

## Appendix

### openSMILE Standardised Feature Sets

*All's well that ends well.*

—William Shakespeare.

In Table [A.1](#) the LLDs and functionals and their frequency across the four openSMILE standard feature sets as were mentioned in this book are given.

LLDs are processed by simple moving average (SMA) low-pass filtering.

Delta regression coefficients are added per LLD. The total number of features is—in principle—obtained by multiplying the number of LLD times two times the number of functionals. However, for the two larger feature sets exceptions hold from this strict brute-forcing rule as are indicated. This prevents creation of non-sense features.

**Table A.1** openSMILE standard features sets by LLDs and functionals

| Feature  | EC  | PC             | SSC   | AVEC           |
|--|-----|----------------|-------|----------------|
| Frequencies  |     |                |       |                |
| # LLDs   | 16  | 38             | 59    | 31             |
| # Functionals  | 12  | 22             | 41    | 42             |
| # Features   | 384 | 1 582          | 4 368 | 1 941          |
| LLDs   |     |                |       |                |
| RMS energy   | ✓   |                | ✓     |                |
| Sum of auditory spectrum (loudness)                  |     | ✓ <sup>a</sup> | ✓     | ✓              |
| Sum of RASTA-sytle filtered auditory spectrum        |     |                | ✓     |                |
| ZCR  | ✓   |                | ✓     | ✓              |
| Energy in bands from 250–650 Hz, 1–4 kHz             |     |                | ✓     | ✓              |
| Spectral roll-off points 25 %, 50 %, 75 %, 90 %      |     |                | ✓     | ✓              |
| Spectral flux  |     |                | ✓     | ✓              |
| Spectral entropy                                     |     |                | ✓     | ✓              |
| Spectral variance                                    |     |                | ✓     | ✓              |
| Spectral skewness                                    |     |                | ✓     | ✓              |
| Spectral kurtosis                                    |     |                | ✓     | ✓              |
| Spectral slope                                       |     |                | ✓     |                |
| Psychoacousitc sharpness                             |     |                |       | ✓              |
| Harmonicity  |     |                |       | ✓              |
| MFCC 0   |     | ✓              |       |                |
| MFCC 1–10  | ✓   | ✓              | ✓     | ✓              |
| MFCC 11–12   | ✓   | ✓              | ✓     |                |
| MFCC 13–14   |     | ✓              |       |                |
| Log Mel frequency band 0–7                           |     | ✓ <sup>a</sup> |       |                |
| LSP frequency 0–7                                    |     | ✓              |       |                |
| RASTA-style auditory spectrum bands 1–26 (0 – 8 kHz) |     |                | ✓     |                |
| $F_0$ (ACF based)                                    | ✓   |                |       |                |
| $F_0$ (SHS based)                                    |     | ✓              |       |                |
| $F_0$ (SHS based followed by Viterbi smoothing)      |     |                | ✓     | ✓              |
| $F_0$ envelope                                       |     | ✓              |       |                |
| Probability of voicing                               | ✓   | ✓              | ✓     | ✓              |
| Jitter   |     | ✓              | ✓     | ✓              |
| Jitter (delta: ‘jitter of jitter’)                   |     | ✓              | ✓     | ✓              |
| Shimmer  |     | ✓              | ✓     | ✓              |
| Logarithmic HNR                                      |     |                |       | ✓              |
| Functionals  |     |                |       |                |
| Positive arithmetic mean                             |     |                |       | ✓ <sup>d</sup> |
| Arithmetic mean                                      | ✓   | ✓              | ✓     | ✓ <sup>d</sup> |
| Root quadratic mean                                  |     |                |       | ✓              |
| Contour centroid                                     |     |                | ✓     |                |
| Standard deviation                                   | ✓   | ✓              | ✓     | ✓              |
| Flatness   |     |                |       | ✓              |

(continued)

**Table A.1** (continued)

| Feature  | EC | PC             | SSC            | AVEC             |
|--|----|----------------|----------------|------------------|
| Skewness   | ✓  | ✓              | ✓              | ✓                |
| Kurtosis   | ✓  | ✓              | ✓              | ✓                |
| Quartiles 1, 2, 3  |    | ✓ <sup>a</sup> | ✓              | ✓                |
| Inter-quartile ranges 2-1, 3-2, 3-1                          |    | ✓ <sup>a</sup> | ✓              | ✓                |
| Percentile 1 %, 99 %   |    | ✓ <sup>a</sup> | ✓              | ✓                |
| Percentile range 1–99 %                                      |    | ✓              | ✓              | ✓                |
| % frames above minimum + 25%, 50% of range                   |    |                |                | ✓                |
| % frames above minimum + 75 % of range                       |    | ✓ <sup>a</sup> |                |                  |
| % frames above minimum + 90 % of range                       |    | ✓ <sup>a</sup> | ✓              | ✓                |
| % frames below minimum + 25 % of range                       |    |                | ✓              |                  |
| % frames rising  |    |                | ✓              | ✓                |
| % frames falling   |    |                | ✓              |                  |
| % frames left, right curvature                               |    |                | ✓ <sup>f</sup> |                  |
| % frames that are non-zero                                   |    |                | ✓ <sup>b</sup> |                  |
| Linear regression offset                                     | ✓  | ✓ <sup>a</sup> |                |                  |
| Linear regression slope                                      | ✓  | ✓ <sup>a</sup> | ✓              | ✓ <sup>c</sup>   |
| Linear regression approximation error (MAE)                  |    | ✓ <sup>a</sup> |                | ✓ <sup>c</sup>   |
| Linear regression approximation error (MSE)                  | ✓  | ✓ <sup>a</sup> | ✓              |                  |
| Quadratic regression coefficient <i>a</i>                    |    |                | ✓              | ✓ <sup>c</sup>   |
| Quadratic regression coefficient <i>b</i>                    |    |                | ✓              |                  |
| Quadratic regression approximation error (MAE)               |    |                |                | ✓ <sup>c</sup>   |
| Quadratic regression approximation error (MSE)               |    |                | ✓              |                  |
| Maximum, minimum   | ✓  |                |                |                  |
| Maximum–minimum (range)                                      | ✓  |                |                |                  |
| Rising, falling slopes (min to max) mean, standard deviation |    |                |                | ✓ <sup>c</sup>   |
| Inter maxima distances mean, standard deviation              |    |                | ✓              | ✓ <sup>c</sup>   |
| Amplitude mean of maxima relative to mean                    |    |                |                | ✓ <sup>c</sup>   |
| Amplitude range of minima relative to mean                   |    |                |                | ✓ <sup>c</sup>   |
| Amplitude range of maxima relative to mean                   |    |                |                | ✓ <sup>c</sup>   |
| Relative position of maximum, minimum                        | ✓  | ✓ <sup>a</sup> |                |                  |
| LP gain  |    |                | ✓              | ✓ <sup>c,e</sup> |
| LP coefficients 1–5  |    |                | ✓              | ✓ <sup>c,e</sup> |
| Peak value arithmetic mean                                   |    |                | ✓              |                  |
| Peak value arithmetic mean–arithmetic mean                   |    |                | ✓              |                  |
| Segment length mean, max, min, standard deviation            |    |                | ✓ <sup>b</sup> | ✓ <sup>e</sup>   |
| Input duration in seconds                                    |    | ✓ <sup>b</sup> | ✓ <sup>b</sup> |                  |

EC:INTERSPEECH 2009 Emotion Challenge, PC:INTERSPEECH 2010 Paralinguistic Challenge, SSC:INTERSPEECH 2011 Speaker Trait Challenge, AVEC:Audio/Visual Emotion Challenge 2011

<sup>a</sup>Only used for the TUM AVIC baseline (PC)

<sup>b</sup>Only applied to  $F_0$

<sup>c</sup>Not applied to delta coefficient contours

<sup>d</sup>For delta coefficients the mean of only positive values is applied, otherwise the arithmetic mean is applied

<sup>e</sup>Not applied to voicing related LLDs

<sup>f</sup>Only applied to voicing related LLDs. For the PC feature set, the two additional features turn duration and number of voiced segments ( $F_0$  onsets) were added

# Index

## A

Accuracy, 133  
Acoustic event classification, 305  
Acoustic model, 122  
AdaBoost, 126  
Advanced front-end, 151  
aGender corpus, 200  
ALC database, 206  
Anti-formants, 59  
Audio activity detection, 45  
Audio chord estimation, 264  
Audio data, 23  
Audio drum detection, 226  
Audio key detection, 252  
Audio mood classification, 274  
Audio onset detection, 233  
Audio tempo extraction, 239  
Auto correlation, 49  
Automatic speech recognition, 170, 176  
Autoregressive model, 160

## B

Backpropagation, 112  
Backtracking, 124  
Bag of words, 75  
Bagging, 125  
Balancing, 131  
Ballroom dance style  
    classification, 239  
Baum–Welch estimation, 120  
Beam search, 123  
Bello database, 235  
Blind source separation, 226  
Boosting, 126  
BRD database, 243

## C

CENS features, 67  
Cepstral mean subtraction, 154  
Cepstrum, 50  
ChoRD database, 265  
Chord-based features, 70  
Chords, 276  
CHROMA features, 67  
Chunking, 41  
Circle of fifths, 258  
Comb filter, 244  
ConceptNet, 78  
Conditional random field, 159  
Confusion matrix, 135  
Continuous speech recognition, 176  
Correlation coefficient, 135  
COSINE corpus, 176  
Cross-validation, 131

## D

Decision trees, 99  
Digitalisation, 41  
Dynamic bayesian network, 118

## E

Emotional find sounds database, 35  
Ensemble learning, 124  
Evaluation measures, 133  
Expectation maximisation, 120

## F

F-Measure, 134  
False negatives, 135

False positives, 134  
 FAU aibo emotion corpus, 194  
 Feature brute-forcing, 83  
 Feature enhancement, 153  
 Feature reduction, 90  
 Feature selection, 90  
 Feed forward neural network, 112  
 FindSounds database, 34  
 Formants, 58  
 Functional, 91

## G

General inquirer, 79  
 Gold standard, 24  
 Ground truth, 24

## H

Harmonics-to-noise ratio, 60  
 Hidden Markov model, 118  
 Histogram equalisation, 155  
 HU-ASA database, 300

## I

Information gain, 102  
 Intensity, 47

## J

Jitter, 63

## K

Kernel function, 107  
 Key databases, 254

## L

Language model, 122  
 Leave one out, 131  
 Line spectral pairs, 56  
 Linear prediction, 52  
 Long short-term memory, 115  
 Low-level descriptor, 47  
 Lyrics, 278

## M

Mean absolute error, 136  
 Mean and variance normalisation, 155  
 Mean linear error, 136  
 Metacritic database, 185

Metadata, 277  
 Metre features, 247  
 Metre recognition, 249  
 MFCC features, 51  
 MPEG-7 low level descriptors, 73  
 Multilayer perceptron, 112  
 Multistream models, 129  
 Music information retrieval, 225

## N

N-gram, 77  
 Neural network, 110  
 Non-linguistic vocalisation, 180  
 Non-negative matrix deconvolution, 140  
 Non-negative matrix factorisation, 140  
 NTWICM corpus, 28

## O

On-line knowledge sources, 77  
 Onset, 233  
 OpenBLISSART, 143  
 OpenSMILE, 83  
 Opinion mining, 185  
 Out of vocabulary, 74

## P

Paralinguistics, 183  
 Partitioning, 130  
 Peak detection, 237  
 Perceptual linear prediction, 56  
 Pitch, 60  
 Pitch class profiles, 64  
 Posterior estimation, 158  
 Precision, 134  
 Probe tone ratings, 70  
 Pruning, 123

## R

Random forests, 103  
 Recall, 133  
 Recurrent neural network, 113

## S

Scale-based features, 68  
 Self-similarity matrix, 242  
 Semi-supervised learning, 305  
 Sentiment analysis, 185  
 Shimmer, 63  
 Singer trait recognition, 283

- SLC database, 207
- Softmax function, 112
- Sound emotion recognition, 308
- Source filter model, 50
- Source-filter production model, 50
- Speaker age recognition, 198
- Speaker gender recognition, 198
- Speaker height recognition, 198
- Speaker interest recognition, 193
- Speaker intoxication recognition, 204
- Speaker sleepiness recognition, 204
- Spectral features, 276
- Speech activity detection, 46
- Speech emotion recognition, 193
- Speech production, 50
- Stacking, 127
- Stemming, 73
- Stopping, 73
- Structural segmentation, 269
- Support vector machines, 103
- Support vector regression, 109
- Switching models, 156

## T

- Tandem learning, 129
- Tatum features, 246
- TI 46 isolated word corpus, 170
- TIMIT database, 201

- True negatives, 135
- True positives, 134
- TUM AVIC corpus, 26

## U

- UltraStar database, 284
- Unsupervised spectral subtraction, 151
- Unweighted accuracy, 134

## V

- Vanishing gradient problem, 115
- Viterbi algorithm, 119
- Vocabulary, 75
- Voice activity detection, 46

## W

- Weighted accuracy, 133
- Windowing, 43
- Word accuracy, 134
- WordNet, 79

## Z

- Zero crossing rate, 48