Whodunnit – Searching for the Most Important Feature Types Signalling Emotion-Related User States in Speech: Appendix

Anton Batliner [a] Stefan Steidl [a] Björn Schuller [b] Dino Seppi [c] Thurid Vogt [d] Johannes Wagner [d] Laurence Devillers [e] Laurence Vidrascu [e] Vered Aharonson [f] Loic Kessous [g] Noam Amir [g]

[a] *FAU: Lehrstuhl für Mustererkennung, Universität Erlangen-Nürnberg, Germany*

[b] *TUM: Institute for Human-Machine Communication, Technische Universität München, Germany*

[c] *FBK: Fondazione Bruno Kessler – irst, Trento, Italy*

[d] *UA: Multimedia Concepts and their Applications, University of Augsburg, Germany*

[e] *LIMSI-CNRS, Spoken Language Processing Group, Orsay Cedex, France*

[f] *AFEKA: Tel Aviv Academic College of Engineering, Tel Aviv, Israel*

[g] *TAU: Dep. of Communication Disorders, Sackler Faculty of Medicine, Tel Aviv University, Israel*

Abstract

In this appendix, which is not enclosed in the article but can be downloaded from the web, we present the feature encoding scheme that has been developed within the initiative CEICES within the European Network of Excellence humaine.

Key words: feature types, feature selection, automatic classification, emotion

Email address: batliner@informatik.uni-erlangen.de (**Anton Batliner).**

[1] The initiative to co-operate was taken within the European Network of Excellence (NoE) HUMAINE under the name CEICES (Combining Efforts for Improving automatic Classification of Emotional user States). This work was partly funded by the EU in the projects PF-STAR under grant IST-2001-37599 and HUMAINE under grant IST-2002-50742. The responsibility lies with the authors.

References

Schröder, M., Devillers, L., Karpouzis, K., Martin, J.-C., Pelachaud, C., Peter, C., Pirker, H., Schuller, B., Tao, J., and Wilson, I. (2007). What Should a Generic Emotion Markup Language Be Able to Represent? In A. Paiva, R. Prada, and R. W. Picard, editors, *Affective Computing and Intelligent Interaction*, pages 440–451, Berlin-Heidelberg. Springer.

A Feature Encoding Scheme

To our knowledge, there is no agreed-upon taxonomy of feature types yet, let alone a machine-readable feature encoding scheme. A common understanding may be guaranteed within a single group; this is, however, not possible across groups and different research traditions. Just to give one simple example: we realized that some partners conceived features representing the temporal alignment of pitch or energy extrema – which are measured in ms. – still as representing pitch or energy. This is not wrong because pitch or energy extrema without such an alignment are not very meaningful. However, when classifying prosodic boundaries, accentuation, or emotion, we have realized that that there is a high correlation between these parameters and genuine duration parameters such as overall duration of phonetic/linguistic entities. Thus, we define such features as duration features – but of course, we still want to keep in mind that they are modelling pitch or energy contours. Another example is 'mixed' parameters: there has to be an agreement whether to attribute them to the one or the other category, or to a 'mixed' or 'not-attributable' category. All these questions are part of a general taxonomy of acoustic or linguistic parameter.

If a database has been thoroughly processed and annotated, another problem pertains to size of units, manual vs. automatic processing, etc. To give another simple example: pitch values can be obtained from different pitch detection algorithms (PDA), or from fully manual procedures, or they can be based on manual corrections of PDA results. These questions are part of a partly general, partly specific taxonomy of types of processing.

The feature encoding scheme which has been developed within the CEICES initiative is one step towards a standardisation, comparable to first attempts to define standards such as SAMPA for phonetic coding or EARL, cf. Schröder *et al.* (2007), for emotion coding.

The scheme is realized in plain ASCII representation; fields are delimited by full stops. Each position (column) has its specific semantics. Each line describes fully one of the 4244 features used in our experiments.

Identifier	**S1**
Site Id	**I12345**
Mic quality	**M1** 1 close-talk, 2 reverberated, 3 room, 9 other
Domain	**D1** 1 fixed frame-rate, 2 subword, 3 word, 4 chunk, 5 turn, 6 speaker, 9 other
Reference	**R1 domain segm. I: basic** 0 none, 1 acoust., 2 syntact., 3 emotional, 5 syntact. + acoust., 9 other **2 domain segm. II: strategy** 0 implicit, 1 manual + heuristics, 2-3 dummy manual, 4 other manual, 5-7 dummy automatic, 8 other automatic, 9 other **3 subdomain segm.** 0 none, 1 manual, 2 automatic, 9 other **4 LLD/orthography/non-verbals** 1 manual, 2 automatic, 3 mixed, 9 other
Linguistic	**L123456**
Acoustics	**A12. duration (1 interval, 2 position)** **34. energy** **56. F0** **78. spectral** **90. formant (9 number, 0 type)** **ab. cepstral (number)** **cd. voice quality** **ef. wavelets (number)** **gh. pause** **ij dummy**

Table A.1
Feature Coding Scheme: Overview; the alpha-numeric number assigns the position of the respective column within that field

Tables A.1, A.2, and A.3 display the coding. Note that due to space restrictions, a high number of abbreviations was preferred over easy readability.

Comp. context	**C1 voiced, 2 unvoiced, 3 fixed frame, 4 sub word, 5 word,** **6 chunk, 7 turn, 8 speaker, 9 whole db, 0 outside sub-corpus** 0 not appl., 1 applicable, 2 incl. l local ctxt, 3 incl. r local ctxt, 4 incl. l+r local ctxt, 6 only l local ctxt, 7 only r local ctxt, 8 only l+r local ctxt
Functionals	**F12.34.56**
Norm. context	**N12**
etXetera	**X1234567890**
Text	**T1234567890** string
Previous name	**P1234567890.....** alphanumeric

Table A.2
Feature Coding Scheme: overview, continued

00 non-applicable
extreme values
01 min, 02 max, 03 min pos, 04 max pos, 05 range, 06 mean min dist, 07 mean max dist, 08 min slope, 09 max slope, 50 on-pos, 51 off-pos, 59 other pos
mean
10 arithmetic, 11 quadratic, 12 geometric, 13 harmonic, 14 absolute, 15 conf. ivl both, 16 conf. ivl upper, 17 conf. ivl lower, 18 centroid
percentiles
20 quartile 1, 21 quartile 2 (median), 22 quartile 3, 23 quartile range 21, 24 quartile range 32, 25 quartile range 31 (iqr), 29 percentile other
higher statistical moments
30 std dev, 31 variance, 32 skewness, 33 kurtosis, 34 length, 35 sum, 36 zcr, 37 most frequent value (mode)
specific functionals/regression
38 up level time, 39 down level time, 40 micro variation, 41 #segments, 42 #rvl points, 43 #peaks, 44 mean dist rvl points, 45 std dev dist of rvl points, 46 mean peak distance, 47 ratio, 48 error, 49 other stat functional, 60 reg error, 61 lin reg coeff 1, 62 lin reg coeff 2, 63 quad reg coeff 1, 64 quad reg coeff 2, 65 quad reg coeff 3, 66 #ivls, 67 mean ivl length, 68 #positive ivls, 69 #negative ivls 70 DCT coeff 1, 71 DCT coeff 2, 72 DCT coeff 3, 73 DCT coeff 4, 74 DCT coeff 5, 79 other spectral coefficient
genetic functions
80 cf absolute value, 81 cf signum, 82 cf log, 83 cf reciprocal value, 84 cf power, 85 cf add, 86 cf minus, 87 cf mult, 88 cf div, 89 cf other
linguistic functionals
90 boolean TF, 91 word count TF, 92 log word count TF, 99 other TF functional

Table A.3
Functionals in detail by type and code. Abbreviations: cf combination functional, ivl interval, lin linear, pos position, quad quadratic, rvl reversal, std dev standard deviation, zcr zero crossing rate