INTSINT: a new algorithm using the OMe scale

Daniel Hirst

Laboratoire Parole et Langage (LPL), CNRS and Aix-Marseille University, France

https://doi.org/10.36505/ExLing-2018/09/0012/000345

Abstract

This presentation reports work in progress on an improved and simplified algorithm for coding the output of the Momel algorithm using the INTSINT alphabet, building on recent work which proposed the Octave-Median scale (ome = log2(Hz/Median)) as a natural scale for the representation of pitch. Preliminary results comparing the output of the new algorithm with that of the standard version shows that more values are less than 1 semitone from the Momel output and the RMSD value is also lower. Further work is needed to improve this new algorithm.

Key words: Intonation, symbolic coding, INTSINT, algorithm, evaluation

Introduction

Official presentations of ToBI (e.g. Beckman et al 2005) have adopted the position that *symbolic tone labels in the ToBI framework are intended to 'tag' the intonation contour and not to 'encode' it.*The authors contrast this with the approach of INTSINT (an International Transcription System for INTonation) which had been proposed as a first approximation of a prosodic equivalent of the IPA.

The original version of the INTSINT system (Hirst 1987) was based on an inventory of minimal pitch contrasts found in published descriptions of intonation patterns. The aim was to provide a tool for the systematic description of these intonation patterns, something along the lines of a narrow transcription using the International Phonetic Alphabet (IPA) Like the IPA, it was intended that INTSINT could be used for preliminary descriptions of intonation patterns, even for languages which had not previously been described.

Notice that this aim is very different from that of the ToBI system (Silverman et al. 1992), which presupposes that the inventory of intonation patterns for the language being described has already been established.

The official website for ToBI (ToBI website) makes this particularly explicit:

Note: ToBI is not an International Phonetic Alphabet for prosody. Because intonation and prosodic organization differ from language to language, and often from dialect to dialect within a language, there are many different ToBI systems, each one specific to a language variety and the community of researchers working on that language variety.

In recent years, though, there has been a revival of interest among researchers working within the ToBI framework, in the development of such a coding system (Hualde & Prieto 2016). There have also been two recent workshops on the subject: one at the 2015 ICPhS, in Glasgow and another at the 2018 Speech Prosody Conference, Poznan, showing that there is a growing interest and need for a tool of this type.

This presentation reports work in progress on developing an improved and simplified algorithm for automatically coding the output of the Momel algorithm using the INTSINT alphabet, building on recent work (De Looze & Hirst 2014) which proposes the Octave-Median scale (ome = log2(Hz/Median)) as a natural scale for the representation of pitch.

Methodology

Campione et al (2000) compared different alternatives to the INTSINT algorithm and found that two versions provided closer fits to the Momel anchor points than the standard model. These versions (*Ampli3* and *Levels*), however, introduced tones, which unlike those of INTSINT, were not derived from phonological descriptions of intonation. Both models used three absolute tones (T, M, B) and six relative tones. This required optimising a total of 15 parameters. If the aim is simply to provide a close copy of the original anchor-points, it would be far simpler and more economical to code each directly in semitones, since a span of 15 semitones covers most of the pitch range of unemphatic utterances. Models like this do not provide a useful coding which could be used in a rule-based model of intonation.

In the most recent implementation of INTSINT (Hirst 2007), a Perl script is used to optimise both the automatic coding of the Momel anchor points and two parameters: *key* and *span,* which are used to interpret the coding.

In this presentation, I explore the possibility of coding INTSINT tones using the Octave-Median scale. Each tone is here defined by a formula, as in (1), using as variables only the median value of the pitch curve and/or the value of the preceding anchor-point (P). The tone *t*, for example[1], is defined as half an octave above the median, while the tone *h* is defined as the geometric mean of the preceding anchor point and the value of *t*. New values *t+* and *b-* are introduced as extreme values more than 2 semitones above/below the values for *t* and *b*. This is motivated by the fact that there seems to be much greater variability in the value of the tones coded by *t* in the standard system than for the other tones.

m = median;	t = m * sqrt(2)	b = m / sqrt(2)
h = sqrt(P * t)	s = P	l = sqrt(P * b)
u = sqrt(P * h) -		d = sqrt(P * l)
t+ = t * 2^(1/6)		-b- = b / 2^(1/6)

(1) Formulas for calculating pitch values corresponding to INTSINT tone labels using the new algorithm. P represents the value in Hz of the preceding anchor point.

Coding each Momel anchor-point is then simply a question of comparing the value of the current point to the current values of each tone and choosing the closest value. This represents a considerable simplification of the standard algorithm which used an iteration of 400 different codings to optimise the two parameters *key* and *span*. In the new algorithm, the value of *key* is calculated directly as the median of the pitch values and the value of *span* is taken as fixed at 1 octave.

Evaluation of this new coding algorithm is currently being carried out on recordings from the OMProDat database (Hirst et al 2013) and compared to that of the standard INTSINT coding. Results reported here are based on the analysis of the two corpora *omprodat-eng01* and *omprodat-cmn01,* each of which contain recordings of 40 5-sentence passages read by 10 speakers (5 male and 5 female).

Results and discussion

Preliminary results on two corpora of read speech, one in English and one in Mandarin Chinese show that, besides being much simpler to implement, the new algorithm gives results which are closer to the output of the Momel algorithm than with the standard version of INTSINT, when measured as RMSD (the square root of the average of squared errors in semitones) or as the number of anchor points less than one semitone from the Momel output. The results are slightly worse when measured as the number of anchor points less than 2 semitones from the Momel output for the Chinese data but not for the English data where the values were not significantly different between the two implementations.

Results for the corpus OMProDat Cmn01 (Mandarin Chinese): 10 speakers, 40 5-sentence passages.

	old version	new version	paired t	p
% < 1 st	71.35	73.85	-6.191	1.489 e-09
% < 2 st	93.88	93.03	4.323	1.949 e-05
RMSD (sts)	1.008	0.984	3.6033	3.541 e-05

Results for the corpus OMProDat Eng01 (British English): 10 speakers, 40 5-sentence passages.

	old version	new version	paired t	p
% < 1 st	83.95	87.44	-12.296	< 2.2e-16
% < 2 st	97.40	97.13	1.767	0.078 (ns)
RMSD (sts)	0.756	0.732	3.226	2.685 e-07

Notes

[1]In recent publications, I have taken the step of using lower case letters for the INTSINT symbols to distinguish them from the symbols used by other more abstract coding systems such as ToBI.

References

Beckman, M.E., Hirschberg J., Shattuck-Hufnagel, S. 2005. The original ToBI system and the evolution of the ToBI framework. In Jun, S.-A. (Ed.), Prosodic models and transcription: Towards prosodic typology, 9-54. London & New York: OUP.

Campione, E., Hirst, D.J., Véronis, J. 2000. Automatic stylisation and modelling of french and italian intonation. In Botinis, A. (Ed.), Intonation, 185-208. Dordrecht: KAP.

De Looze, C., Hirst, D.J. 2014. The OMe (Octave-Median) scale. In Campbell, N, Gibbon, D., Hirst, D.J. (Eds.), Proc. SP7, Dublin, Ireland.

Hirst, D.J. 2007. A Praat plugin for Momel and INTSINT with improved algorithms for modelling and coding intonation. Proc. XVIth ICPhS 1233-36, Saarbrucken, Germany.

Hirst, D.J. 1987. La représentation linguistique des systèmes prosodiques: une approche cognitive. Thèse de Doctorat d'Etat, Université de Provence.

Hirst, D.J., Bigi, B., Cho, H.-S., Ding, H., Herment, S., Wang, T. 2013. Building OMProDat, an open multilingual prosodic database. In Proc. TRASP, Tools and Resources for the Analysis of Speech Prosody 11-14, Aix-en-Provence, France.

Hualde, J.I., Prieto, P. 2016. Towards an International Prosodic Alphabet (IPrA). Laboratory Phonology: Journ. Association for Laboratory Phonology, 7(1), 5.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., Hirschberg. J. 1992. TOBI: A Standard for Labeling English Prosody. In Second ICSLP 867-870, Banff, Alberta, Canada.

ToBI website (consulted 2018-07-28) https://www.ling.ohio-state.edu/~tobi/.