Removing micromelody from fundamental frequency contours

Uwe D. Reichel, Raphael Winkelmann

Institute of Phonetics and Speech Processing
University of Munich, Germany
{reichelu|raphael}@phonetik.uni-muenchen.de

Abstract

In this paper we describe a new method to diminish microprosodic components of fundamental frequency contours by applying weight functions linked to microprosodically classified phone combinations. For vowel segments in obstruent environments our algorithm outperforms standard smoothing algorithms like Moving-Average filtering, Savitzky-Golay filtering or MOMEL in diminishing F0 variations related to microprosodic factors while retaining significant differences related to macroprosody.

Index Terms: microprosody, smoothing, intonation

1. Introduction

Smoothing is a crucial preprocessing step for preparing F0 contours for intonation research. It serves to weaken F0 measurement errors and to at least partially remove the influence of micromelody which consists in F0 perturbations caused by the segmental phonetic level.

1.1. Micromelody

Micromelody is widely considered to be non-intentional and universal (while showing a high degree of variability among speakers) [1, 2]. It is primarily related to the segmental phonetic level and not to higher macroprosodic events such as accentuation and prosodic phrasing. It can be divided into *intrinsic* and *co-intrinsic pitch* (*IF0* and *CF0*).

Intrinsic Pitch IF0 is related to phonetic segment categories. For example a positive correlation between vowel height and IF0 is extensively documented (e.g. [3]), which has been found to be most prominent within the vowel centers [1]. No uniform findings are reported for tense vs. lax vowels [4].

While these general tendencies can be attributed to the segmental level, the amount of perturbation is strongly speaker dependent and furthermore related to macroprosodic factors such as register and phrase-level accent. For example greater segment-induced IF0 effects are reported in accented and utterance initial syllables [5], and generally in higher F0 registers [6]. These findings complicate the task of separating micro- and macrointonation.

Co-intrinsic pitch CF0 is related to segment transitions, especially to CV sequences. It was found that F0 in a vowel rises in the vicinity of voiceless obstruents as opposed to voiced ones (e.g. [3]), and that fricatives are more influential than stops [7]. Often place of articulation turned out not to be influential [8]. While in [8] co-intrinsic pitch effects are assumed not to span the whole vowel segment but just its peripheral parts, other studies as [9] report CF0 influence of a neighbouring obstruent on the whole vowel.

Generally micromelodic effects are more prominent in isolated words controlled for macroprosody than in prosodically uncontrolled connected speech [10].

1.2. Smoothing procedures

Among the most popular smoothing procedures to remove F0 disturbances not related to macrointonation are *Moving-Average* and *Savitzky-Golay* filtering [11] as well as MOMEL [12].

Moving-Average Moving-Average filtering replaces each F0 value y_t by the arithmetic mean within a time window of length $2n + 1$ centered on t: $y_t = \text{mean}(y_{t-n \ldots t+n})$. The higher n the smoother the resulting contour.

Savitzky-Golay In Savitzky-Golay filtering each y_t is replaced by a value derived from polynomial fitting: $y_t = \text{polyfit}(y_{t-n \ldots t+n})_{n+1}$. The lower the chosen polynomial order, the smoother the resulting contour. In general Savitzky-Golay filtering is more appropriate to preserve the original contour extrema than Moving-Average.

MOMEL While these smoothing methods are general and not initially developed to address intonation research problems, MOMEL (*MOdelisation MELodique*) was designed specially for this purpose: each F0 segment in an *analysis window* is iteratively approximated by a parabola p. At each iteration step original F0 values with a distance from the fitted parabola exceeding a chosen threshold Δ are removed. This iteration terminates as soon as no F0 value differs from p by more than Δ. The extrema of the parabolas derived this way form target candidates which are further reduced within *reduction windows* dependent on their deviance from local mean values. The remaining targets finally serve as nodes for a quadratic spline function for F0 smoothing.

1.3. Goal of this paper

All these smoothing methods have the advantage not to depend on any prior phonetic segmentation. On the other hand, none of them explicitely addresses the issue of segment relatedness of micromelody since the whole signal is processed uniformly. Therefore none of these methods can innately guarantee (a) to remove micromelody, and (b) not to affect macrointonation. Our goals have therefore been to directly face the issue of micromelody and to test explicitely for our method and the ones described in the previous section to what extent they fulfil the formulated criteria. In this first attempt we restrict ourselves to vowel segments in obstruent environments.

2. Data

The used data consists of parts of the Kiel Corpus [13] containing about 6.5 hours of spontaneous spoken dialogues of 128

speakers. The data is hand-segmented and prosodically annotated within the Kiel intonation model framework. F0 was extracted with a sample rate of 200 Hz using the Schaefer-Vincent algorithm [14]. No manual F0 correction was carried out.

MOMEL smoothing was done with Praat 5.0.29 by means of a freely available script provided by [15]. Δ was set to 5%, and the lengths of the analysis and reduction windows to 300 and 200 ms respectively, as suggested by [12]. Severe F0 breakouts caused by cubic spline interpolation were bridged by linear interpolation. For the other two smoothing procedures a window size of 25 ms (5 samples) was chosen. Savitzky-Golay filtering was carried out by third order polynomials.

3. Removing micromelody

The *WAM* model developed in this study (*W*eights *A*gainst *M*icromelody) treats the task of micromelody removal as a multiplication of a vowel segment's F0 contour by factors derived from phoneme sequence dependent weight functions.

3.1. Segment classification

Initially, vowel segments are classified with respect to the following three microprosodically relevant factors:

- **HGT:** Vowel height (*high vs. mid vs. low*)
- **VOI1:** Voicing of the preceeding obstruent (*voiced vs. unvoiced*)
- **VOI2:** Voicing of the following obstruent (*voiced vs. unvoiced*)

Given 3x2x2 factor steps, this categorisation yields 12 vowel classes.

3.2. Weight function

Preprocessing In order to abstract from a segment's length and to yield F0 contours y of uniform length, which is required by the subsequent operations, the following preprocessing steps are carried out: In each vowel segment the F0 contour is time-normalised to the interval $[-1\ 1]$ and polynomially approximated. The order of the polynomial is adjusted dynamically to *contour length minus one* to entirely conserve the contour's shape. Within the time normalised interval the contour is then mapped on a representation consisting of 10 time-equidistant samples which are derived by the fitted polynomial.

Base F0 removal For each speaker an F0 base value y_b is calculated by taking the median of all his F0 values less or equal the 2nd percentile – a procedure we consider to be robust against outliers. This speaker-dependent y_b is then subtracted from the F0 contours.

Function development Since micromelodic effects vary strongly among speakers [1], weight functions are derived separately for each speaker. For each vowel class i a set Y_i of F0 contours is given, which were preprocessed and separated from the base F0 as described above. From Y_i the centroid F0 sequence m_i is calculated by taking the arithmetic mean of the values below the 98th percentile (again to achieve robustness against outliers). A reference centroid r is computed the same way from F0 values Y of all vowel segments of the speaker.

Class-related weight sequences w_i are then derived by pointwise division of r by the corresponding centroid m_i. Finally the class-specific weight function $\mathrm{wf}_i(t)$ is derived by fitting a third-order polynomial to map normalised time t to w_i.

reference $r := \mathrm{median}(Y)$
foreach vowel class i

centroid $m_i := \mathrm{median}(Y_i)$
weight sequence $w_i := \frac{r}{m_i}$
weight function $\mathrm{wf}_i(t) := \mathrm{polyfit}(t, w_i)$

end

3.3. Application

In application each vowel segment first has to be classified with respect to the factors *HGT*, *VOI1*, and *VOI2* to choose the appropriate weight function $\mathrm{wf}_i(t)$ for F0 modification. After determination of the F0 base value y_b as explained above, the time-normalised F0 contour y is then adjusted by subtracting y_b, pointwise multiplication of the residual by the weights derived from the appropriate weight function $\mathrm{wf}_i(t)$, and adding y_b to the product:

$$y_{\text{smoothed}} \quad = \quad y_b + (y - y_b) \cdot \mathrm{wf}_i(t)$$

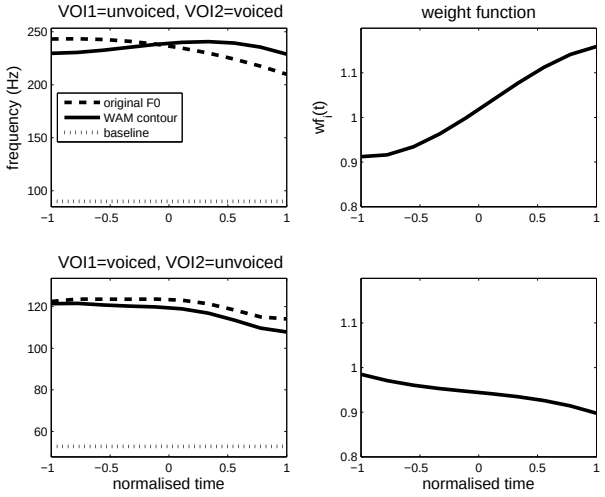

Figure 1: **Left:** *Micromelody removal for F0 contours of mid vowels in two different obstruent contexts (***top:** *unvoiced-voiced,* **bottom:** *voiced-unvoiced) by multiplication with time-dependent weight factors derived from the weight functions* $wf_i(t)$ *shown on the* **right**.

4. Evaluation

4.1. Method

In order to evaluate the smoothing methods with respect to their capability of removing micromelody from F0, we tested the influence of the microprosodic factors *HGT, VOI1*, and *VOI2* on the original and the smoothed F0 contours. Furthermore we tested the effect of a macroprosodic factor *ACC* derived from the prosodic corpus annotation: *MP (middle peak; higher F0 values)* vs. *MV (middle valley; lower F0 values)*.

Appropriate smoothing methods should:

- remove the influence of the microprosodic factors, and
- retain the influence of the macroprosodic factor.

An analysis of variance with following dependent variables was carried out:

- the F0 mean value calculated over the whole vowel segment, and
- the mean values for three slices of equal size in the beginning, middle, and end of the vowel segment, overlapping by the factor 0.2.

The latter variables serve to examine the time course of micromelody. The independent variables are given by *HGT, VOI1, VOI2*, and *ACC*.

F0 was normalised to $[0\ 1]$ for each speaker with respect to his F0 range (0th–98th percentile) in order to cancel out speaker-dependent variation.

4.2. Results

Figures 2, 3, 4, and 5 show the mean normalised F0 contours related to tongue height, voicing of the preceeding and following consonant, and to the accent type, respectively.

Micromelody As one can see, Savitzky-Golay and Moving-Average filters are not capable of removing the micromelody since the corresponding mean contours show only minor differences from the originals. MOMEL as well as WAM clearly show the tendency of micromelody neutralisation except of the MOMEL treatment of following consonant voicing (see Figure 4). These observations are confirmed by the results of the analysis of variance to test the significance of differences in the mean values which are shown in Table 1.

In the original contours whole segment F0 differences related to tongue height are significant for all level pairs (Tukey-Kramer post hoc). The same holds for the vowel center and the offset region. Near the onset high and mid vowels do not differ significantly with respect to mean F0.

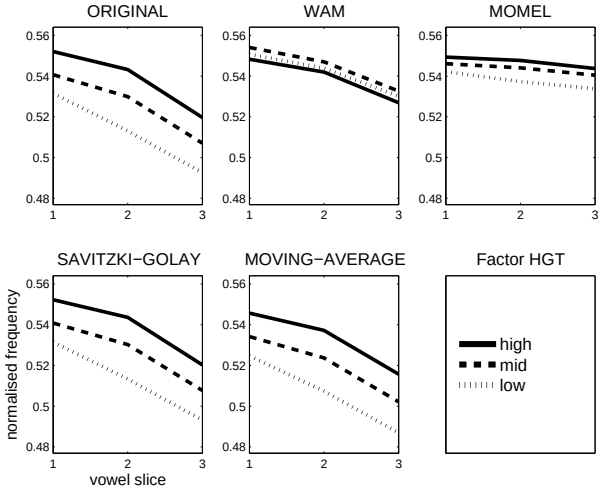

Figure 2: *Vowel segment means related to different vowel heights (factor HGT).*

Macrointonation As can be seen in Figure 5 all examined smoothing procedures are capable to conserve F0 differences related to macroprosody, although differences are a bit reduced in the MOMEL output. Nevertheless, the factor *ACC* clearly keeps its influence on F0 mean differences for all smoothing methods (Anova, $\alpha = 0.001$).

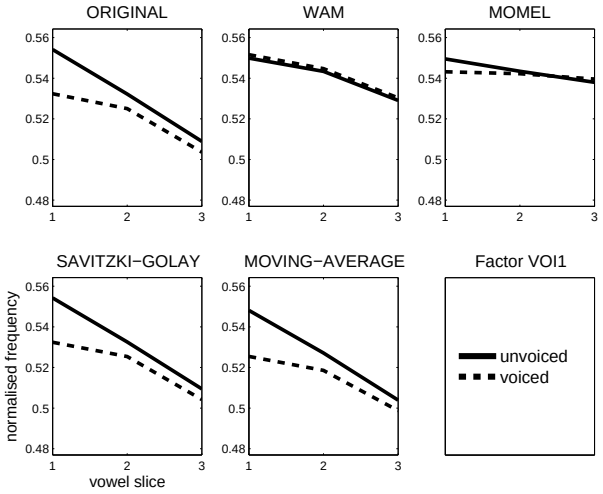

Figure 3: *Vowel segment means related to voicing of the preceeding obstruent (factor VOI1).*

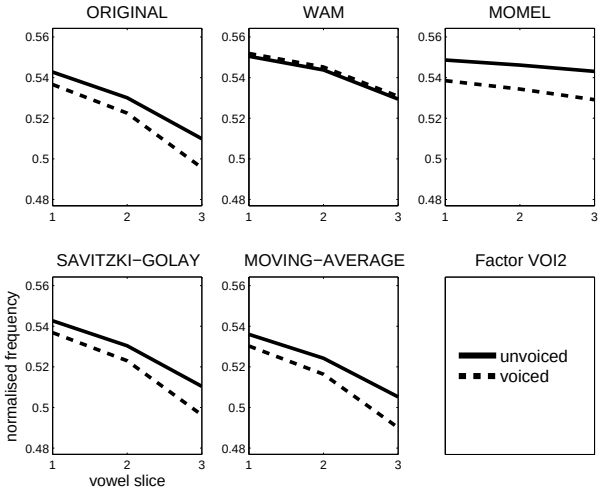

Figure 4: *Vowel segment means related to voicing of the following obstruent (factor VOI2).*

5. Discussion and Conclusions

5.1. General micromelodic trends

As can be seen in the original F0 mean values measured for three consecutive slices in the vowel segments in Figures 2, 3, and 4, the general trends already discovered in former studies mentioned in the Introduction are confirmed here: F0 is positively correlated with vowel height, and voiceless obstruents cause raised F0 values as opposed to voiced ones. The influence of the neighbouring obstruents is visible throughout the entire vowel segment, but in this study only sustainedly significant for pre-vocalic obstruents.

5.2. Comparative evaluation

While Moving-Average and Savitzky-Golay smoothing do not contribute to the removal of micromelody, WAM and MOMEL are both capable of reducing the micromelodic influence, WAM in all contexts, MOMEL concerning vowel height and voicing

Table 1: *Significance levels for global and slice-related F0 mean differences for each smoothing method for the factors HGT, VOI1, and VOI2.*

HGT	global	slice 1	slice 2	slice 3
original	0.001	0.001	0.001	0.01
WAM	*n.s.*	*n.s.*	*n.s.*	*n.s.*
MOMEL	n.s.	n.s.	n.s.	n.s.
Savitzky-Golay	0.001	0.001	0.001	0.001
Moving-Average	0.001	0.001	0.001	0.001
VOI1				
original	0.001	0.001	0.01	0.05
WAM	*n.s.*	*n.s.*	*n.s.*	*n.s.*
MOMEL	n.s.	0.05	n.s.	n.s.
Savitzky-Golay	0.001	0.001	0.001	0.01
Moving-Average	0.001	0.001	0.001	0.001
VOI2				
original	0.05	n.s.	n.s.	0.01
WAM	*n.s.*	*n.s.*	*n.s.*	*n.s.*
MOMEL	0.05	n.s.	0.05	0.01
Savitzky-Golay	0.05	n.s.	n.s.	0.05
Moving-Average	0.05	n.s.	n.s.	0.001

of the pre-vocalic obstruent. MOMEL has the advantage of not requiring any preceding phone sequence classification as is the case for WAM. However, since segmental influence is not explicitely covered, MOMEL reduces micromelodic perturbations to a lesser extent than WAM and is not capable of treating post-vocalic obstruents appropriately.

5.3. Composition of micro- and macromelody

In accordance with [1] we have chosen a multiplicative composition of micro- and macromelody which of course is not per se obligatory. Nevertheless, justification for multiplication is provided by the finding that microprosodic effects tend to be more pronounced in higher F0 registers [6]. Likewise relations between micro- and macroprosody cannot be accounted for by an additive composition.

5.4. Future Research

So far our research has been concentrated on obstruent-vowel-sequences which are microprosodically the most extensively examined and for which prominent micromelodic effects have been reported. Future research will include the application of our method to further phone combinations. Since the number of needed parameters rises exponentially with the number of possible combinations, a research focus has to be put on micromelodically equivalent behaviour of phone sequence types in order to keep the number of microprosodically motivated classes as low as possible.

Furthermore, a comparative evaluation with more current smoothing methods like *PfitzingerSmooth* [16] is to be carried out.

6. Acknowledgements

This research was funded by the DFG (STRETTS project, HA 3512/5–2).

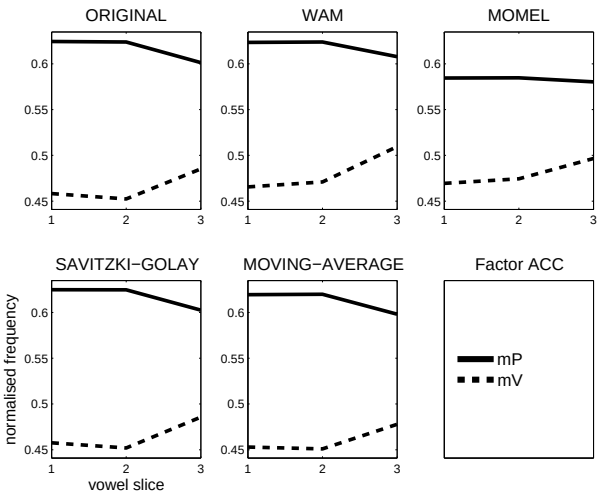

Figure 5: *Vowel segment means related to macroprosodic events Middle Peak vs. Middle Valley.*

7. References

[1] A. Di Cristo and D. Hirst, "Modelling french micromelody: Analysis and synthesis," *Phonetica*, vol. 43, pp. 11–30, 1986.

[2] D. Whalen and A. Levitt, "The universality of intrinsic F0 of vowels," *J. Phonetics*, vol. 23, pp. 349–366, 1995.

[3] I. Lehiste, "Some basic considerations in the analysis of intonation," *J. Acoust. Soc. Am.*, vol. 33, no. 4, pp. 419–425, 1961.

[4] D. Pape and C. Mooshammer, "Intrinsic pitch in German – Examining the whole fundamental frequency contour of the vowel," in *Proc. CFA/DAGA*, Strasbourg, 2004, pp. 897–898.

[5] R. Petersen, "Variation in inherent f0 level differences between vowels as a function of position in utterance and the stress group," in *Annual Report of the Inst. of Phonetics, University Copenhagen*, 1979, vol. 13, pp. 317–354.

[6] J.-M. Hombert, "Development of tone from vowel-height," *J. Phonetics*, vol. 5, pp. 9–16, 1977.

[7] R. Petersen, "The effect of consonant type on fundamental frequency and larynx height in danish," in *Annual Report of the Inst. of Phonetics, University Copenhagen*, 1983, vol. 17, pp. 55–86.

[8] B. Möbius, A. Zimmermann, and W. Hess, "Microprosodic fundamental frequency variations in German," in *Proc. ICPhS*, vol. 1, Tallinn, 1987, pp. 146–149.

[9] A. Löfquist, "Intrinsic and extrinsic f0 variations in Swedish tonal accents," *Phonetica*, vol. 31, pp. 228–247, 1975.

[10] N. Umeda, "Influence of segmental factors on fundamental frequency in fluent speech," *J. Acoust. Soc. Am.*, 1981.

[11] J. Jan Van Santen, T. Mishra, and E. Klabbers, "Estimating Phrase Curves in the General Superpositional Intonation Model," in *Proc. ISCA Speech Synthesis Workshop*, Pittsburgh, 2004.

[12] D. Hirst and R. Espesser, "Automatic modelling of fundamental frequency using a quadratic spline function," in *Travaux de l'Institut de Phonetique d'Aix*, 1993, vol. 15, pp. 71–85.

[13] K. Kohler, "Labelled data bank of spoken standard German – the Kiel Corpus of read/spontaneous speech," in *Proc. ICSLP*, 1996, pp. 1938–1941.

[14] K. Schaefer-Vincent, "Pitch period detection and chaining: Method and evaluation," *Phonetica*, vol. 40, pp. 177–202, 1983.

[15] B. Remijsen, "momel_modif.psc," http://www.ling.ed.ac.uk/˜bert/praatsripts.html, 2004.

[16] H. Pfitzinger, H. Mixdorff, and J. Schwarz, "Comparison of Fujisaki-model extractors and F0 stylizers," in *Proc. Interspeech*, Brighton, 2009, pp. 2455–2458.