Hilbert-Huang Transform for Non-Linear Characterization of Speech Rhythm

Fabien Ringeval and Mohamed Chetouani

Université Pierre et Marie Curie – Paris 6,
Institut des Systèmes Intelligents et de Robotique, 4 place Jussieu,
75005 Paris, France.
Fabien.Ringeval@isir.fr , Mohamed.Chetouani@upmc.fr

Abstract. A method for non-linear and non-stationary characterisation of speech rhythm is presented using Hilbert Huang Transform (HHT) of 'Speech Unit Intervals' (SUI) signals. SUI signals are supported by intervals duration between given speech units such as vowel, consonant, or syllable. While HHT is based on the combination of the Empirical Mode Decomposition (EMD) and the Hilbert transform of the provided Intrinsic Mode Functions (IMFs). Since EMD is a data-driven approach which includes both signal-dependent and time-variant filtering, HHT analysis on the SUI signals makes it possible non-linear and non-stationary characterisation of the speech rhythm. Investigations on the HHT based rhythmic features are presented in this paper: emotional speech classification is individually performed on rhythmic features, and obtained classification probabilities are fused with those provided by a typical state-of-the-art emotion recognition system based on acoustic and prosodic features sets.

Keywords: Speech rhythm modelling, Hilbert-Huang Transform, Emotion recognition.

1 Introduction

Majority of recent approaches used for modelling speech rhythm employ interval durations to describe the temporal patterns of speech [1,2]. Interval durations were defined by the onsets of linguistic units (syllable/stress timing), suggesting a relative isochrony of interval durations across languages [3,4]. Since the isochronous units failed in finding cross-linguistic differences [5,6,7], a reconsideration of the speech rhythm was engaged: temporal properties of vocalic and consonantal units were then employed for the rhythmic characterisation of speech [8]. Concerning the metrics, two measures have mainly been proposed: global statistics such as %V and ΔC [9], and statistics derived from the pair-wise variability indices (PVI: differences between pair-wise intervals duration) [10]. Both languages and dialects discrimination could have been achieved by many authors with these metrics [9,10,11]; strong correlations were then fund between them. Statistics such as mean and standard-deviation from linguistic speech unit seem to perform well enough for language discrimination. Yet, speech rhythm is chaotic [12,13] and includes thus non-linear properties. Indeed, the rhythm of speech refers to the alternations of short/long durations interval and

weak/strong perceptual levels, with additional variations across both different speech units and languages. Moreover, pauses between groups of speech units as well as specific acoustic shapes conveyed by the latter can also be considered as rhythmic events, e.g. similar results were obtained by %V-ΔC and PVI metrics and by a sonority measure of the obstruency in the speech signal in [14]. Furthermore, the ability to perceive rhythmic differences in both time and space lies on cognition (motor action in response to an outer source) and is individual [13]. Hence the speech rhythm is non-linear.

One could therefore wonder if measuring statistics directly from interval durations of given speech units is an appropriate way for the characterisation of speech rhythm. Because these statistics are not designed for capturing non-linear phenomena conveyed by the speech rhythm. Likewise, Fourier analysis from 'p-centers' filtered signal [15] may not be a suitable measure for rhythmic modelling also since it requires stationary signals.

In this paper, we propose a new method for modelling non-linear phenomena of speech rhythm by using Hilbert Huang Transform (HHT) of "Speech Unit Intervals" (SUI) signals [16]. SUI signals are built on intervals duration between speech units, i.e. vowel, consonant, or syllable. While HHT is based on the combination of the Empirical Mode Decomposition (EMD) and the Hilbert transform of the provided Intrinsic Mode Functions (IMFs). Since the EMD algorithm is a data-driven approach involving both signal-dependent and time-variant filtering, HHT analysis of the SUI signals leads to the characterisation of non-linear and non-stationary phenomena conveyed by the speech rhythm. The aim of this paper is therefore to determine if proposed HHT based rhythmic features can improve emotion recognition performances of a typical state-of-the-art system composed of acoustic and prosodic features sets. Figure 1 illustrates the approach used for the experiments.

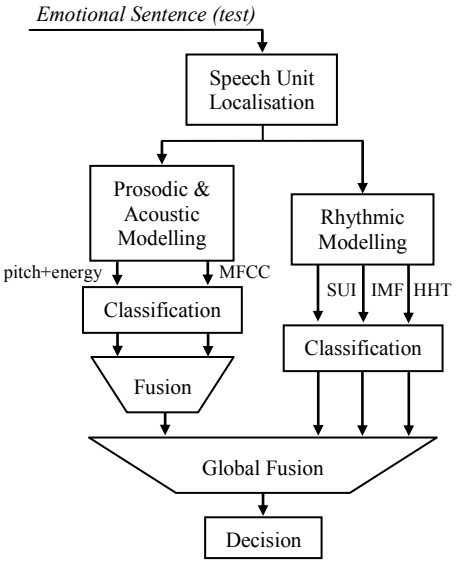

Fig. 1. Assessment of the HHT based rhythmic features by individual classification and fusion with typical state-of-the-art based emotion recognition system

The remainder of this paper is organized as follows: the SUI signals, EMD and HHT are described in section 2. Emotion classification results from acoustic, prosodic and HHT features sets as well as the fusion of these approaches are presented in section 3.

2 Proposed Method for Rhythmic Modelling

The following sections are dedicated to the description of the proposed method for the non-linear characterisation of speech rhythm. First, we describe how the SUI signals are constructed, then we present HHT analysis which is composed of both EMD and Hilbert transform.

2.1 Speech Unit Intervals Signals

SUI signals are used for extracting HHT based rhythmic features. These signals are provided by a resampling process (cubic spline, Fs = 32 Hz) on intervals duration computed from central positions of localised speech unit SU in the time domain (figure 3). The choice of the sampling frequency is motivated by the fact that we want to model rhythmic features for different linguistic units (i.e. vowels, consonants and syllables). Consequently, the lowest interval duration that can be found on these data must be obtained by the phonetic units, for which the frequency range is known to vary from 1 Hz to 16 Hz [17]. Since 16 Hz (Fs/2) is equivalent to 1 phoneme each 6.3 ms, a value close to the lowest duration of the consonant, all rhythmic frequencies provided by the studied speech units will be thus available for the analysis.

2.2 Empirical Mode Decomposition

The first step of Hilbert-Huang transform consists of empirical mode decomposition. EMD is a data-driven approach wherein a time series x(t) is decomposed into a finite number of its individual characteristic oscillations, termed Intrinsic Mode Functions (IMFs) [16]. IMFs are iteratively extracted through a local representation of the signal x(t), considered as the sum of an oscillating component d(t) – high-frequency part – and a local trend m(t) – low frequency-part. IMF components are obtained by a sifting process that requires two conditions: zero-mean and either identical number of both extrema and zero-crossing, or differ by one. The signal x(t) is thus represented as a sum of N Intrinsic Mode Functions d_k and the final residual components r_k:

$$x(t) = \sum_{k=1}^{N} d_k(t) + r_k(t) \ . \tag{1}$$

Considering that over-iteration leads to over-decomposition, [18] proposed a new criterion for stopping the sifting process of the EMD. This criterion is based on two thresholds θ_1 and θ_2 which ensures globally small fluctuations in the mean while taking into account local large excursions. In this approach, the EMD sifting process

is iterated until an evaluation function σ remains below θ_1 for a fraction (1-α) of the total duration, and below θ_2 for the remaining fraction. Where the evaluation function σ is defined as the absolute ratio between the mean m(t) and the residual d(t). Default coefficient values are set to α = 0.05, θ_1 = 0.05 and θ_2 = 10 θ_1. The EMD algorithm can be summarised as follows [16]:

1. Extract all the extrema of x(t)
2. Interpolate between minima (resp. maxima) to obtain two envelopes: e_{min}(t) and e_{max}(t)
3. Compute the mean: $m(t) = (e_{min}(t) + e_{max}(t))/2$
4. Extract the detail: $d(t) = x(t) - m(t)$
5. Iterate on the residual if all sifting conditions unsatisfied

2.3 Hilbert Transform

The second step of HHT analysis is Hilbert transform. Starting from the fact that IMFs are narrow-band signals with decreasing frequencies, Huang et al. proposed to apply the Hilbert transform on all IMFs provided by the EMD [16]. This makes it possible to localise well both the instantaneous frequency and temporal envelop of a real signal x(t) in the time-frequency domain. The Hilbert transform of x(t) is defined as:

$$y(t) = \frac{1}{\pi} p.v. \int_{-\infty}^{+\infty} \frac{x(\tau)}{t-\tau} d\tau \ . \tag{2}$$

where p.v. refers to the Cauchy principal value. The analytic signal of x(t) can be then defined as:

$$z(t) = x(t) + iy(t) = a(t)e^{i\theta(t)} \ . \tag{3}$$

where

$$a(t) = \sqrt{x^2(t) + y^2(t)} \text{ and } \theta(t) = \arctan\left(\frac{y(t)}{x(t)}\right). \tag{4}$$

The Hilbert transform is a powerful method for extracting instantaneous attributes of a non-linear and non-stationary time series, especially the envelop and frequency. The envelop of the signal x(t) is provided by the amplitude a(t) of its complex Hilbert transform z(t), while instantaneous frequency f(t) is obtained by differentiation on the phase θ(t):

$$f(t) = \frac{1}{2\pi} \frac{d\theta(t)}{dt}. \tag{5}$$

When Hilbert-Huang transform is applied to the SUI signals (figure 3), some precautions have to be taken for the estimation of the instantaneous frequency. Indeed, negative values can be obtained on the first and last part of the signal due to

finite observation lengths. Introducing zeros in the first and last values of SUI signals before its resampling shifts the problem of frequency estimation on these values. Concerning errors which occurs during the estimation of the extrema in EMD, and which are related to the frequency *Fs* used for resampling, we did not found any significant differences in the reconstruction error (equation 6) for increasing *Fs* values (figure 2).

$$\frac{e(Fs)}{2} = \log\left(x(t) - \sum_{k=1}^{N} d_{k,Fs} + r_{k,Fs} \right) - \log\left(\sum_{k=1}^{N} d_{k,Fs} + r_{k,Fs} \right). \quad \textbf{(6)}$$

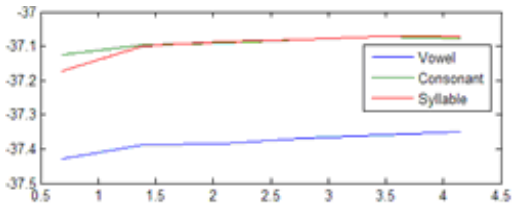

Fig. 2. Reconstruction error *e*(*Fs*) according to the log of the sampling frequency *Fs*

Local fluctuations on the phase due to resampling involves wider time variations on the instantaneous frequency because of differentiation (equation 5). Two passes of both mean and median Simple Moving Average (SMA) filters on the instantaneous phase and frequency signals (respectively) reduces considerably the local fluctuations as depicted in figure 3.

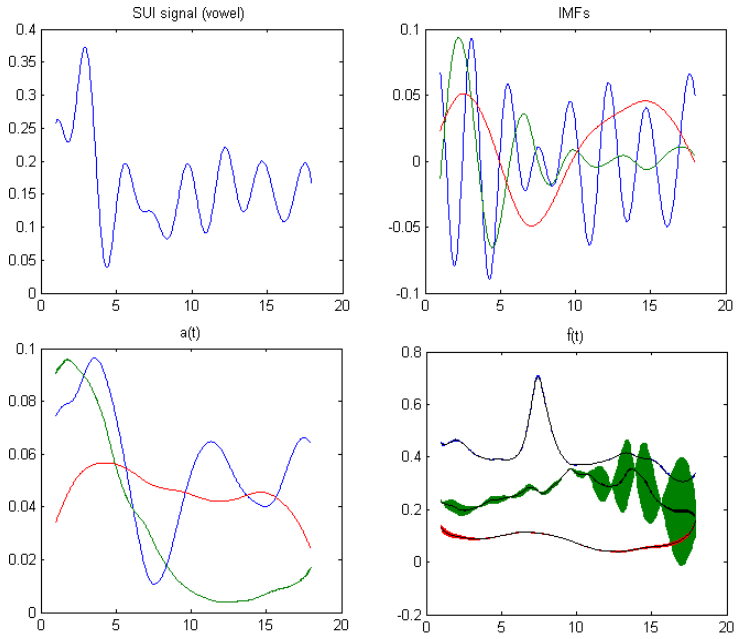

Fig. 3. HHT analysis on a SUI signal: IMFs and their rhythmic envelops *a*(t) and instantaneous frequency *f*(t). Output of the two-passes of the median filter are plotted in black colour for *f*(t).

3 Emotion Recognition

The database used for the experiments is presented in section 3.1. Features used for the characterisation of acoustic, prosodic and rhythmic signals are described in section 3.2. Classification approaches and optimisation techniques as well as results are given in section 3.3, while fusion is analysed in section 3.4.

3.1 Database

The Berlin corpus [19] contains German emotional speech (six primary emotions plus a 'neutral' style) and is commonly used in the emotion recognition [20, 21, 22]. 10 utterances (five short and five long) which could be used in everyday communication have been emotionally coloured by ten gender equilibrated native German actors, with high quality recording equipment (anechoic chamber). 535 sentences marked as min. 60% natural and min. 80% recognisable by 20 listeners in a perception test have been kept and phonetically labelled in a narrow transcription. Both phonemes (vowel and consonant) and syllable based timing synchronisation are included in these data.

3.2 Features Extraction

Three different sets of features are based on acoustic, prosody and rhythm and are extracted on the emotional sentences provided by the Berlin database.

Acoustic features. Sentences are segmented into 32 ms frames with an overlap ratio equal to ½. 22 MFCC coefficients (Mel Frequency Cepstrum Coding) are computed on frames with hamming temporal window and triangular filter banks configurations.

Prosodic features. Prosodic features are composed of statistics from both pitch and energy. Pitch is automatically extracted each 10ms by an autocorrelation based method (KTH snack toolbox). Obtained values are converted into half-tones to better cope with the human perception system (f ref. = 55Hz). Energy is provided in *dB* with the same frame rate than pitch. Two post-processes are applied on both pitch and energy signals: minimal segmental duration threshold (30ms) and median SMA filter (30ms window).

Rhythmic features. SUI signals (vowels, consonants and syllables) and their HHT analysis including IMFs, envelop and instantaneous frequency constitute the rhythmic features. The number of maximum intrinsic mode functions provided by the EMD was fixed to 3. Rhythmic features are characterised by the same statistics used to model both pitch and energy (table 1).

Table 1. Set of statistics used for both prosodic and rhythmic characterisation; rp_max/min: relative position of the maximum/minimum; rp_adif: |rp_max – rp_min|; on_v: onset value; tar_v: target value; off_v: offset value; *range is normalised by rp_adif

Statistic	n	Statistic	n
max	1	3rd quartile	13
rp_max	2	IQR	14
min	3	|IQR – std|	15
rp_min	4	jitter	16
rp_adif	5	regression slope	17
range*	6	onset value	18
mean	7	target value	19
std	8	offset value	20
skewness	9	|tar_v – on_v|	21
kurtosis	10	|off_v – on_v|	22
1st quartile	11	|off_v – tar_v|	23
median	12		

3.3 Classification

Two different optimisation techniques were involved at the feature level: ZN and ZA normalisation and features selection for both prosodic and rhythmic based features. Z-scores normalisations are either based on neutral emotional data (ZN) or all emotional data (ZA). Features from each gender (male and female) are normalised separately. The *RELIEF-F* algorithm [23] provided by the Weka toolbox [24] was employed for feature selection. RELIEF-F is based on the computation of both *a priori* and *a posteriori* entropy of the features according to the emotional classes.

A segment based approach (SBA) was employed for classifying the features based on acoustic, while a turn based approach (TBA) was used for both prosodic and rhythmic features [21,25]. SBA consists of classifying the MFCC frames according to their corresponding time-synchronised speech unit (e.g. vowel). The mean of the obtained classification probabilities is computed to get one vector $p(C_i \mid SU_x)$ for each speech unit SU_x. Duration characteristics of *SU* are used as weights for each vector of emotional probabilities. Emotion decision from the seven classes C_i included in the Berlin database is took by an *argmax* function on the mean of the final weighted vectors for a given emotional sentence:

$$E = \arg\max_i \left(\frac{1}{N} \sum_{x=1}^{N} p(C_i \mid SU_x) * length(SU_x) \right) . \quad (7)$$

TBA considers classification of one feature vector F per sentence:

$$E = \arg\max_i \left(p(C_i \mid F) \right) . \quad (8)$$

A k-nearest-neighbours distance based classifier was used for the experiments with k = 5, and a 10-fold stratified cross validation scheme was employed for the scoring computation. Classification results of the studied approaches are given in table 2.

Table 2. Emotion recognition performances of acoustic, prosodic and rhythmic based features according to the different speech units: vowel, consonant and syllable. The number of selected features are given in small size characters.

Data	Vowel			Consonant			Syllable		
	Raw	**ZN**	**ZA**	**Raw**	**ZN**	**ZA**	**Raw**	**ZN**	**ZA**
Acoustic	**72.0**	68.4	71.2	67.0	66.0	**70.0**	71.2	**74.0**	71.2
Prosodic	57.8_{37}	52.4_{21}	60.2_{23}	65.4_{23}	59.2_{36}	69.8_{32}	65.6_{30}	60.4_{16}	63.2_{38}
SUI	36.8_{5}	35.2_{8}	40.4_{17}	37.2_{14}	37.2_{16}	37.6_{5}	32.6_{23}	27.6_{16}	31.2_{18}
IMF 1	34.4_{5}	30.8_{3}	32.6_{17}	33.4_{10}	32.0_{4}	33.2_{6}	32.4_{9}	29.4_{6}	28.6_{5}
IMF 2	30.6_{3}	30.8_{7}	31.4_{23}	37.0_{4}	34.7_{6}	37.0_{9}	31.4_{18}	34.3_{16}	35.0_{9}
IMF 3	35.6_{23}	35.9_{8}	34.5_{23}	-	-	-	30.0_{1}	31.3_{3}	36.3_{23}
Mod 1	33.2_{4}	33.0_{9}	34.8_{12}	31.8_{4}	32.2_{4}	32.2_{6}	30.8_{5}	26.8_{16}	29.8_{7}
Mod 2	32.0_{3}	30.3_{3}	29.0_{10}	36.3_{3}	33.0_{4}	35.7_{8}	35.4_{3}	33.9_{4}	35.7_{7}
Mod 3	39.7_{1}	35.2_{10}	35.5_{3}	-	-	-	37.5_{11}	37.5_{11}	35.0_{11}
Freq 1	26.2_{23}	26.2_{21}	30.2_{12}	29.6_{10}	28.4_{12}	31.4_{19}	26.0_{23}	24.6_{6}	24.8_{10}
Freq 2	25.3_{5}	26.5_{5}	28.8_{19}	31.3_{5}	29.7_{20}	34.3_{12}	25.7_{20}	27.1_{5}	24.3_{3}
Freq 3	29.7_{5}	27.9_{3}	32.4_{7}	-	-	-	31.3_{7}	26.3_{23}	31.3_{2}
Rhythm	31.6_{34}	30.0_{31}	31.0_{18}	38.8_{8}	36.2_{8}	40.6_{26}	38.6_{4}	35.2_{6}	39.6_{16}

Best performances are obtained by the acoustic data, while HHT based rhythmic features give the lowest scores. The third IMF component provided by the consonant based SUI signals does not provide enough information for classifying the data. This lets us suppose that consonant based SUI signals convey low non-linearity properties compared to the other *SU*. Z-score normalisation depends on the studied speech units since best scores are obtained by the ZN approach for consonant, ZA for syllables and raw data for vowels.

3.4 Fusion

Figure 1 illustrates the method used for the fusion of the classification probabilities provided by the rhythmic based features. Fusion is based on a linear combination of the classification probabilities obtained by two given approaches. Two weights are then used for realising fusion wherein an optimal configuration is searched for a number of given iterations (50). Results presented in table 3 are very interesting since improvement of the state-of-the-art fused system is accomplished by many HHT based rhythmic features, even if the former performs more than twice better than the latter. With HHT based rhythmic features a mean relative error improvement of 5.4% was obtained for all studied approaches (z-normalisation and *SU*). However, not each rhythmic features increase emotion recognition performances of the state-of-the-art based system. While fusion with the features based on both IMFs components and their temporal envelops perform well, fusion with the instantaneous frequency does not provide any improvement. Different scoring patterns are obtained according to the speech units: vowels seems to convey more rhythmic features related to emotional speech than syllables since few approaches led to an improvement of the results on the latter.

Table 3. Emotion recognition performances of both acoustic and prosodic fusion (state-of-the-art results) and fusion of the rhythmic based features with state-of-art system according to the speech units: vowel, consonant and syllable. Optimal configurations of the weights are given in small size characters; – means than no improvement could have been obtained (either lower or equal performances).

Fusion	Vowel			Consonant			Syllable		
	Raw	**ZN**	**ZA**	**Raw**	**ZN**	**ZA**	**Raw**	**ZN**	**ZA**
A. & P.	$75.6_{9/1}$	$70.5_{9/1}$	$75.2_{8/2}$	$75.2_{7/3}$	$71.4_{8/2}$	$78.1_{7/3}$	$76.2_{9/1}$	$73.4_{7/3}$	$\mathbf{78.5}_{9/1}$
SUI	–	$72.1_{9/1}$	$75.4_{9/1}$	$75.6_{9/1}$	–	$\mathbf{78.8}_{9/1}$	$77.3_{9/1}$	–	–
IMF 1	$\mathbf{77.0}_{9/1}$	–	–	$75.4_{9/1}$	–	–	$76.8_{9/1}$	–	–
IMF 2	–	$72.2_{9/1}$	$76.5_{9/1}$	–	–	–	–	–	–
IMF 3	–	–	–	–	–	–	–	$73.7_{9/1}$	–
a(t) 1	$76.0_{9/1}$	–	–	$75.9_{9/1}$	–	–	–	–	–
a(t) 2	$76.0_{9/1}$	$70.8_{9/1}$	$75.5_{9/1}$	–	–	–	$76.5_{9/1}$	–	–
a(t) 3	–	–	–	$76.3_{9/1}$	–	–	–	–	–
f(t) 1	–	–	–	–	–	–	–	–	–
f(t) 2	–	–	–	–	–	–	–	–	–
f(t) 3	–	–	–	–	–	–	–	–	–
HHT	–	$72.9_{9/1}$	$75.6_{9/1}$	$77.3_{9/1}$	$73.8_{9/1}$	–	–	–	–

4 Conclusion

A new method for extracting non-linear and non-stationary characteristics of the speech rhythm is presented in this paper: Speech Unit Intervals (SUI) signals are analysed by Hilbert-Huang transform (HHT). This approach leads to characterise many useful features, such as intrinsic mode functions and their temporal envelops a(t) and instantaneous frequencies f(t). Investigations on the HHT based rhythmic features are presented in this paper: emotional recognition is performed on these features individually, and obtained classification probabilities are then fused with those provided by a typical state-of-the-art emotion recognition system including fusion of both acoustic and prosodic features sets. Results show that even if the state-of-the-art system performs more than twice better than the HHT based rhythmic features (table 2), fusion from these two approaches leads to a mean improvement of 5.4% of the relative error for all the studied speech units as well as classification configurations (table 3). This reveals the interest of non-linear modelling of the speech rhythm, namely for the emotion recognition.

References

1. Cummins, F.: Speech rhythm and rhythmic taxonomy. In Proc. of Speech Prosody (2002) 121–126
2. Campbel, W. N.: Syllable-based segmental duration. In G. Bailly et al. (eds.) Talking Machines. Theories, models and designs, Elsevier Science Publishers (1992) 211–224

3. Pike, K. L.: The intonation of American English. University of Michigan Press, Ann Arbor (1945)
4. Abercrombie, D.: Elements of general phonetics. Edinburgh University Press (1967)
5. Bolinger, D.: Aspects of language. Harcourt, Brace and World, New York (1968)
6. Lehiste, I.: Isochrony reconsidered. Journal of Phonetics, Vol. 5, no. 3 (1977) 253–263
7. Roach, P.: On the distinction between 'stress-timed' and 'syllable-timed' languages. In D. Crystal, Linguistic Controversies, Arnold, London (1982)
8. Dauer, R. M.: Stress-timing and syllable-timing reanalyzed. Journal of phonetics, Vol. 11 (1983) 51–62
9. Ramus, F., Nespor, M. and Mehler, J.: Correlates of linguistic rhythm in the speech signal. Cognition, Vol. 73 (1999) 265–292
10. Grabe, E. and Low, E. L..: Durational variability in speech and the rhythm class hypothesis. C. Gussenhoven and N. Warner (eds.) Papers in Laboratory Phonology 7, Berlin, New York: Mouton de Guyter (2002)
11. O'Rourke, E.: Correlating speech rhythm in Spanish: evidence from two Peruvian dialects. In Proc. of the 10th Hispanic Linguistics Symposium (2008) 276–287
12. Zellner Keller, B. and Keller, E.: The chaotic nature of speech rhythm: hints for fluency in the language acquisition process. In Delcloque, Ph., Holland, V.M. (eds.) Integrating Speech Technology in Language Learning, Swets & Zeitlinger, In Press (2000)
13. Evans, J.R. and Clynes, M.: Rhythm in psychological, linguistic and musical processes. Springfield, Charles C. Thomas (1986)
14. Galves, A., Garcia, J., Duarte, D. and Galves, C.: Sonority as a basis for rhythmic class discrimination. In Proc. of Speech Prosody (2002) 11–13
15. Tilsen, S. and Johnson, K.: Low-frequency Fourier analysis of speech rhythm. The Journal of the Acoustic Society of America, Vol. 124, Issue 2 (2008) 34–39
16. Huang, N., Shen, Z., Long, S. et al.: The empirical mode decomposition and Hilbert spectrum for nonlinear and nonstationary time series analysis. In Proc. R. Soc. London, Ser. A, Vol. 454 (1998) 903–995
17. Drullman, R., Festen, J. M. and Plomp, R.: Effect of temporal envelope smearing on speech reception. Journal of the Acoustical Society of America, Vol. 95 (1994) 1053–1064
18. Riling, G., Flandrin, P. and Gonçalvès, P.: On empirical mode decomposition and its algorithms. In Proc. of IEEE-EURASIP workshop on NSIP (2003).
19. Burkhardt, A., Paeschke, M., Sendlmeier, W. and Weiss, B.: A Database of German Emotional Speech. In Proc. of Interspeech (2005) 1517–1520
20. Schuller, B., and Rigoll, G.: Timing levels in segment-based speech emotion recognition. In Proc. of Interspeech-ICSLP (2006) 1818–1821
21. Schami, M. and Verhelst, W.: An evaluation of existing supervised machine learning approaches to the classification of emotions in speech. Speech Communications, Vol. 49, no. 3 (2007) 201–212
22. Ringeval, F. and Chetouani, M.: A vowel based approach for acted emotion recognition. In Proc. of Interspeech (2008) 2763–2766
23. Robnik, M. and Konenko, I.: Theoretical and empirical analysis of ReliefF and RReliefF. Machine Learning Journal, Vol. 53 (2003) 23–69
24. Witten, I. H. and Frank, E.: Data mining: pratical machine learning tools and techniques, 2nd edition, Morgan Kaufmann, San Fransisco (2005)
25. Vlasenko, B., Schuller, B., Wendermuth, A. and Rigoll, G.: Frame vs. turn-level: emotion recognition from speech considering static and dynamic processing. In Proc. of ACII, Springer-Verlag publisher (2007) 139–147