

Llisterri, J. (Ed.). (1996). *Prosody tools efficiency and failures*. WP 4 Corpus. T4.6 Speech markup and validation. Deliverable 4.5.2. Final Report. LRE Project 62-050 MULTEXT.

[http://liceu.uab.cat/~joaquim/publicacions/Prosody\\_tools\\_96.pdf](http://liceu.uab.cat/~joaquim/publicacions/Prosody_tools_96.pdf)

**MULTEXT - LRE Project 62-050**

# **Prosody Tools Efficiency and Failures**

**WP 4 Corpus  
T4.6 Speech Markup and Validation  
Deliverable 4.5.2**

**Final version**

**15 October 1996**

**Edited by**

**Joaquim Llisterri**

Departament de Filologia Espanyola  
Universitat Autònoma de Barcelona

**Contributors:**

**Corine Astesano, Robert Espesser  
and Daniel J. Hirst**

(Laboratoire Parole et Langage, CNRS)

**Lorraine Baqué**

(Departament de Filologia Francesa i Romànica, UAB)

**Mònica Estruch and Juan M. Garrido**

(Departament de Filologia Espanyola, UAB)

**Eva Strangert and Anna Aasa**

(Department of Linguistics, Umeå University)

## **1.- Introduction**

This report describes the work carried out by three different groups (Laboratoire Parole et Langage, CNRS, Aix-en-Provence, France; Grup de Fonètica, Departament de Filologia Espanyola, Universitat Autònoma de Barcelona, Spain; Department of Phonetics, Umeå University, Sweden) within task 4.6. (Speech Markup and Validation), devoted to assess the efficiency as well as the failures of the prosody tools developed under tasks 2.6 (Prosody Tools) and 2.7 (Post-editing Tools).

The report is organized in six parts. After the introduction, a brief description of the multilingual corpus used for the task is provided in part two. Part three describes the markup and validation tasks that have been carried out with selected sub-corpora in 5 languages, and the results concerning the efficiency and failure of the tools used are presented in part four. The individual reports for each language are collected in part five. Finally, a general summary of the results and some suggestions for improvement are presented in part 6.

The editor of the report wishes to acknowledge the technical assistance provided by Robert Espesser and the coordinating work carried out by Corine Astesano, both at the Laboratoire Parole et Langage (CNRS, Aix-en-Provence, France).

## 2.- The multilingual speech corpus

### 2.1.- EUROM

One of the tasks within MULTEXT has been the enhancement with prosodic annotation of materials from the EUROM.1 speech database produced under the ESPRIT projects SAM 2589 *Multilingual Speech Input/output Assessment, Methodology and Standardisation* and SAM-A 6819 *Speech Technology Assessment for Multilingual Applications*.

EUROM.1 is a set of corpora originally recorded for Danish, Dutch, English, French, German, Italian, Norwegian and Swedish. Greek, Portuguese and Spanish were added later. The full corpus comprises, for each language, the following materials:

- C(C)VC(V) combinations in isolation and in context covering between 60 and 100 combinations per language;
- 100 numbers covering the phonotactic distribution found in this type of material for each language;
- 40 short passages composed of thematically linked sentences, and
- 50 filler sentences to compensate for lack of phonemic coverage in the passages.

These materials are recorded by 60 subjects for each language, divided into three different groups:

- Many Talker set: each subject read 4 passages, 5 sentences and 100 numbers;
- Few Talker set: each subject read 15 passages, 25 sentences, CVC isolated words, and repeated the 100 numbers five times
- Very Few Talker set: each subject read the CVC words embedded in 5 different carrier sentences and repeated the carrier words from the carrier phrases 5 times.

The recordings are of high acoustic quality (sampling rate 16 KHz, 16 bit sampling, recorded in anechoic room) and were specifically designed for use in speech technology assessment.

A general overview of EUROM can be found in Chan *et al.* (1995) and a complete description is available in Sherwood and Fuller (1992). The EUROM Home Page is located at URL: <http://www.phon.ucl.ac.uk/resource/eurom.html>.

## 2.2.- EUROM subset annotated and validated in MULTEXT

### 2.2.1.- Overview

It is beyond the scope of MULTEXT to provide annotation for all the EUROM corpus in all languages. Thus, only some of the languages which are represented within the consortium have been retained: English, French, German, Spanish and Swedish. Since the aim is to provide a prosodic annotation that could be linked to other levels of linguistic annotation, only the 40 passages are considered. Each passage is composed of 5 thematically linked sentences, and it is then necessary to introduce further restrictions on the material to be annotated so that the size of the corpus is proportional to the efforts allocated to the task. The amount of materials annotated and validated in MULTEXT is summarized in table 1:

Language	Passages in EUROM.1	Sentences	Speakers
English	O0-O9+P1-P5+Q0-Q9+R1-R5+P0+P6-P9+R0+R6-R9 = 40 passages	75 sentences x 2 speakers + 25 sentences x 2 speakers = 200 sentences	3 male 1 female
French	O0-O9+P0-P9+Q0-Q9+R0-R9 = 40 passages	50 sentences x 4 speakers = 200 sentences	1 male 3 female
German	01+04+05+06+09+14+15+16+17+19+20+23+27+28+33+34+36+37+38+40 = 20 passages	100 sentences x 2 speakers = 200 sentences	1 male 1 female
Spanish	O0-O9+P0-P9+Q0-Q9+R0-R9 = 40 passages	75 sentences x 2 speakers + 50 sentences x 1 speaker = 200 sentences	1 male 1 female
Swedish	5 passages	25 sentences x 8 speakers = 200 sentences	4 male 4 female
<b>Total</b>	<b>145 passages</b>	<b>1000 sentences</b>	<b>17 speakers</b>

Table 1: Total amount of materials annotated and validated in MULTEXT

### 2.2.2.- Collection of the materials

It was expected at the beginning of the project that the full EUROM corpus would be available in CD-ROM by Autumn 1993. Unfortunately for reasons external to the MULTEXT consortium the production of CD-ROMs has not been possible for all languages and, consequently, much time and effort has been spent in locating the different Institutes which had copies of the materials, contacting the relevant persons, trying to cope with a variety of formats and making arrangements for sending the recordings to CNRS, who had to make some preparations to make the files compatible

with the tools before distributing them to all the partners involved in the task. The status of the EUROM corpus for each language at the time of performing the task described in this report is explained below.

#### 2.2.2.1.- English

The English corpus has been copied from CD-ROMs, which were made available to MULTEXT by University College London, coordinator of the SAM consortium.

#### 2.2.2.2.- French

The French corpus was made available to the CNRS by the *Institut de la Communication Parlée* (Grenoble).

#### 2.2.2.3.- German

The German data was located at Bielefeld University through contacts with Saarbrueken University while CD-ROMs were still in preparation. The final CD-ROMs were made available to MULTEXT by University College London, coordinator of the SAM consortium.

#### 2.2.2.4.- Spanish

The Spanish recordings have been made available to UAB by the Universitat Politècnica de Catalunya (Barcelona). They were transferred from DAT tapes to CD-ROMs by the Computer Center at the UAB and were also sent to CNRS for central storage of the corpus.

#### 2.2.2.4.- Swedish

Swedish data was provided to Umeå University by KTH (Stockholm).

The assistance received from the partners of the SAM consortium and specially of Adrian Fourcin, coordinator of the project, is to be gratefully acknowledged.

### **2.2.3.- Preparation and distribution to the partners**

Recordings and orthographic transcriptions of the five languages were centralized at the CNRS (Aix-en-Provence), where they were converted - with the technical assistance of Robert Espesser- into in a file format which is compatible with the MULTEXT prosodic tools and were redistributed to the partners, keeping a central archive at CNRS.

## **3.- Prosodic markup, alignment and validation**

### **3.1.- Definition of the task**

#### **3.1.1.- Markup levels**

As specified in the MULTEXT technical annex and in the reports describing the tools developed in task 2.6, two levels of markup have been applied to the subset of the EUROM corpus described in part 2.:

- Fundamental frequency (F0) detection in order to obtain the melodic contour - F0 curve- of each sentence by means of the DETECT-F0\_PEIGNE program.
- Automatic modelling of the fundamental frequency curves as a sequence of target points (Hz, ms), obtaining a stylized representation of the melodic contour by means of the MOMEL (*MOdélisation de MELodie*) program (Hirst & Espesser, 1993).

#### **3.1.2.- Alignment**

Alignment has been performed manually, using the post-editing tools developed in task 2.7 by CNRS. Two elements of the orthographic transcription have been aligned with the speech signal using the speech editor MES:

- Word boundaries.
- Onset of stressed vowels, except in language such as French where lexical accent does not exist.

#### **3.1.3.- Validation**

Since the alignment has been manually performed, the validation task has been only concerned with the automatic modelling of the fundamental frequency contours performed by MOMEL. During the validation process, the sentence is resynthesized with the PSOLA algorithm using the target points automatically obtained by MOMEL. A perceptual comparison between the result and the original sentence is carried out, and the target points are adjusted, if necessary, in order to obtain a melodic contour close to the original sentence and perceptually acceptable to a native speaker.

### **3.2.- Procedure**

#### **3.2.1.- Training**

On January 28 1995 a one-day workshop took place in Aix-en-Provence organized by CNRS. The aim of the workshop was to gain hands-on experience in using MULTEXT tools for prosody annotation using the speech corpora collected by that time. It was attended by the CNRS and the UAB and work was carried out for French and Spanish.

The workshop constituted a first attempt at using EUROM materials combined with the tools developed by CNRS under task 2.6 and 2.7. and provided a good opportunity to have a realistic view of the problems related to corpus preparation and distribution and to the common decisions that had to be taken regarding the methodology for annotation and validation. It also provided a test of the time that was needed for the task and the amount of work that could be performed considering the available resources.

The tools developed at CNRS were tested on French and Spanish samples of EUROM by two UAB phoneticians who were afterwards involved in the annotation and validation tasks. No problems were found during the process, and it was shown that the annotation and post-editing tools could be efficiently used with a minimum training. The F0 detection tool (DETECT-F0\_PEIGNE), and the automatic modelling tool (MOMEL) were tested together with the general-purpose speech editor (MES) with the files are used in the project, trying to reproduce the actual task described in figure 2. It was shown that they can perform an adequate F0 extraction, modelling and resynthesis for perceptual validation; the speech editor provides good facilities for alignment with the text. Given the experience in these two languages, no difficulties were foreseen concerning the applications of the tools to other languages.

### **3.2.2.- Work for individual languages**

Work for English and German was performed at the CNRS (Aix-en-Provence, France), French and Spanish were analysed at the *Universitat Autònoma de Barcelona* (Spain) and work on Swedish was carried out at Umeå University (Sweden). Where possible, the same methodology has been applied for all languages, following the procedures described in figure 1. The Swedish group has also provided an INTSINT (International Transcription System for Intonation) (Hirst & Di Cristo, in press) coding and the corresponding evaluation.

A detailed description of the work carried out for each language is provided in part 5 of the report.

The process of markup, alignment and validation that has been applied to the data is described in figure 2. The first step consists on the automatic detection of the F0 contour with the DETECT-F0\_PEIGNE algorithm. The result is a series of F0 points showing the temporal evolution of the fundamental frequency in the original sentence. An automatic stylization is then performed by MOMEL, detecting a certain number of target points (TPs). In order to validate the correct detection of target points, the stylized sentence is resynthesized using the PSOLA algorithm and a perceptual evaluation is carried out. If necessary, target points are manually corrected and the process of resynthesis and evaluation is repeated until the synthesized sentence is acceptable to a native speaker and perceived as a close equivalent of the original one.

Once the stylization has been validated, the sentence is manually aligned with word boundaries and with the onset of stressed vowels in the orthographic transcription.

The final result is a stylized representation of the melodic contour which has been perceptually validated and which is also aligned with word boundaries and onsets of stressed vowels.

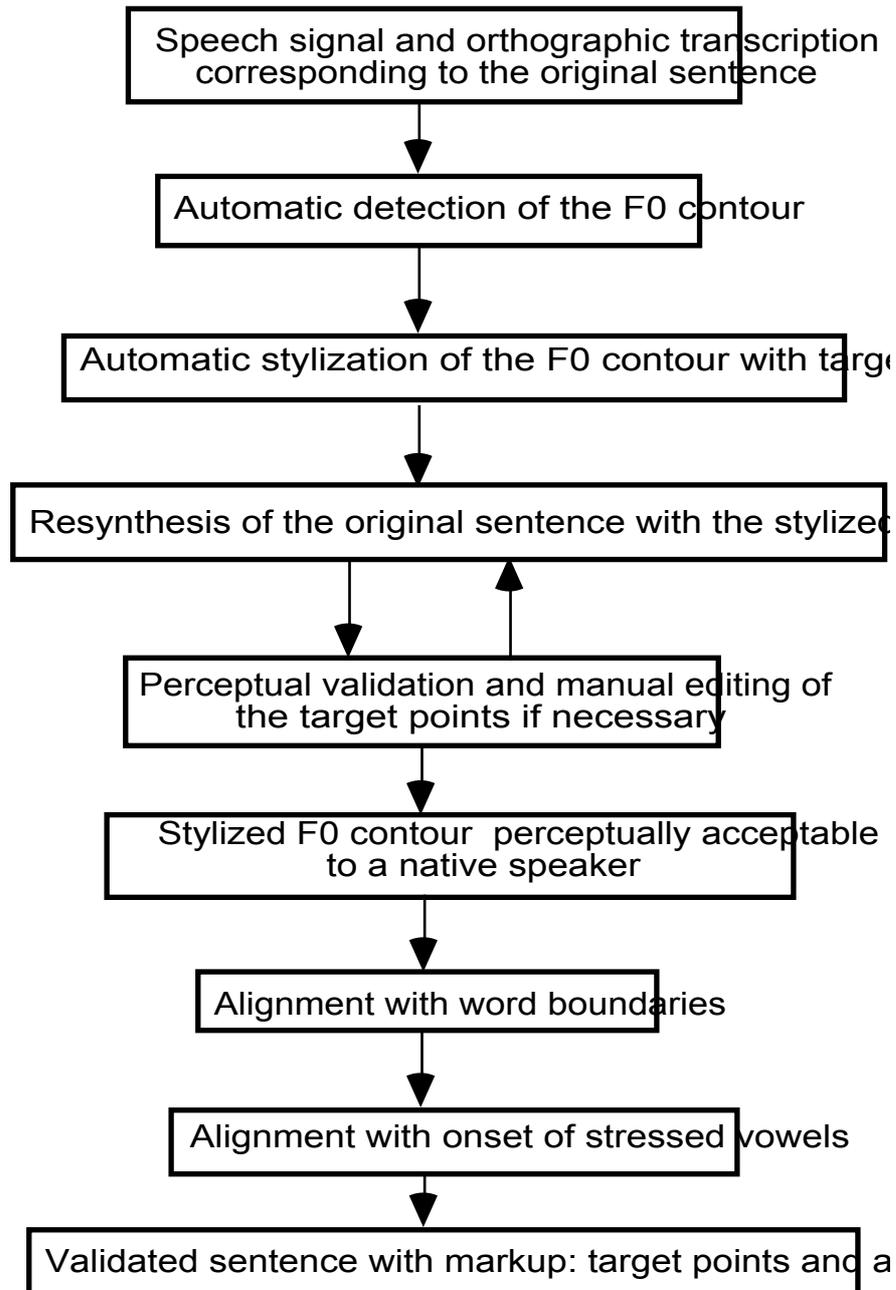


Figure 2: Markup, alignment and validation of MULTTEXT data

A further step not described in figure 2 would be the automatic generation of another level of prosodic representation which consists in a prosodic coding using a set of symbols designed to describe F0 movements either in absolute or in relative terms. Such coding is known as INTSINT (International Transcription System for Intonation) (Hirst and Di Cristo, in press) and can be used to generate a new synthesized version of the sentences which can also be perceptually evaluated in order to validate the symbolic coding. This process has been applied to Swedish within MULTTEXT, and the results are presented in section 5.5. of this report.

## 4.- Tools efficiency and failures

The validation process has been applied to the automatic stylization performed by MOMEL. In order to obtain comparable results for the different languages considered, three categories of problems related to the detection and placement of target points have been defined:

- Missing target points: target points have to be manually added in positions where the automatic detection performed by MOMEL fails to introduce them and the presence of these target points is necessary to obtain a stylized representation which sounds acceptable to a native speaker when compared with the original sentence.

Three different contexts have been defined for those cases:

- Missing target points in sentence-initial position
  - Missing target points in middle-sentence position
  - Missing target points in sentence-final position  
Missing points in rising contours and missing points in falling contours have been distinguished in this context.
- Extra target points: target points introduced by MOMEL which have to be manually suppressed in order to obtain a stylized representation which is judged as acceptable by a native speaker when compared to the original sentence.
  - Moved target points: target points which have to be shifted since MOMEL does not detect them in the correct position when comparing the original sentence with the resynthesis of the stylized contour.

Three different moved target points have been defined:

- Too high: target points which have to be shifted towards lower frequencies
- Too low: target points which have to be shifted towards higher frequencies
- Horizontally: target points which have to be shifted in the time domain

### 4.1.- Results for individual languages

This section presents a summary of the main results for each of the languages considered in the project. A more complete evaluation of the performance of the prosodic tools in each of the languages can be found in the individual language reports compiled in part 6.

### **4.1.1. English**

Errors in the automatic detection of F0 target points in the English corpus have been detected at major syntactic boundaries and before and after pauses. These errors are more frequent in final rising contours than in final falling ones.

### **4.1.2. French**

It is observed for French that most of the errors in the automatic detection of F0 target points correspond to cases where a target point was not detected and was judged as necessary in order to obtain a close approximation to the original sentence acceptable to a native speaker. Within this category, the highest number of errors corresponds to target points located in final rising contours (29.5%), followed by target points in initial position (12.9%) and by target points in final falling contours (9.14%). It should be noted that these positions occur before or after pauses.

### **4.1.3. German**

The highest number of errors in the automatic detection of F0 target points in the German corpus occurs in final rising contours (44.44%), followed by target points in initial position (16.83%). In those cases, target points not detected by MOMEL had to be added in order to obtain a close approximation to the original sentence. Major problems are then found at sentence and prosodic boundaries due to the presence of pauses.

### **4.1.4. Spanish**

Most of the errors in the Spanish corpus were located in target points in final rising contours which were not detected by MOMEL and had to be manually added in order to obtain a perceptually good approximation to the original sentence (72.85%); this category is followed by detection errors corresponding to missing target points in initial position (15.43%). Those errors are linked to beginning and end of sentences in the passages where a pause was realized by the speakers.

### **4.1.5. Swedish**

The work for Swedish has been mainly concerned with the evaluation of the automatic INTSINT coding, judging the match between original sentences and sentences resynthesized using the information from the prosodic coding. Deviations between original and coded sentences were judged considering shifts of meaning between both versions.

Stylistic changes, changes of prominence relations and shifts of word accent have been detected in the comparisons. In the first case, missing F0 target points in rising contours at the beginning of the sentence and missing points at falling contours at the end are detected. The most frequent errors involve prominences; in those cases, a degrading of prominence resulting from missing F0 peaks in the contours is observed in the sentences resynthesized using INTSINT coding. A too heavy smoothing of the F0 contour is also observed as a general problem.

## 4.2.- Comparison across languages

In order to perform a general comparison across languages, the differences between the original F0 contour and the stylized contour derived from automatically detected F0 target points has been computed for each language. The results are shown in table 2.

Language	Number of passages	% of deviation between the original F0 contour and the modeled contour	Total number of target points	Total duration of all the passages	Number of target points per second
English	150	5.6	8.680	2635	3.29
French	100	6	6747	2190	3.08
German	200	4.7	13.995	4420	3.17
Spanish	40	4.8	2.495	863	2.89
Swedish	150	6.1	10869	3257	3.32

Table 2. Cross-language comparison of the differences between the original F0 contour and the contour derived the target points automatically detected by MOMEL. Differences have been computed over the total number of passages recorded for each language (with 10 speakers, 5 male and 5 female) except for the case of Spanish in which 40 passages have been used.

It can be seen from the table above that differences between the original and the modeled contours never exceed 6.1%, and are, on average, in the region of 5.4%. German is the language for which the best match is obtained in general terms, followed by Spanish, while the lowest scores are found for Swedish and French

Since French, German and Spanish have been evaluated taking into account the same typology of errors described in 4, results for the three languages can be easily compared.

The first category of errors concerns target points which were not detected by MOMEL and which had to be manually introduced in order to obtain an approximation to the original F0 contour acceptable to a native speaker. The results (averaged across speakers) for each language are shown in figure 3:

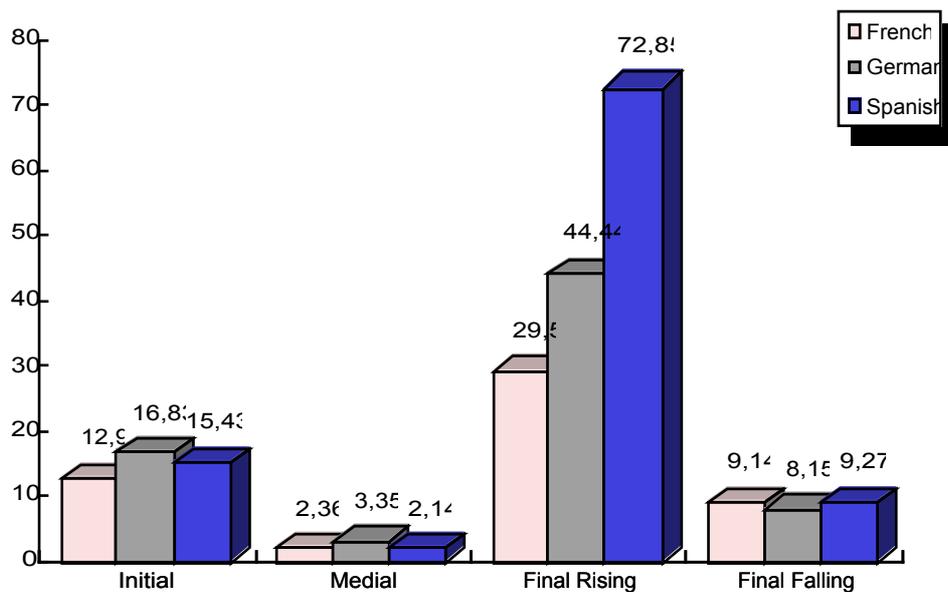


Figure 4.1: Percentage of target points not detected by MOMEL  
(Averaged value for the speakers validated for each language)

It can be observed in figure 3 that in the three languages taken into account, the highest number of errors appears in target points located in final rising contours, followed by errors in target points located in initial position and in final falling contours. It is important to note that a very low number of errors is found for target points in middle sentence position. Differences between languages are found in final rising contours, while in the other positions the performance of MOMEL does not seem to be highly dependent on the language.

The second category of errors is related to target points which were detected by MOMEL and which were judged as unnecessary by the validators. The results for each of the three languages considered here are presented in figure 4.

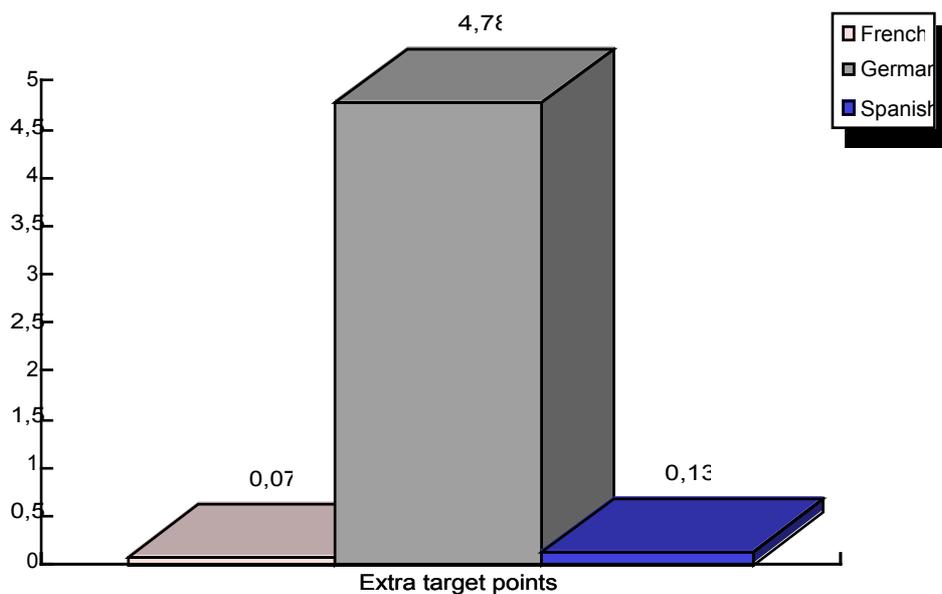


Figure 4: Percentage of extra target points added by MOMEL  
(Averaged value for the speakers validated for each language)

A clear difference between German on the one hand and Spanish and French on the other can be observed in the figure. However, as remarked in the report for German (cf. 5.3.5.5) those extra target points did not often hinder the understanding of the sentence. The low number of errors encountered in this category shows a good performance of MOMEL in the modelization of F0 curves as far as extra target points are concerned.

The third category of problems concerns the target points that were moved by the validators in order to obtain a better approximation to the original sentence. A comparison between French, German and Spanish can be seen in figure 5.

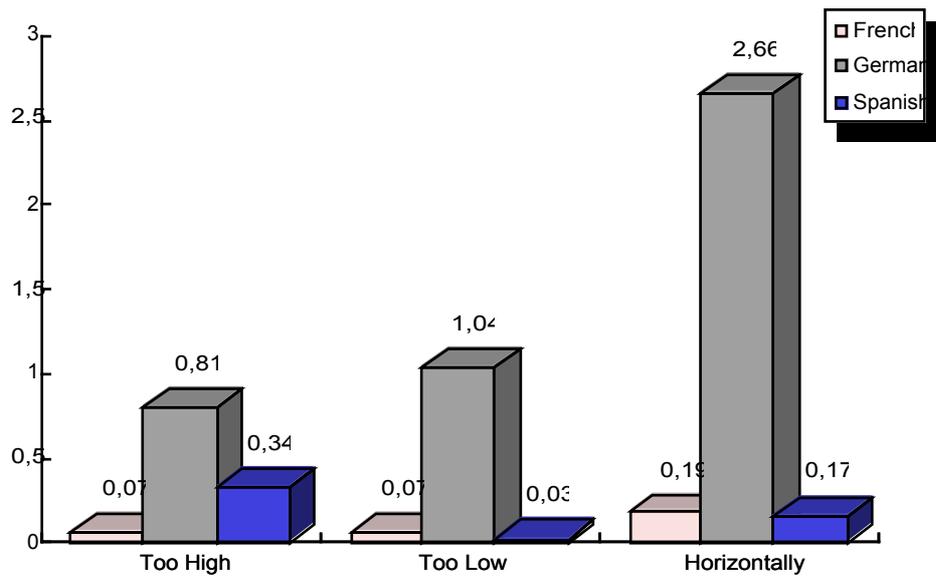


Figure 5: Percentage of moved target points  
(Averaged value for the speakers validated for each language)

German is the language in which more target points were moved when compared to French and Spanish, but it has to be noted that the criterion followed by the validator for this language (cf. 5.3.5.6.) was to move target points rather than to create new ones. However, even in the case of German, the percentage of moved target points is lower than 3%, showing a good performance of MOMEL in this respect.

## 5.- Individual languages reports

### 5.1.- English

Corine Astesano

Institut de Phonétique, URA 261 CNRS, Université de Provence

29 Av. Robert Schuman, F-13621 Aix-en-Provence CEDEX 1, France

Fax : +33 - 42 59 50 96; Tel : 42 95 36 37; E-mail : corine.astesano@lpl.univ-aix.fr

With contributions from Andrea Levitt (Haskins Laboratories and Wellesley College, USA) and Sophie Herment (Laboratoire Parole et Langage, CNRS, Aix-en-Provence)

#### 5.1.1.- Corpus and speakers

The 4 speakers used for the English analysis are taken from the Few Talker set. The distribution of passages and speakers is presented in table 3.

Speaker	Block	Speaker	Block	Speaker	Block
FA	O0	FB	Q0	FC	P0
FA	O1	FB	Q1	FC	P6
FA	O2	FB	Q2	FC	P7
FA	O3	FB	Q3	FC	P8
FA	O4	FB	Q4	FC	P9
FA	O5	FB	Q5	FD	R0
FA	O6	FB	Q6	FD	R6
FA	O7	FB	Q7	FD	R7
FA	O8	FB	Q8	FD	R8
FA	O9	FB	Q9	FD	R9
FA	P1	FB	R1		
FA	P2	FB	R2		
FA	P3	FB	R3		
FA	P4	FB	R4		
FA	P5	FB	R5		

Table 3: Distribution of speakers and passages in the English corpus

#### 5.1.2.- Methodology

The English analysis has been done on a different methodological basis, namely more qualitative than quantitative. Nevertheless, it is interesting to note that the expert comes to the same general conclusions as for the other languages involved in the project.

#### 5.1.3.- Results

The major detection problems noticed were located at major prosodic boundaries and at the beginning or the end of sentences, namely before and after the realization of pauses. Those problems were more numerous in the cases of final risings than final fallings. However, those detection problems never hindered the understanding or the meaning of the sentences.

The expert suggest as well that the presence of pauses should be taken into consideration by the MOMEL algorithm.

## 5.2.- French

Lorraine Baqué

Departament de Filologia Francesa i Romànica, Universitat Autònoma de Barcelona

Facultat de Filosofia i Lletres. Edifici B, 08193 Bellaterra, Barcelona, Spain

Tel : +34.3.581.14.10; Fax : +34.3.581.20.01; E-mail : lorraine@prosodia.uab.es

### 5.2.1.- Introduction

The aim of this task was to make a perceptual validation of the automatic modeling of the F0 curve from the speech signal. The prosody tools developed within MULTTEXT derive automatically a symbolic representation of the intonation contour from the speech signal.

The process for the analysis includes the following steps:

- automatic modeling of the F0 curve from the speech signal
- automatic stylisation of the curve with a sequence of target points (Hz, ms)
- perceptual validation of the synthesis of the paragraphs
- hand validation and correction of the errors done by the system

### 5.2.1.- Corpus

The EUROM.1 prompting texts consist of three parts from which, in this case, only passages have been analyzed. There are 40 passages consisting each one of five task-related sentences whereas the sentences are designed to compensate the uneven diphone distribution of the passages. Each passage was displayed as a single block and speakers tried to produce the block with natural intonation.

The forty passages were distributed as shown in table 4:

Speaker	Block	Number	Speaker	Block	Number	Speaker	Block	Number
BF	O0	1282	BO	P5	1647	VI	R0	1214
BF	O1	1283	BO	P6	1659	VI	R1	1215
BF	O2	1295	BO	P7	1660	VI	R2	1227
BF	O3	1296	BO	P8	1672	VI	R3	1228
BF	O4	1308	BO	P9	1673	VI	R4	1240
BF	O5	1309	SL	Q0	1145	VI	R5	1241
BF	O6	1321	SL	Q1	1146	VI	R6	1253
BF	O7	1322	SL	Q2	1158	VI	R7	1254
BF	O8	1334	SL	Q3	1159	VI	R8	1266
BF	O9	1335	SL	Q4	1171	VI	R9	1267
BO	P0	1620	SL	Q5	1172			
BO	P1	1621	SL	Q6	1184			
BO	P2	1633	SL	Q7	1185			
BO	P3	1634	SL	Q8	1197			
BO	P4	1646	SL	Q9	1198			

Table 4: Distribution of speakers and passages in the French corpus

## 5.2.2.- Speakers

Four speakers have been analyzed, three males (BF, BO, SL) and one female (VI).

## 5.2.3.- Results

As a result of this analysis, some problems have been found and classified in three different tables: cases in which target points were missed, cases in which there were extra target points and finally, cases in which the existent target points had to be moved. The results have also been classified according to the four speakers which have been analysed.

### 5.2.3.1.- Missing target points

The first category of problems found are cases in which there should be target points in places where the system has not detected them.

Three different positions have been distinguished: initial (for the beginning of sentences), final (for the end of sentences, distinguishing between cases with a rising end from the ones with a falling one) and middle (for points in the middle of sentences).

Table 5 summarizes the results:

	Initial	Middle	Final	
			Rising	Falling
Speaker BF	6/64 (9.38%)	11/509 (2.16%)	7/16 (43.75%)	3/48 (6.25%)
Speaker BO	12/90 (13.33%)	16/479 (3.34%)	11/30 (36.67%)	6/60 (10.00%)
Speaker SL	12/81 (14.81%)	11/568 (1.94%)	4/26 (15.38%)	6/55 (10.91%)
Speaker VI	10/71 (14.08%)	9/444 (2.03%)	4/18 (22.22%)	5/53 (9.43%)

Table 5: Missing target points (in absolute numbers and in percentages) in the French corpus

The same results are visually presented in figure 6:

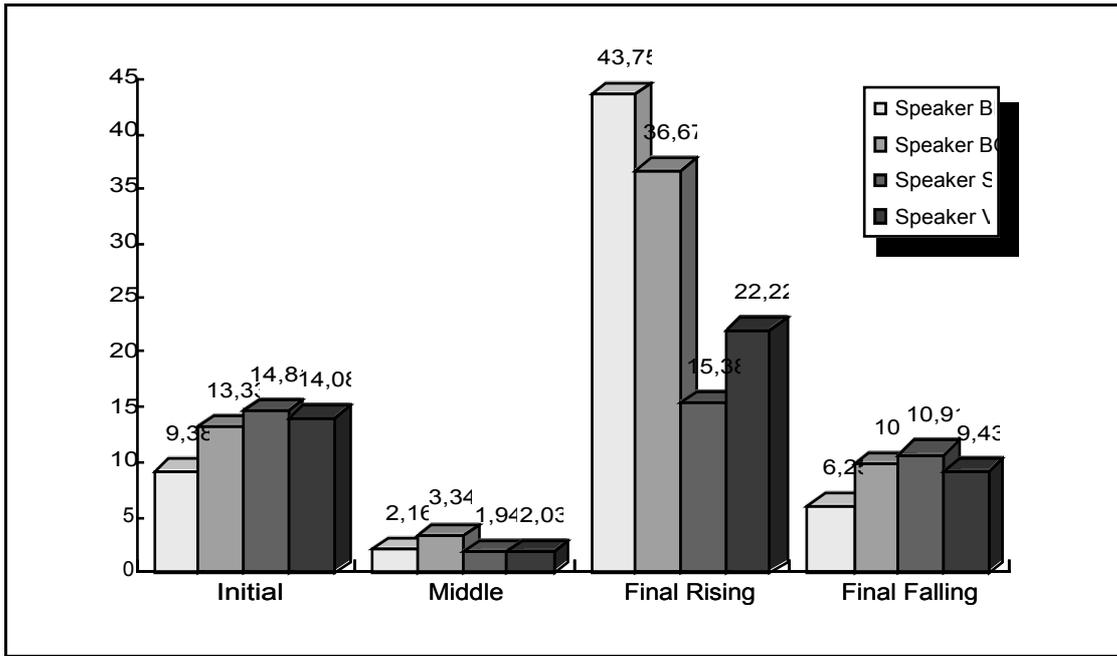


Figure 6: Percentage of missing target points in the French corpus

Let us give an example of each case:

- initial position: BO P5 1647 (*Il est recommandé...*)
- middle position: BF O7 1322 (*à Saint-Barnabé*)
- final 'rising' position: SL Q5 1172 (*de service*)
- final 'falling' position: VI R8 1266 (*plains pas*)

### 5.2.3.2.- Extra target points

This second category is much more reduced than the first one. It includes cases in which the system has recognised too many target points, that is to say, target points in places where they should not appear.

The number of errors found is shown in table 6 and in figure 7.:

<b>Speaker BF</b>	2/637 (0.31%)
<b>Speaker BO</b>	0/659 (0%)
<b>Speaker SL</b>	0/730 (0%)
<b>Speaker VI</b>	0/586 (0%)

Table 6: Extra target points  
(in absolute numbers and in percentages) in the French corpus

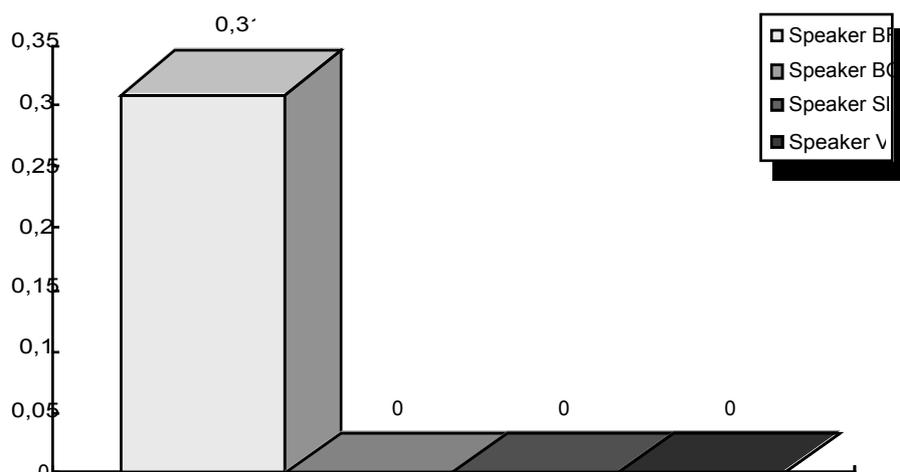


Figure 7: Percentage of extra target points in the French corpus

An example:

BF O8 1334 (*ma femme*)

### 5.2.3.3.- Moved target points

Finally, another group of errors includes the cases in which the target points were detected by the system but in an incorrect position. Therefore, these target points have been moved. The classification has been done according to the situation of the target point: whether it was too high or too low, or whether it had to be moved horizontally in the time domain.

	Too High	Too Low	Horizontally
Speaker BF	2/637 (0.31%)	0/637 (0%)	1/637 (0.16%)
Speaker BO	0/659 (0%)	1/659 (0.15%)	2/659 (0.30%)
Speaker SL	2/730 (0.27%)	1/730 (0.14%)	1/730 (0.14%)
Speaker VI	0/586 (0%)	0/586 (0%)	1/586 (0.17%)

Table 7. Moved target points (in absolute numbers and in percentages) in the French corpus

Examples of these cases are the following:

- too high: SL Q0 1145 (*bronzage*)
- too low: SL Q1 1146 (*assis sur*)
- horizontally: BF O8 1334 (*Elle a horreur*)

The same data is visualized in figure 8.:

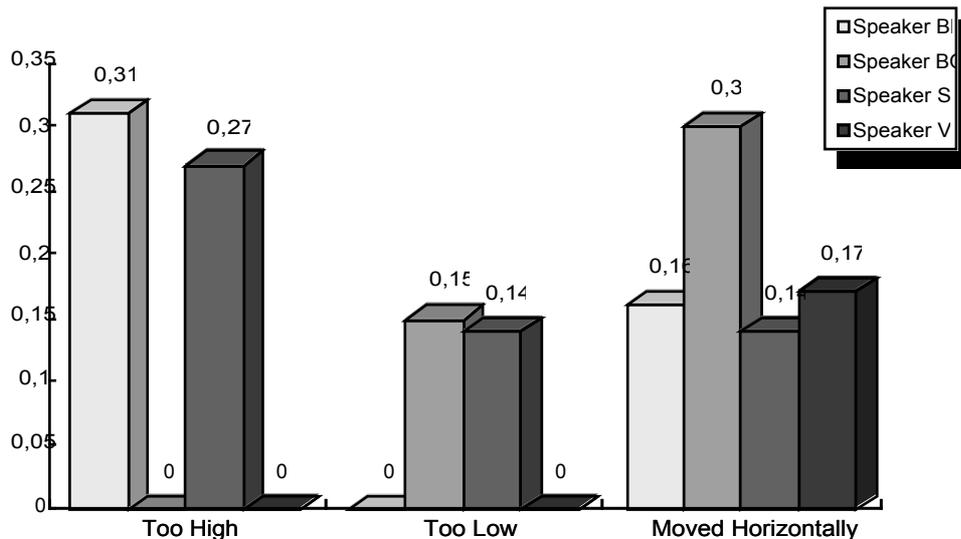


Figure 8: Percentage of moved target points in the French corpus

## 5.3.- German

Corine Astesano

Institut de Phonétique, URA 261 CNRS, Université de Provence

29 Av. Robert Schuman, F-13621 Aix-en-Provence CEDEX 1, France

Fax : +33 - 42 59 50 96; Tel : 42 95 36 37; E-mail : corine.astesano@lpl.univ-aix.fr

### 5.3.1.- Corpus

The EUROM.1 corpus for German is organised by blocks, i.e. not according to the speakers, like the Spanish or French corpuses. It comprises 40 blocks (01 to 40) read by 10 speakers of the Few Speaker set. However, the speakers are subdivided in two groups, each reading 20 passages: the first group is constituted by 2 female speakers and 3 male speakers; the second group is constituted by 3 female speakers and 2 male speakers.

### 5.3.2.- Speakers

For the analysis of the German corpus, we have chosen 2 speakers of the first group:

- Female speaker: initials: GA; sex: F; age: 27; nationality: German.

- Male speaker: initials: SM; sex: M; age: 25; nationality: German.

Both speakers read the same 20 blocks, namely: 01, 04, 05, 06, 09, 14, 15, 16, 17, 19, 20, 23, 27, 28, 33, 34, 36, 37, 38, 40.

### 5.3.3.- Methodology

We used the same methodology as the Barcelona team for Spanish and French, namely:

- automatic modeling of the F0 curve from the speech signal
- automatic stylisation of the curve with a sequence of target points (Hz, ms)
- perceptual validation of the synthesis of the paragraphs

- hand validation and correction of the errors done by the system.

To be more precise, we corrected by hand the MOMEL file 'COURANT.CB' and created a file 'EVALUATION.CB' which we consider to be closer to the original F0 curve. Also, for both speakers, we proceeded to the word-alignment. However, we neither proceeded to the alignment of the accented vowels, nor to the perceptual validation of the INTSINT transcription system.

### 5.3.4.- Results

As a result of this analysis, we classified the different problems encountered in three categories:

missing target-points, extra target-points and finally moved target-points.

#### 5.3.4.1.- Missing target points

The results concerning missing target points that were manually introduced are presented in table 8 and in figure 9.

	Initial	Middle	Final	
			Rising	Falling
<b>Speaker GA</b>	19/101 (18,81%)	48/1130 (4,25%)	4/9 (44,44%)	9/92 (9,78%)
<b>Speaker SM</b>	15/101 (14,85%)	26/1060 (2,45%)	4/9 (44,44%)	6/92 (6,52%)

Table 8: Missing target points (in absolute numbers and in percentages) in the German corpus

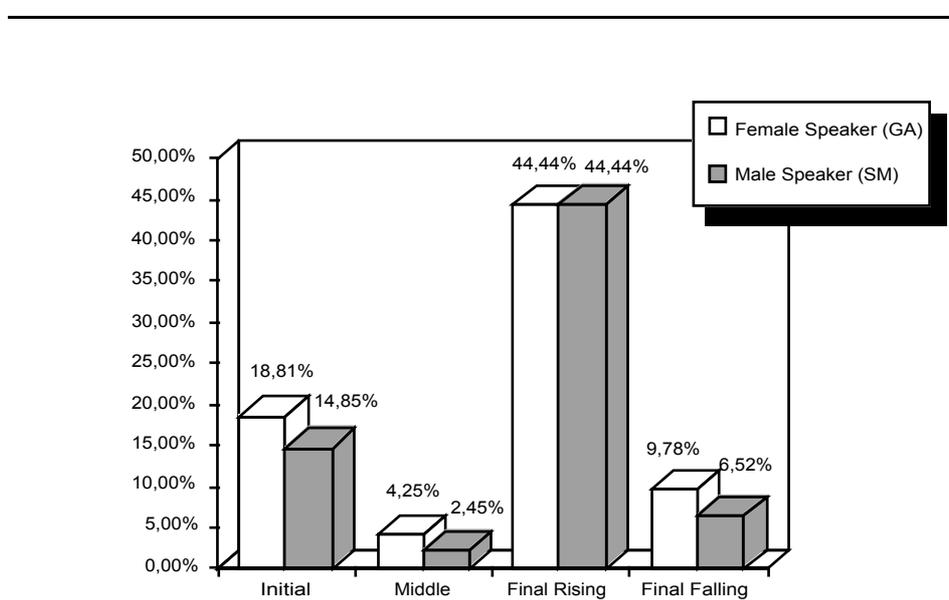


Figure 9: Percentage of missing target points in the German corpus

#### 5.3.4.2.- Extra target points

The absolute number and the percentage of extra target points introduced by MOMEL are shown in table 9 and in figure 10:

<b>Speaker GA</b>	65/1332 (4,88%)
<b>Speaker SM</b>	59/1262 (4,68%)

Table 9: Extra target points  
(in absolute numbers and in percentages)  
in the German corpus

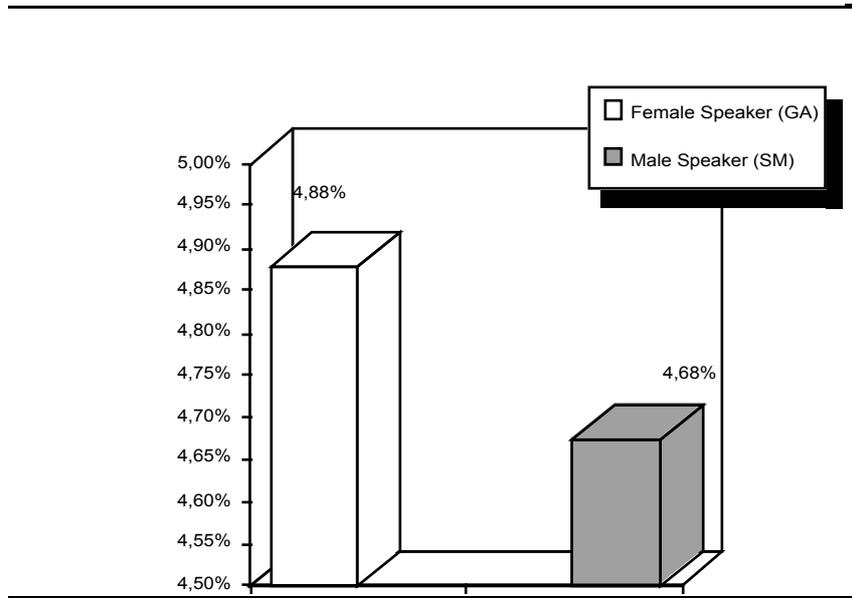


Figure 10: Percentage of extra target points in the German corpus

### 5.3.4.3.- Moved target points

Table 10 and figure 11 present the percentage and absolute number of target points which were manually moved in order to obtain a good perceptual approximation to the original sentences.

	<b>Too High</b>	<b>Too Low</b>	<b>Horizontally</b>
<b>Speaker GA</b>	9/1332 (0,68%)	12/1332 (0,90%)	37/1332 (2,78%)
<b>Speaker SM</b>	12/1262 (0,95%)	15/1262 (1,19%)	32/1262 (2,54%)

Table 5.3.3. Moved target points (in absolute numbers and in percentages) in the German corpus

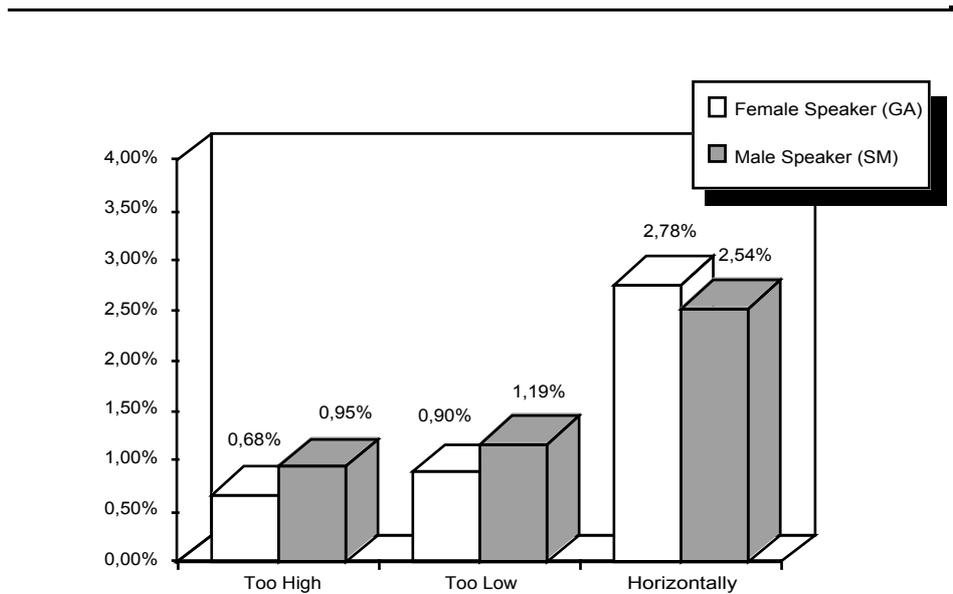


Figure 11. Percentage of moved target points in the German corpus

### 5.3.5.- General comments and conclusions

#### 5.3.5.1.- Speakers

The male speaker (SM) has a faster speaking rate than the female speaker, leading to the realization of fewer breathing or empty pauses. Considering that the major target-points detection problems are located at the beginning or the end of sentences, or at major prosodic boundaries followed or preceded by pauses, the detection of target-points is generally much better for the male speaker.

We might consider too that the lower pitch range of the male speaker can be taken into consideration for the better results of MOMEL. The F0 variations are indeed less sharp.

Those two parameters (faster speaking rate and lower pitch range) may also be correlated. As a result, it is interesting to note that the female speaker has a total of more target points (TPs) detections as the male speaker and more modifications of TPs are necessary. In the final MOMEL file (EVALUATION.CB), the number of total TPs for the female speaker is far superior to the male speaker's; this is essentially due to missing TPs.

#### 5.3.5.2.- Language specification

German is a language with lexical stress involving quick micro-variations of the F0 curve, sometimes even in the syllabic frame. MOMEL tends to 'polish' those micro-variations, maybe because it confuses them with micro-prosodic variations.

Also, German contains many unvoiced obstruents and plosives, as well as glottal stops which are sometimes difficult to detect properly for the F0 detection algorithm; hence a less efficient stylisation of F0.

Finally, it would be interesting to compare the total number of TPs detected across the different languages. Languages with lexical stress may need more TPs to account for their prosodic organisation than languages without lexical stress like French for example.

#### 5.3.5.3.- Sentence length

It seems that the TP detection is much more efficient the shorter the sentences are. Much more modifications are necessary when the sentences are long or contain pauses or major prosodic boundaries.

#### 5.3.5.4.- Missing TPs

Most of the missing of TPs are due to the realization of the lexical stress in German. It is sharp and the F0 of the unaccented syllables preceding it is often very low before the stress. It is often the case at the beginning of a sentence in the case of Article+Noun for example. MOMEL tends to interpolate directly with the peak on the lexical stress.

There are interesting cases of missing TPs in the middle of sentences but located at major prosodic boundaries followed or preceded by a breathing or an empty pause. They are more numerous for the female speaker (12 cases against 3 for the male speaker), as we explained earlier. Of course, the majority of missing TPs are located at Final rising or falling corresponding to the orthographic sentences, but a good proportion of the middle missing TPs are located at prosodic boundaries, and not just anywhere in the middle of the sentence.

As for the missing TPs located at Final risings, one could wonder if the problem of the MOMEL algorithm is the pitch range or the timing of the rise.

#### 5.3.5.5.- Extra TP

Most of the time, the extra TPs are those which are 'stuck' to one another, which is clearly a problem of the detection algorithm. It often doesn't hinder the understanding of the sentence, but extra TPs produce 'noise' (too much information for the computer).

#### 5.3.5.6.- Moved TP

The expert has often the choice between creating or moving an existing TP. For the reason cited above, it is preferable to move TPs. Interestingly, we can note that many moved TPs are located at sentence or prosodic boundaries.

#### 5.3.5.7.- Production: particular cases

In passage 34, both speakers produce 5 sentences, but they don't produce the major prosodic boundaries (final falling) at the same place. This is due to the graphic style of the sentences (many ';'). In passage 38, both speakers produces 6 sentences due to the same graphic problem.

#### 5.3.5.8.- General conclusions

Contrary to Swedish, the detection problems in German do not impair the understanding of the meaning of the sentences. There are nevertheless 2 cases per speaker in which the

modality of the sentences (question) is mistaken for another (assertion). However, the problem of MOMEL sometimes 'polishing' some lexical stresses doesn't hinder the understanding.

More important, we can conclude that the major problems of detection of TPs are located at sentence and prosodic boundaries. These could be a lot lessened if the algorithm could take pauses into consideration.

## 5.4.- Spanish

Mònica Estruch & Juan M. Garrido

Departament de Filologia Espanyola, Universitat Autònoma de Barcelona

Facultat de Filosofia i Lletres. Edifici B, 08193 Bellaterra, Barcelona, Spain

Tel : +34.3.581.19.12; Fax : +34.3.581.16.86; E-mail : {monical|juanma}@liceu.uab.es

### 5.4.1.- Introduction

The aim of this task was to make a perceptual validation of the automatic modelling of the F0 curve from the speech signal. The prosody tools developed within MULTTEXT derive automatically a symbolic representation of the intonation contour from the speech signal.

The process of analysis includes the following steps:

- automatic modelling of the F0 curve from the speech signal
- automatic stylisation of the curve with a sequence of target points (Hz, ms)
- perceptual validation of the synthesis of the paragraphs
- hand validation and correction of the errors done by the system

### 5.4.2. Database

The EUROM.1 prompting texts consist of three parts from which, in this case, only passages have been analyzed. There are 40 passages consisting each one of five task related sentences whereas the sentences are designed to compensate the uneven diphone distribution of the passages. Each passage was displayed as a single block and speakers tried to produce the block with natural intonation.

The forty passages were distributed as follows:

Speaker	Block	Number	Speaker	Block	Number	Speaker	Block	Number
RA	O0	0739	CA	P5	0493	TA	R0	0599
RA	O1	0740	CA	P6	0494	TA	R1	0600
RA	O2	0741	CA	P7	0495	TA	R2	0601
RA	O3	0742	CA	P8	0496	TA	R3	0615
RA	O4	0743	CA	P9	0497	TA	R4	0616
RA	O5	0763	CA	Q0	0498	TA	R5	0617
RA	O6	0764	CA	Q1	0499	TA	R6	0629
RA	O7	0765	CA	Q2	0500	TA	R7	0630
RA	O8	0766	CA	Q3	0501	TA	R8	0631
RA	O9	0767	CA	Q4	0502	TA	R9	0632
RA	P0	0768	CA	Q5	0503			
RA	P1	0769	CA	Q6	0504			

RA	P2	0770	CA	Q7	0505			
RA	P3	0771	CA	Q8	0506			
RA	P4	0772	CA	Q9	0507			

Table 11: Distribution of passages and speakers in the Spanish corpus

### 5.4.3. Speakers

Three speakers have been analyzed, the two Very Few speakers and the third one selected among the Few Speakers set. A code has been assigned to each speaker based on age, sex and talker group:

Sex	Age Group	Few Speaker	Very Few Speaker
Male	30-40		RA
Female	20-30		CA
Male	40-50	TA	

Table 12: Speakers selected for the Spanish corpus

### 5.4.4.- Results

As a result of this analysis, some problems have been found and classified in three different categories: cases in which target points were missed, cases in which there were extra target points and finally, cases in which the existent target points had to be moved. The results have also been classified according to the three speakers which have been analysed.

#### 5.4.4.1.- Missing target points

The first category of problems found are cases in which there should be target points in places where MOMEL has not put them.

Three different positions have been distinguished: initial (for the beginning of sentences), final (for the end of sentences, distinguishing between cases with a rising end from the ones with a falling one) and middle (for points in the middle of sentences).

Table 13 summarizes the results:

	Initial	Middle	Final	
			Rising	Falling
Speaker CA	6/73 (8,22%)	10/817 (1,22%)	4/4 (100%)	2/69 (2,89%)
Speaker RA	15/76 (19,73%)	21/822 (2,55%)	11/14 (78,57%)	7/62 (11,29%)
Speaker TA	9/49 (18,36%)	13/487 (2,67%)	2/5 (40%)	6/44 (13,63%)

Table 13: Missing target points (in absolute numbers and in percentages) in the Spanish corpus

Let us give an example of each case:

- initial position: RA O6 0764 (*Por favor...*)
- middle position: TA R6 0629 (*de la calle*)

- final 'rising' position: TA R2 0601 (*por teléfono*)
- final 'falling' position: CA Q5 0503 (*otros animales*)

These cases can also be seen in figure 13:

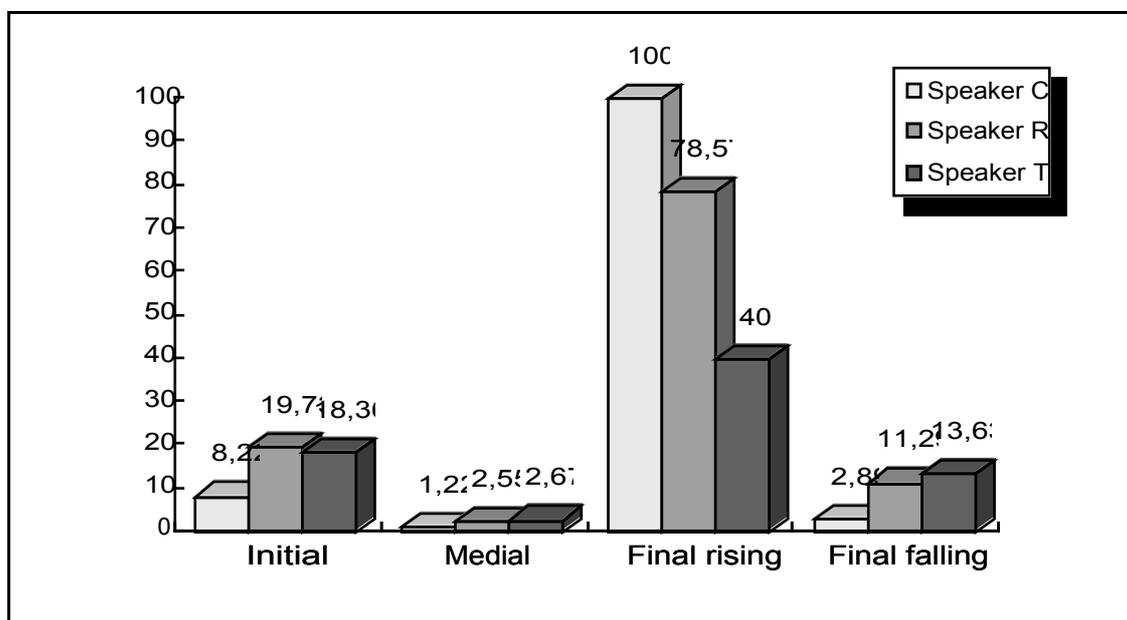


Figure 13: Percentage of missing target points in the Spanish corpus

#### 5.4.4.2.- Extra target points

This second category is much more reduced than the first one. It includes cases in which the system has recognised too many target points, that is to say, target points in places where they should not appear.

The number of errors found is shown in the next table:

<b>Speaker CA</b>	2/963 (0,20%)
<b>Speaker RA</b>	2/974 (0,20%)
<b>Speaker TA</b>	0/585 (0%)

Table 14: Extra target points  
(in absolute numbers and in percentages)  
in the Spanish corpus

An example:

CA P7 0495 (*de Griñón*)

The same results are shown in figure 14.:

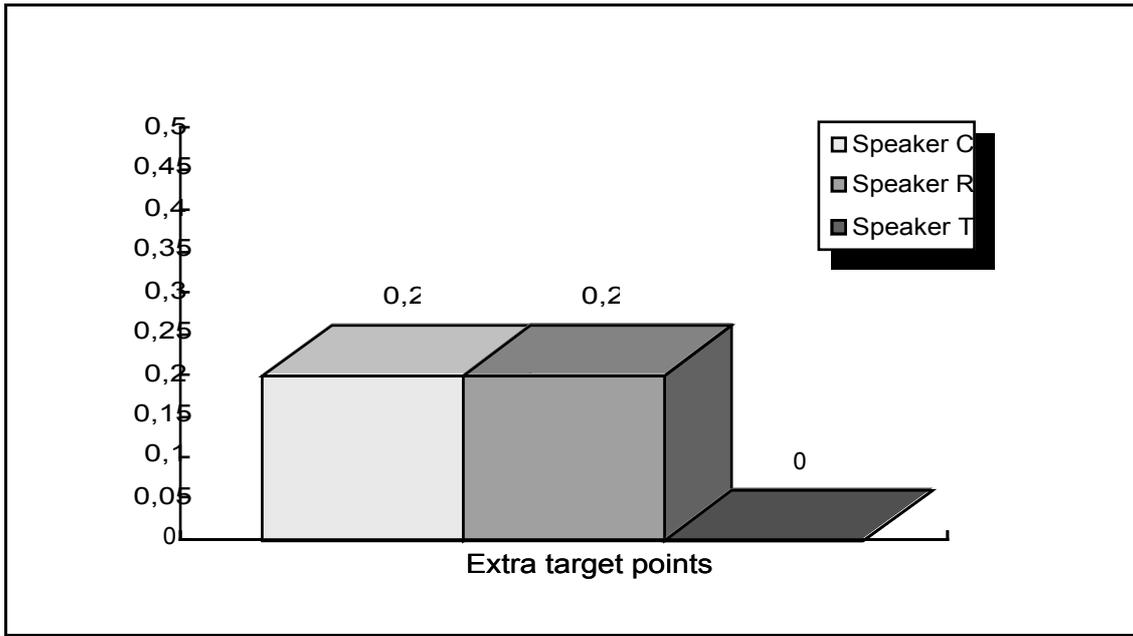


Figure 14: Percentage of extra target points in the Spanish corpus

#### 5.4.4.3.- Moved target points

Finally, another group of errors includes the cases in which the target points were detected by MOMEL but in an incorrect position. Therefore, these target points have been moved. The classification has been done according to the situation of the target point: whether it was too high or too low, or whether it had to be moved horizontally in the time domain.

	Too High	Too Low	Horizontally
Speaker CA	7/963 (0,72%)	1/974 (0,10%)	1/585 (0,17%)
Speaker RA	2/963 (0,21%)	0/974 (0%)	1/585 (0,17%)
Speaker TA	1/963 (0,10%)	0/974 (0%)	1/585 (0,17%)

Table 15: Moved target points (in absolute numbers and in percentages) in the Spanish corpus

Examples of these cases are the following:

- too high: CA Q2 0500 (*fantástico*)
- too low: CA Q1 0499 (*soy*)
- horizontally: CA Q0 0498 (*Tenía*)

These results can also be observed in figure 15:

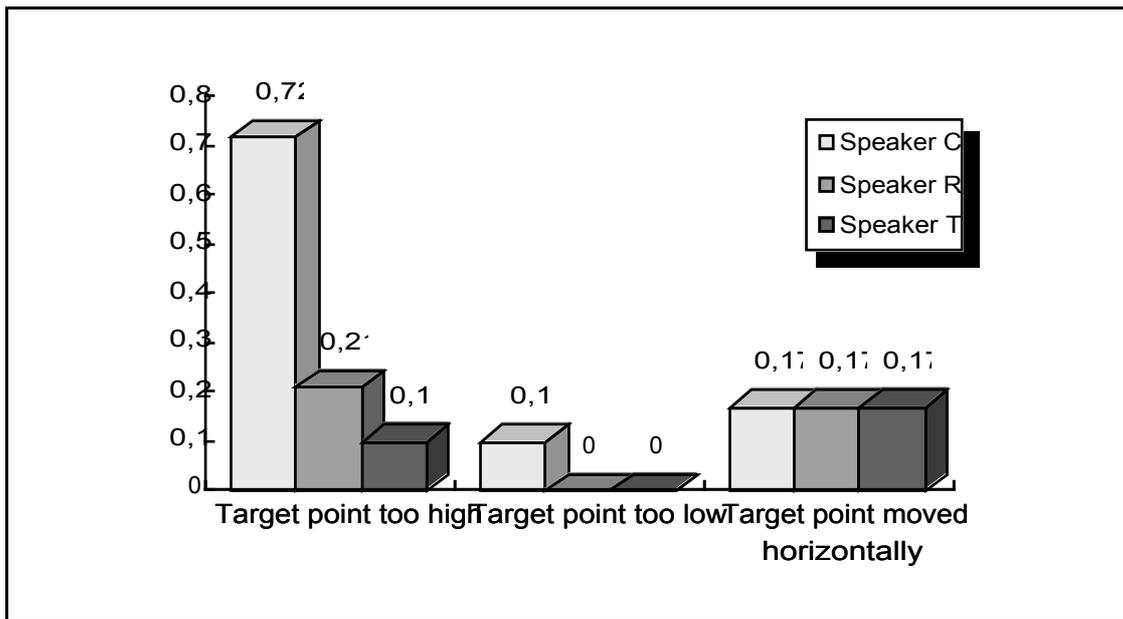


Figure 15: Percentage of extra target points in the Spanish corpus

#### 5.4.5.- General evaluation

In general, it has to be remarked that the automatic stylisation done by MOMEL produces good results.

Nevertheless, there are some errors, specially cases of missing target points. A consideration of possible causes has led us to think that this may have a solution if the system would recognise pauses. In that case, the F0 curve would be interrupted at least at the end of each sentence and would also have a new start at the beginning of the sentences. Most of the errors of misdetection of target points at the beginning and the end of sentences would probably disappear.

#### 5.4.6.- Alignment

There was another task consisting of the alignment of word boundaries on the one side, and the stressed syllables on the other, with the speech signal. The alignment has been performed for one speaker (CA), and for the 15 paragraphs recorded by this speaker.

### 5.5.- Swedish

Eva Strangert & Anna Aasa

Department of Linguistics, Umeå University

S-901 87 Umeå, Sweden

Tel +46 (0)90 16 56 80; Fax +46 (0)90 16 63 77; E-mail: strangert@ling.umu.se

Report published in TMH-QPSR 2/1996 (*Speech, Music and Hearing - Quarterly Progress and Status Report*), KTH, Stockholm, Sweden, pp. 37-40. (Papers presented at the Swedish Phonetics Conference held in May 1996).

## Abstract

Based on speech data collected within the SAM project, tools for the analysis and coding of F0 used within the MULTEXT project were evaluated using 8 Swedish speakers. Generally, comparison of original and modelled versions of the same utterances using INTSINT coding revealed fairly successful matching. The three main types of mismatches that occurred were: stylistic changes, changes of prominence relations, and shifts of word accent.

### 5.5.1.- Introduction

MULTEXT (*MULTilingual TEXT Tools and corpora*) is a European project in the area of Linguistic Research and Engineering. It started in January 1994 with Jean Véronis (Laboratoire Parole et Langage, CNRS & Université de Provence) as coordinator and a number of participants representing six languages - English, German Dutch, French, Spanish and Italian. In 1994, Eva Ejerhed obtained funding from NUTEK for work on the Swedish language to be carried out at the Department of Linguistics, Umeå, as an Associated Partner in the MULTEXT project.

The MULTEXT project consists of three main areas: standardization, development of corpus-handling tools and multilingual corpora (written and spoken), and industrial validation.

Work in the MULTEXT-SW project is going on in the second of these areas to produce: 1) a lexicon of the 10 000 most frequent word forms in the MULTEXT-SW text corpus, with their analyses; 2) a Swedish text corpus of 2 M words of financial newspaper text from 1993, annotated for part of speech by a fully automatic, probabilistic tagger; and 3) a Swedish speech corpus, collected by KTH, Stockholm, as part of the European project SAM (*Speech Assessment and Methodology*), automatically analyzed and semiautomatically annotated for prosody in Umeå, using the prosody tool INTSINT (INternational Transcription System for intonation) developed at Aix-en-Provence (Hirst and Espesser, 1993; Hirst and Di Cristo, 1996).

The aim of this paper is to contribute to the evaluation of the prosody tool, based on its application to Swedish data. Evaluations like the Swedish one have also been undertaken for English, French, Spanish and Polish.

### 5.5.2.- Prosody tools

The evaluation concerns software tools for the analysis and coding of F0 in continuous text. The analysis is handled by an automatic F0 modelling program, MOMEL, the output of which is a sequence of target points (ms; Hz). These can be used to generate a quadratic spline function which in turn can be used as input for PSOLA resynthesis (see Hirst, Nicolas and Espesser, 1991, and further references there). F0 targets points are then modelled by using the INTSINT system that has been developed in an attempt to set up a pitch transcription system which can be utilized for any language (Hirst and Di Cristo, 1996). This coding includes the absolute symbols T (Top), M (Mid) and B (Bottom) and the relative symbols H (Higher), L (Lower), S (Same), U (Upstepped) and D (Downstepped).

The INTSINT coding is performed automatically using an algorithm which optimizes the coding so that the absolute tones are modelled by the mean value of their category

and the relative tones are modelled using linear regression as a linear function of the preceding target point. The symbolic coding together with the means or regression coefficients makes it possible to generate a new set of target points from the symbolic coding which can then in turn be used to generate an F0 curve that can be input into PSOLA resynthesis. The modelling is made on stretches of speech that have been segmented into single intonation units.

### **5.5.3.- The Swedish corpus**

The evaluation was based on recordings of speech data carried out at KTH within the framework of the SAM project. Of the 70 speakers recorded, we chose 4 women and 4 men from the 'Few Speaker Set' (containing subjects used to reading diverse types of material under controlled conditions). They were all from the Stockholm area and spoke Standard Middle Swedish. Each speaker read 5 passages, each made up of 5 task-related sentences. Thus the corpus evaluated included 200 sentences (8 subjects x 5 passages x 5 sentences).

### **5.5.4.- Evaluation**

The match between the synthesized and the original version of the sentences was judged by listening to the material through headphones. A single Swedish subject, experienced in phonetics, judged all the material.

The evaluation was made in two steps. First, deviations from perfect matching with the original were noted for the synthesized versions of all sentences. Here the aim was to listen for deviations of any kind, whether minor or major. Second, the synthesis generated from the symbolic INTSINT coding and the statistical coefficients for each tone were checked to identify major deviations, that is, those cases in which the modelled version brought about a change of meaning compared to the original. In this paper we will restrict ourselves to reporting on this second step of the evaluation.

### **5.5.5.- Results**

Generally, the INTSINT modelling appears to be fairly successful as applied to the Swedish data. However, mismatches involving shifts of meaning between the synthesized and the original versions of sentences occurred with all speakers, although for some there were fewer errors than for others. There were no apparent differences between the male and female groups, although the speaker with the best fit with the modelled version was a male. This good fit, however, might be a consequence of the rather small range of F0 observed for this speaker.

We have observed three main types of errors: stylistic changes, changes of prominence relations and shifts of word accent.

#### **5.5.6.1.- Stylistic changes**

The stylistic changes are emotive in character. They involve a change in mood or attitude, mostly in the direction of a more neutral or even depressed style of speaking than in the original. There are also utterances which sound ironic, demanding and surprised rather than neutral as in the original. One example fragment of a modelled sentence conveying a feeling of sadness is shown in Figure 16 together with the more

neutral-sounding original. The figure also contains some corrections to the modelled version that were made in order to achieve a better match with the original.

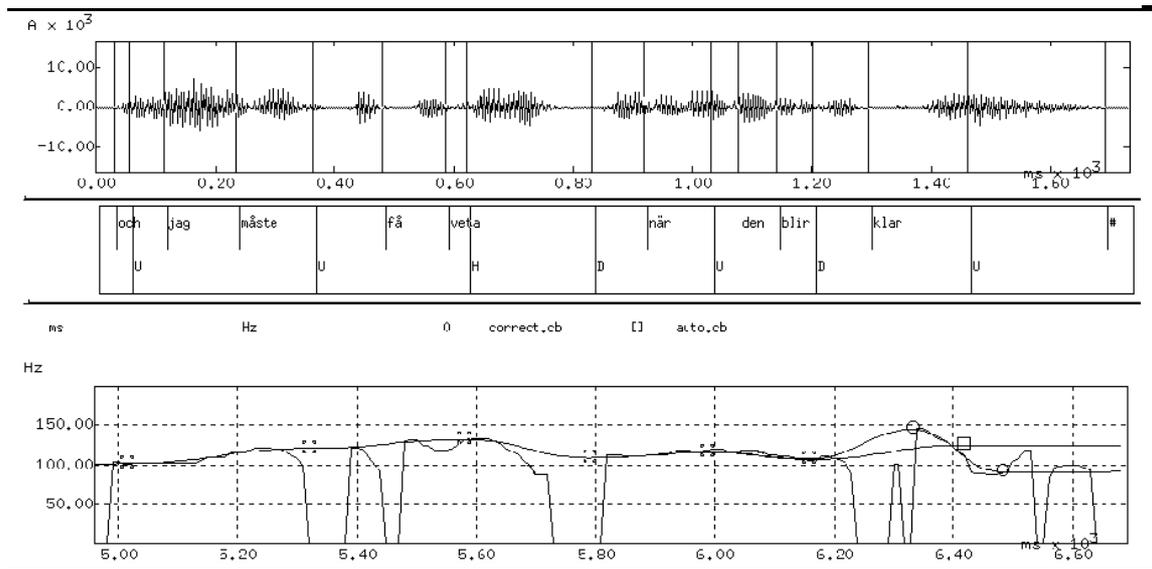


Figure 16: Example of a change in speech mood in the sentence fragment  
 '...och jag måste få veta när den blir klar.' - '...and I have to know when it will be ready.'  
 Original: neutral; Modelled (—): sad; Corrected (---): neutral.

The acoustic basis for these mismatches very often seems to be F0 deviations either at the beginning or the end of sentences. The most frequent patterns appear to be a missing F0 rise at the beginning and/or a missing fall at the end. In the example in Figure 1 the synthesized version missed a H close to the end and the final F0 lowering.

### 5.5.6.2.- Changes of prominence relations

The far most frequently occurring mismatches involve prominences. Usually prominences are weakened, exhibiting a degrading along the prominence scale with originally accented syllables being deaccented and focussed syllables transformed into accented or even deaccented syllables. (For Swedish four prominence levels are usually assumed: unstressed, stressed (deaccented), accented or focus accented, see e.g. Bruce 1990.)

In Swedish F0 is assumed to be the main acoustic correlate apart from duration for perceiving prominence, and perceived focus accent is assumed to be tied to an F0 rise occurring after a word accent fall (Bruce, 1977). In acute words (see below) this fall may sometimes be missing. This is the case in the example shown in Figure 17 in which the acute, focus-accented, second word in the question *Hur går pjäsen?* 'How is the play doing?' not only lacks the fall but also the focus accent rise (and the connected H) conveying the impression of a non-focussed word. Conversely, we have observed cases where a perceived focus does not co-occur with an F0 rise, which indicates a more complex acoustic basis for perceived focus than just F0.

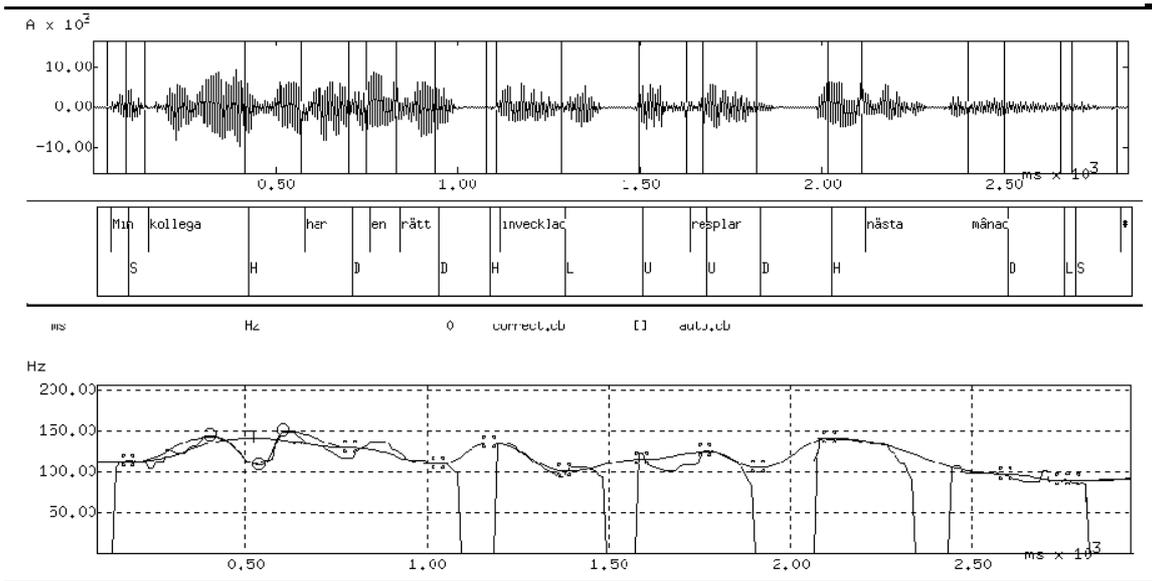


Figure 17: Example of a change of prominence relation in the sentence 'Hur går pjäsen?' - 'How is the play doing?' Original: strong prominence on 'går'; Modelled (□): prominence missing; Corrected (○): prominence restored .

### 5.5.6.3.- Shifts of word accent

There is a small number of word accent shifts, mostly in the direction grave > acute. For example, *kollega*, 'colleague' with grave accent (and focus) is rendered as acute in the modelled version in the example in Figure 18.

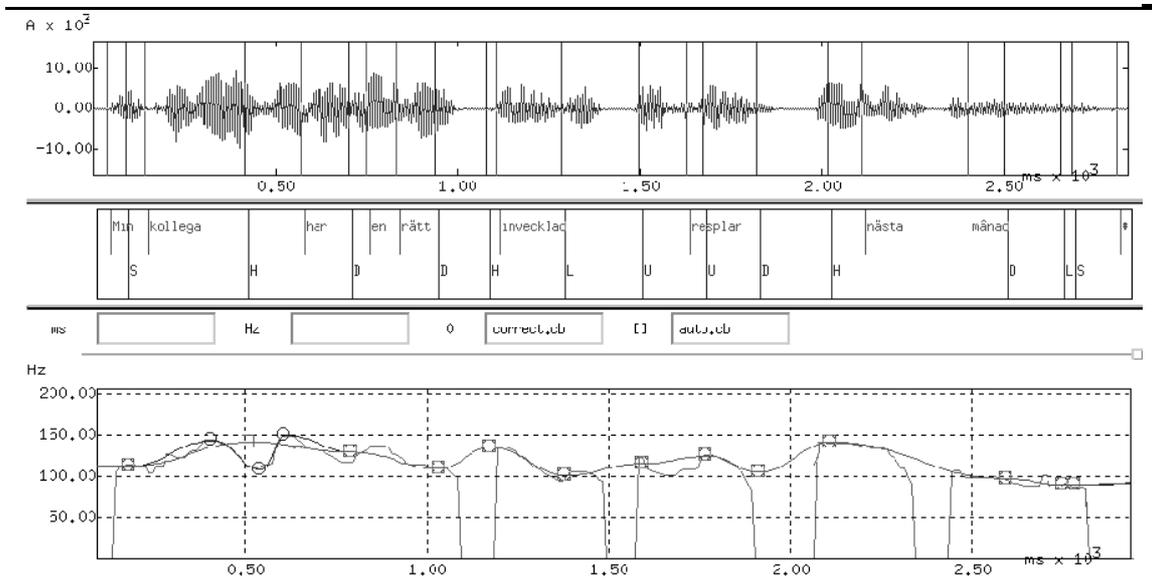


Figure 18: Example of a shift of word accent of the grave word 'kollega' - 'colleague'. Original: grave; Modelled (□): acute; Corrected (○): grave.

The two word accents in Swedish, acute and grave, have to be kept apart, as they sometimes distinguish between otherwise identical words. The distinction rests on the timing of an F0 fall related to the stressed syllable, the fall being later in grave than in acute

words (Bruce, 1977). The word accent fall may also be missing altogether in acute words.

We assume this to be the reason for the impression of the second word, *kollega* as acute in Figure 2, as there is no word accent fall on this word in the modelled version.

Some cases involving a shift of word accent also give the impression of a dialectal change, e.g. a Dalecarlian or Northern Swedish style instead of the original Stockholm Swedish. This might be explained by the various manifestations of the word accents over the different parts of Sweden.

### **5.5.6.- Conclusions**

The evaluation undertaken has shown the INTSINT modelling of intonation to be fairly successful as applied to Swedish. Based on judgements of 200 sentences spoken by 8 Swedish men and women, we can conclude that the majority of errors detected when comparing the modelled and the original sentences are of three types: changes of mood or attitude of speech, changes of prominence relations, and interchange between the two word accents in Swedish. The acoustic basis for most of these errors seems to be a too heavy smoothing of the intonation contour. At the beginning of a sentence, F<sub>0</sub> often starts too high and at the end there may be a missing fall. Finally, the impression that there is a degrading of prominence results from missing peaks in the contour.

### **5.5.7.- Acknowledgements**

Daniel Hirst made suggestions for improvements to an earlier version of this paper. Also, Daniel Hirst and Robert Espesser at CNRS & Université de Provence provided us with a thorough introduction to the project and to technical details. The Department of Speech, Music and Hearing, KTH, Stockholm, supplied the (SAM) material. Ola Andersson served as technical assistant at the Phonetics Laboratory in Umeå. We gratefully acknowledge the contributions of these people.

### **5.5.8.- References**

- Bruce G (1977). Swedish word accents in sentence perspective. *Travaux de l'Institut de Linguistique de Lund*. Lund: CWK Gleerup.
- Bruce G (1990). On the analysis of prosody in spontaneous dialogue. *Working Papers* 36: 37-55. Department of Linguistics, Lund University.
- Hirst D J and Espesser R (1993). Automatic modelling of fundamental frequency using a quadratic spline function. *Travaux de l'Institut de Phonétique d'Aix* 15: 71-85.
- Hirst D J, Nicolas P and Espesser R (1991). Coding the F<sub>0</sub> of a continuous text in French: an experimental approach. In: *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Aix-en-Provence, 1991, 5: 234-237.
- Hirst D J and Di Cristo A (1996). A survey of intonation systems. In: *Intonation systems: a survey of twenty languages*. Cambridge: Cambridge University Press.

## 6.- Conclusions

In order to obtain an overview of the performance of MOMEL, the results for the languages for which quantitative data is available can be summarized according to the three categories of problems that have been identified during the validation process: missing target points, extra target points and target points which had to be moved. Figure 19 presents the percentage of target points corresponding to each of the three categories averaged according to different speakers and positions for French, German and Spanish.

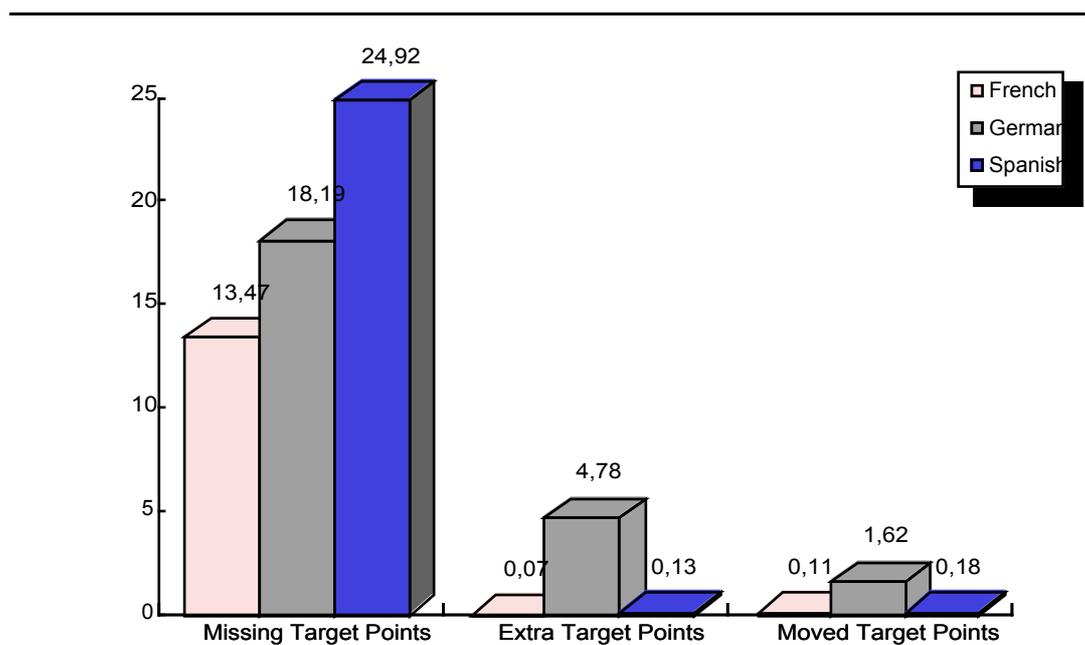


Figure 19: Percentage of missing, extra and moved target points averaged across speakers and positions for French, German and Spanish

The highest number of problems is found in missing target points that had to be manually added during the validation process (18.8% averaged across languages, positions and speakers). Most of these target points are located in initial and final positions specially in rising contours. Problems in this position are also detected for English and Swedish, and are in most cases related to the appearance of pauses at the beginning or at the end of the sentences composing the analyzed passages.

The percentage of target points which had to be manually corrected either adding new ones or moving existing ones is very low, showing a good performance of MOMEL when pauses are not involved.

It is noted for all languages that taking pauses into account would improve the performance of MOMEL, since most of the problems are detected at sentence boundaries.

## 7.- References

- 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", in *Eurospeech'95. Proceedings of the 4th European Conference on Speech Communication and Speech Technology*. Madrid, Spain, 18-21 September, 1995. Vol 1, pp. 867-870.
- HIRST, D.- ESPESSER, R. (1993) "Automatic modelling of fundamental frequency using a quadratic spline function", *Travaux de l'Institut de Phonétique d'Aix 15*: 71-85
- HIRST, D.J.- DI CRISTO, A. (in press) " A survey of intonation systems", in HIRST, D.J.- DI CRISTO, A. (Eds.) *Intonation Systems: A Survey of Twenty Languages*. Cambridge: Cambridge University Press.
- Prosody Encoding Survey*. WP 1. Specifications and Standards. T1.5 Markup Specifications. Deliverable 1.5.3. Final version. Available at: <http://www.lpl.univ-aix.fr/projects/multext/CES/CES2.html>
- Report on Task 2.6 Prosody Tools and Task 2.7. Post-editing tools*. Speech Tools. WP2 Corpus Annotation Tools. Milestone A2 Deliverables. Available at: <http://www.lpl.univ-aix.fr/projects/multext/MUL7.html>
- SHERWOOD, T.- FULLER, H. (1992) *Guide to EUROM.1 Speech Database*. Doc. No. SAM-NPL-102, Final, 21 April 1992. ESPRIT PROJECT 2589 (SAM)