'algorithme qui semble contenir une erreur dans la réduction des cibles candidates, ou par une phase de post-traitement.

4.2.3. Points manquants

Deux types d'erreurs systématiques se produisent dans cette catégorie :

• Avant une pause : ces erreurs représentent la grande majorité des points manquants. Ces erreurs induisent une mauvaise modélisation des contours mélodiques finaux (ascendants et parfois descendants), dont la différence avec la courbe originale de F0 est très importante (figure 10) et changent généralement l'intention intonative (intonation montante au lieu d'une intonation descendante ou inversement).

• Après une pause : ces erreurs sont moins nombreuses mais ont également une influence très forte sur l'intonation perçue. En particulier, l'absence assez systématique de points-cibles en début de phrase provoque une intonation trop haute par rapport au registre du locuteur (figure 10).

Ces deux types d'erreurs pourraient être facilement corrigés par une prise en compte des pauses.

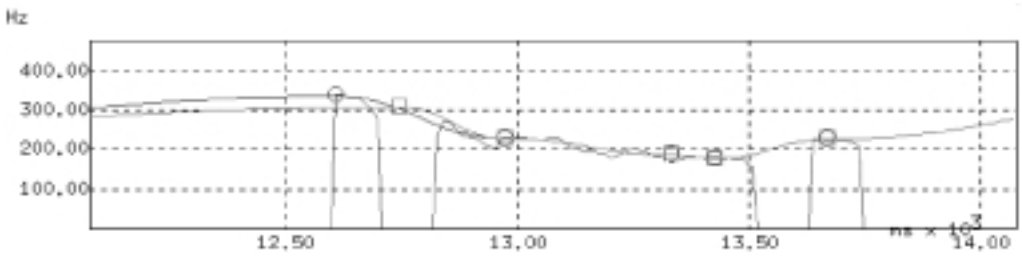

Figure 8.
Exemple de point mal placé

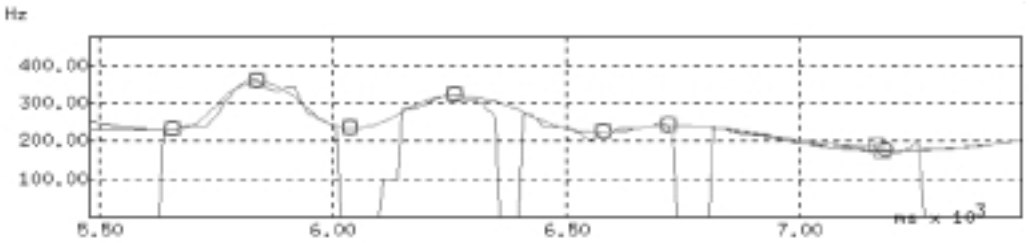

Figure 9.
Exemple de point redondant

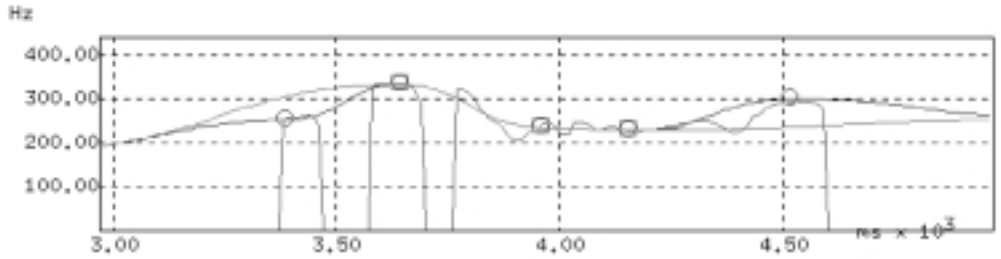

Figure 10.
Exemple de points manquants avant et après pause

Conclusion

Nous avons présenté une évaluation quantitative et qualitative de l'algorithme de stylisation mélodique MOMEL sur un grand corpus de parole en français et en italien. Les résultats obtenus montrent que MOMEL est un algorithme qui produit de bonnes performances puisqu'on n'observe que 3,90% de bruit (points erronés) et 6,09% de silence (points manquants). De plus, l'analyse qualitative révèle des erreurs systématiques (points redondants et points manquants avant ou après les pauses) qui pourraient être aisément corrigées dans l'algorithme, conduisant à une forte réduction du bruit et une quasi-disparition du silence. D'autre part, nous avons constaté une cohérence des stratégies de correction des experts sur leurs langues maternelles mais des différences significatives de la correction sur les langues secondes. Afin de valider ces dernières observations ainsi que les procédures de corrections employées, nous pensons qu'il est nécessaire de réaliser une analyse systématique de la part de variabilité de l'expert sur la totalité des corrections qu'il entreprend, notamment par une étude préalable visant à comparer la correction d'un même échantillon de phrases par différents experts.

Remerciements

Les auteurs remercient Robert Espesser (auteur des programmes d'édition et de manipulation du signal utilisés dans cette étude) pour son aide constante, Corine Astesano et Fabienne Courtois pour la correction d'une partie du corpus, ainsi que Daniel Hirst pour ses conseils. Une pensée émue va à Fabienne, tragiquement disparue avant la publication de cet article.

Bibliographie

AASA, A., STANGERT, E. (1996). Prosodic Analysis of Swedish within the Multext project, *Fonetik* 96, Nässlingen.

ASTESANO, C., ESPESSER, R., HIRST, D.J., LLISTERRI, J. (1997). Stylisation automatique de la fréquence fondamentale : une évaluation multilingue. *4ème Congrès Français d'Acoustique,* Marseille, p. 441-444.

CHAN, D., FOURCIN, A., GIBBON, D., GRANSTRÖM, B., HUCVALE, M., KOKKINAKIS, G., KVALE, K., LAMEL, L., LINDBERG, B., MORENO, A., MOUROPOULOS, J., SENIA, F., TRANCOSO, I., VELD, C., ZEILIGER, J. (1995). EUROM-A spoken language resource for the EU. *Proceedings of the 4th European Conference on Speech Communication and Speech Technology, Eurospeech'95.* Madrid. vol. 1, p. 867-870.

COHEN, A., 't HART, J. (1965). Perceptual analysis of intonation pattern. *Actes du 5ème Congrès International d'Acoustique,* Liège, p. 1-4.

D'ALESSANDRO, C., MERTENS, P. (1995). Automatic Pitch Contour Stylisation Using a Model of Tonal Perception. *Computer, Speech and Language*, 9, p. 257-288.

DI CRISTO, A. (1985). *De la microprosodie à l'intonosyntaxe.* Thèse d'état, Université de Provence.

DI CRISTO, A., HIRST, D. (1986). Modelling French Micromelody : Analysis and Synthesis, *Phonetica,* 43, p. 11-30.

HAMON, C., MOULINES, E., CHARPENTIER, F. (1989). A diphone system based on time-domain prosodic modifications of speech. *Proceedings of ICASSP'89,* p. 238-241.

HIRST, D.J., ESPESSER, R. (1993). Automatic Modelling of Fundamental Frequency using a quadratic spline function, *Travaux de l'Institut de Phonétique d'Aix,* 15, p. 75-85.

HIRST, D.J., DI CRISTO, A., ESPESSER, R. (2000). Levels of representation and levels of analysis for the description of intonation systems. In Horne, M. (ed.), *Prosody: Theory and Experiment,* Dordrecht: Kluwer Academic Publishers.

ROSSI, M. (1971). Le seuil de glissando ou seuil de perception des variations tonales pour les sons de la parole. *Phonetica,* 23, p. 1-33.

ROSSI, M. (1978). La perception des glissando descendants dans les contours prosodiques. *Phonetica,* 35, p. 11-40.

't HART, J., COLLIER, R. (1975). Integrating different levels of intonation analysis. *Journal of Phonetics,* 3, p. 235-255.

't HART, J., COLLIER, R., COHEN, A. (1990). *A Perceptual Study of Intonation : an experimental-phonetic approach to speech melody,* Cambridge : Cambridge University Press.

TAYLOR, P. (1993). Automatic Recognition of Intonation from F0 contours using the Rise/Fall/connection. *Proceedings of Eurospeech 93,* 2, Berlin, p. 789-792.

TAYLOR, P. (1994). The Rise/Fall/Connection Model of Intonation. *Speech Communication,* 15, 1-2, p. 169-186.

VAN RIJSBERGEN, C. J. (1979). *Information Retrieval.* 2nd edition, London : Butterworths.

VÉRONIS, J., HIRST, D.J., ESPESSER, R., IDE, N. (1994). NL and speech in the MULTEXT project, *Proceedings of the AAAI'94 Workshop on the Integration of Natural Language and Speech,* Seattle, p. 72-78.