Computational annotation-mining of syllable durations in speech varieties

Jue Yu[1], *Dafydd Gibbon*[2], *Katarzyna Klessa*[3]

[1] School of Foreign Languages, Tongji University, China
[2] Fakultät für Linguistik und Literaturwissenschaft, Universität Bielefeld, Germany
[2] Institute of Linguistics, Adam Mickiewicz University in Poznań, Poland
gibbon@uni-bielefeld.de, erinyu@126.com, klessa@amu.edu.pl

Abstract

There are many techniques for modelling properties of speech duration patterns, including models of rhythm as oscillation, partial models of rhythm types as departures from isochrony, models of tempo acceleration and deceleration, and models of duration hierarchies and their relation to hierarchies in word and phrase structure. Except for oscillator modelling, many approaches use data extraction from speech annotations, often with mainly manual methods. We employ computational data-mining for phonetic research, as opposed to phonological research on the one hand or speech technological research on the other, and explore the potential of the computational annotation data-mining paradigm for improving efficiency and scope of analysis. We show consistent variation in syllable duration patterns in selected speech varieties in English, Chinese and Polish, chosen for their known different prosodic typological properties. Results include a possible limen of 50ms for relevant timing patterns. For data-mining we use the *Time Group Analysis* (*TGA*) methodology, directly in the *TGA* online tool and integrated into the *Annotation Pro+TGA* desktop software.

Index Terms: prosody, syllable duration, speech style, register, dialect, annotation mining, English, Chinese, Polish

1. Introduction: domain and methods

Inter-variety differences in speech duration patterning have been studied mainly in the context of prosodic typology and native-foreign pronunciation. The present sociophonetic contribution addresses the issue of intra-language variation in syllable durations at the phonology-phonetics interface, in pilot case studies of different registers or speech styles in the same dialect (English, Polish) and different dialects in the same register (Mandarin) using the computational annotation-mining paradigm [1], [2]. We concentrate on syllable duration patterns in interpausal time groups. Annotation practice in this field has been criticised for lack of a precisely specified empirical basis [3], [4], so pause and syllable annotation criteria require comment.

Pauses are typically defined acoustically with a minimal duration criterion such as 100, 150, or 200 ms ([5], [6], [7], [8], [9], [11], [12], [13], [14], [15], [16]): auditorily by holistic annotator perception (whether actual silence, or associated with final lengthening and other item-final features), or functionally (with syntactic boundary, hesitation). The annotations used here are grounded in a heuristic combination of acoustic, visualised and auditory criteria, with actual acoustic pause lengths sometimes much less than the commonly proposed minimum of 100 ms. Explicit functional criteria were not used, in order to avoid circularity in later studies of the relation between interpausal groups and grammatical constituents. The segmental content of so-called 'filled pauses' was not treated as a pause.

Syllable annotation is based on more language-dependent criteria than for pauses. The initial criterion of word boundary as syllable boundary is relatively straightforward for Mandarin and also for English. For Polish the criterion is more complex (cf. proclitics): a modified *Maximal Onset Principle* was used, with two constraints on onset structure: non-decreasing sonority, and attested actual occurrence as word onset [17]. Word-internally, ambisyllabic consonants were annotated as onsets of the following syllable in English and Polish; the issue does not arise in Mandarin phonotactics. We are not concerned with syllable-internal boundaries.

The literature reveals many techniques for measuring properties of speech duration patterns, including acoustic models of rhythm as oscillation, and annotation mining with partial models of rhythm types as degrees of isochrony, models of acceleration and deceleration, and of duration hierarchies and their relation to word and phrase hierarchies. We focus on annotation mining (cf. Section 2). Section 3 presents the results of the three pilot case studies on English, Mandarin and Polish, and Section 4 contains a summary and conclusion, and outlines future research and applications.

2. Annotation data-mining

We define speech annotation data-mining as the extraction of structured information from speech annotations, and use computational annotation mining tools for efficiency, consistency and handling large corpora: the *Time Group Analysis* (*TGA*) online tool [18] and *TGA* functions integrated into *Annotation Pro* [19], [20]. We apply the tools to reliable statistical distributional analysis of syllable duration relations and patterns, many of these properties being relatively inaccessible to manual approaches, at least for large data sets.

2.1. Annotations

Annotations are in general modelled as two-dimensional constructs structured as parallel symbolic information streams (tiers, layers) of event tags. Event tags are represented as label-interval pairs, where the intervals Δt can be represented in a number of ways: (1) as single time-stamps t for the event start or end with a second time-stamp implicitly provided by an adjacent event tag (ESPS, HTK, BOSS); (2) time information pairs: (a) event beginning and end time-stamps t_1, t_2 (Praat, Transcriber, WaveSurfer, ELAN, Anvil), or (b) event beginning time-stamp t_1 and duration Δt for $\Delta t = t_2 - t_1$ (Annotation Pro [20]). Time-stamps are typically represented (1) as sample numbers (with sample-rate stored in the annotation file metadata for conversion to time values),

(2) as clock time. Rarely, time-stamp triples may be defined for event beginning, centre (or peak) and end (SAM).

The parallel symbolic information streams offer two levels of complexity in annotation data-mining: *intra-stream relations* of sequence and hierarchy in single streams, and *inter-stream relations* of overlap or synchronisation [21], with hierarchy as a special case of overlap. Formal models from graph theory [22], event logic [23], automata theory [24] and interval calculus [25] are available for representing and computing with annotations.

Two kinds of intra-stream information can typically be mined from annotations: (1) distributional properties of label sequences and intervals (cf. Section 3.1); (2) interval duration I=Δt and duration difference relations ΔI=ΔΔt such as interval duration dispersions (or inversely: regularity or isochrony[1]), and interval acceleration and deceleration slopes. We focus on interval duration distributions and duration patterns in interpausal syllable sequences.

2.2. Duration dispersion or isochrony

Measures of duration dispersion (or its inverse, relative isochrony) which have been used for various phonetic units include standard deviation and the models shown in Table 1: *Pairwise Irregularity Measure*, *PIM*; *Pairwise Foot Difference*, *PFD*; *raw* and *normalised Pairwise Variability Index*, *rPVI* and *nPVI*; cf. references and discussion in [1]. *PIM* is a ratio model, *PFD* is a simplified variance model. The *nPVI* is a difference limen model 100*ΔI/(I/2). (rather than the usual ΔI/I, yielding an asymptote of 200 for the *nPVI*, not the usual 1.0), and eliminates rate change effects by comparing neighbours, not all intervals.

Table 1: *Definitions of PIM, PFD, PVI measures.*

$PIM(I_{1,\dots n})$ =	$\sum_{i\neq j} \lvert \log \frac{I_i}{I_j} \rvert$
$PFD(foot_{1\dots n})$ =	$\frac{100\times\sum \lvert MFL - len(foot_i)\rvert}{len(foot_{1\dots n})}$ where MFL = ‘mean foot length’
$rPVI(d_{1\dots m})$ =	$\sum_{k=1}^{m-1} \lvert d_k - d_{k+1}\rvert/(m-1)$
$nPVI(d_{1\dots m})$ =	$100\times\sum_{k=1}^{m-1} \lvert \frac{d_k - d_{k+1}}{(d_k+d_{k+1})/2}\rvert/(m-1)$

The models are typically used for specific label types (foot, vocalic and consonantal intervals, syllables), but the formulae are neutral in this respect and may be used for any intervals. The models are not equivalent: Figure 1 shows correlations between the measures for utterances of 5 speakers of Brazilian Portuguese [1]; there is considerable inter-speaker variation in the correlations, and while *corr*(*SD*,*PFD*) is predictably high, *corr*(*PFD*,*nPVI*) and *corr*(*SD*,*nPVI*) are lower, though similar to each other. In general, *PIM* does not relate well to the other measures. The models have been called ‘rhythm metrics’, but they only fulfil one necessary rhythm condition of *relative* (‘*fuzzy*’, ‘*sloppy*’) *isochrony*. They fail on the equally necessary condition of *rhythmic alternation*, since the use of absolute (unsigned) values does not distinguish between negative and positive duration changes [1], [2]. The *SD*, *PFD* and *nPVI* measures (though not *PIM*) are still useful models of relative isochrony (regularity, ‘smoothness’, ‘evenness’) of intervals, however.

Like the formulae, the annotation data-mining approach itself, shown in the pilot case studies in Section 3, is domain-neutral, in that it may be applied to annotations of any segments in speech, to annotations of visual head, hand and postural gesture streams, to both of these combined (or indeed to any comparable empirical time-function). We focus on syllables.

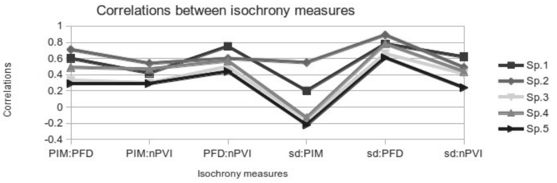

Figure 1: *Correlations between so-called ‘rhythm metrics’ for 5 speakers of Brazilian Portuguese.*

While the basic manual annotation mining studies noted above have been discredited as rhythm measures and have been conceptually overtaken by oscillator models [26], they have proved their worth as irregularity measures. We go a step further in using computational annotation mining for efficiency, consistency and data quantity, and for processing additional complex empirical parameters.

3. Varietal duration patterns: case studies

Contrary to *nPVI* studies, which ignore speech rate and filter out speech rate change, the case studies on English and Polish speech styles focus on acceleration, deceleration and, for Polish, also speech rate. The Mandarin study investigates relations between duration patterns and grammatical items in Beijing and Hangzhou Mandarin. At this stage, the three studies are designed to show the potential of computational annotation mining techniques with typologically different languages applied to relatively large data sets, rather than to pursue typological studies, because language variety corpora of adequate size are not readily available.

3.1. Case study 1: British English genres

The annotation data for pilot studies of annotation mining techniques with British English are taken from a subset of the Aix-MARSEC [27] database of radio speech, covering a range of sub-genres: *A* (*Commentary*), *B* (*News broadcast*), *C* (*Lecture aimed at general audience*), *D* (*Lecture aimed at restricted audience*), *E* (*Religious broadcast including liturgy*), *F* (*Magazine-style reporting*), *G* (*Fiction*), *H* (*Poetry*), *J* (*Dialogue*), *K* (*Propaganda*) and *M* (*Miscellaneous*). The genres *A*, *B*, *C*, *F* and *K* were included in this study due to their relatively similar discourse types. The Aix-MARSEC repository also contains annotation data-mining tools, but not for parameters investigated here.

Figure 2 shows averages of several metrics for the genre data (values are scaled to permit visualisation in the same graph). Except for genres *F* and *K*, values of the measures are consistently very similar, even though the speakers in each case are different and in some cases several speakers per genre are present in the corpus.

Genres *F* and *K* are outliers with regard to slope. The explanation may lie in a difference in discourse functions: genres *A*, *B*, *C* and *D* are typically formal read-aloud or rehearsed genres, while *F* is associated with more spontaneous speech, and *K*, whether read or not, would be expected to contain persuasion oriented rhetorical prosodic

[1] Organisation of an event sequence into equal time intervals; in data transmission engineering (sometimes incorrectly spelled 'isochryny') a particular kind of synchronisation.

features, including syllable lengthenings. Higher positive slope values mean increasing average duration, i.e. speech rate deceleration in interpausal units, contrasting with more constant speech rate in the read-aloud genres. Slope may thus be a useful discourse type marker, along with other prosodic parameters which were not represented in the annotations.

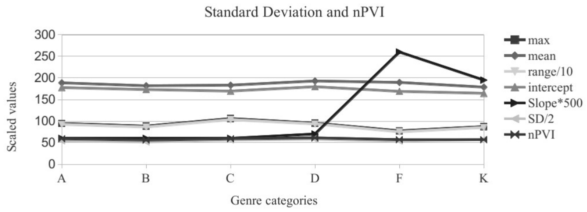

Figure 2: *Scaled annotation mining measures of six sub-genres of British radio speech.*

In addition to the analysis of duration dispersions, the *positive and negative polarities of duration differences* between neighbouring syllables were represented as *tokens*, retaining alternation (unlike the dispersion metrics), and *token sequence* distributions were registered at different *duration difference thresholds*. This technique is a first approximation to identifying actual rhythmic alternation independently of oscillator models, and in contrast to the dispersion measures shown in Table 1.

The token sequences of Table 2 (from the first utterance in the Aix-MARSEC database) show a high proportion of alternations at difference thresholds below about 50 ms; above 50 ms the difference threshold overrides many smaller alternation differences. Whether this transition at 50 ms is perceptually or functionally relevant needs more study.

Table 2: *Ranks of duration change n-grams* (2≤*n*≤5) *at thresholds 0...60* (\: *increase; /: decrease; =: same*); +, #: *word boundaries and pauses*.

Thr = 0		Thr = 20		Thr = 40		Thr = 60	
% (*n*)	**Seq**	**% (*n*)**	**Seq**	**% (*n*)**	**Seq**	**% (*n*)**	**Seq**
24 (65)	/\	20 (55)	/\	15 (41)	/\	17 (46)	==
23 (61)	\/	18 (48)	\/	13 (34)	\/	11 (29)	=\
13 (36)	\\	9 (24)	\#	9 (24)	\=	10 (26)	/=
17 (39)	\/\	13 (31)	\/\	9 (21)	\/\	8 (2)	===
13 (31)	/\/	10 (23)	/\/	7 (17)	/\/	6 (13)	==\
9 (21)	/\\	6 (13)	/\\	5 (11)	=/\	5 (12)	\/=
10 (20)	\/\/	7 (14)	\/\/	5 (10)	/\/\	4 (8)	====
9 (18)	/\/\	7 (14)	/\/\	4 (9)	\/\/	3 (7)	===\
5 (11)	\/\\	4 (8)	=\/\	3 (7)	=\/\	3 (7)	==/\
6 (10)	\/\/\	5 (9)	\/\/\	4 (6)	\/\/\	3 (5)	==/\/
5 (9)	/\/\/	4 (7)	/\/\/	3 (5)	\=/=\	3 (5)	+====

Preliminary studies show that a similar transition at around the 50 ms duration difference threshold can also be found in other languages and language varieties (cf. Section 3.2). However no hard and fast evidence can be given at this time. If this threshold transition at about 50 ms turns out to be generally valid, the result casts doubt on the validity of the raw duration data of previous duration dispersion studies.

3.2. Case study 2: Chinese regional accent

An issue which has not received detailed empirical attention in recent years is the relation between timing in syllable sequences and grammatical units such as words and phrases. A pilot annotation mining experiment was undertaken with recordings of 6 speakers (3 from the Hangzhou area and 3 from Beijing) reading a Mandarin Chinese translation of the IPA standard text 'The North Wind and the Sun' from the CASS corpus [28], [29]. *Time Tree* relations [21], [30] between syllable relations in interpausal groups, and words (one or more characters/syllables) were investigated.

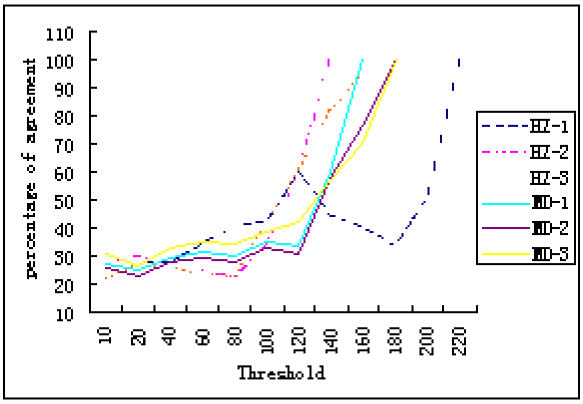

Figure 3: *Relations between duration-based syllable groupings and words for speakers of Beijing and Hangzhou varieties of Mandarin Chinese.*

The constituents were induced automatically from long-short duration patterns, where shorter constituents are prepended to longer constituents with a recursive quasi-iambic (*weak-strong*) *Time Tree* algorithm setting:

1. A syllable is a constituent.
2. A shorter constituent prepended to a longer following neighbour constituent is a constituent.
3. Nothing else is a constituent.

The algorithm was applied with all integer ms thresholds for duration differences from 0 ms to around 200 ms, where the correspondence ratio starts dropping. The following example shows a quasi-iambic *Time Tree* (represented as bracketing) of the Mandarin utterance "zhe4 shi1hou5, lu4 shang5 lai2 le5 ge4 zou3 daor4 de5" (*at that time, on the street came a traveller*), and a grammatical bracketing:

Quasi-iambic Time Tree: (((zhe4 (shi2 hou5)) (((lu4 shang5) (lai2 (le5 (ge4 zou3)))) daor4)) (de5 PAUSE))

Grammatical bracketing: ((zhe (shi hou)), (lu shang) ((lai) (le) (ge) (zou daor de)))

The groups (shi2 hou5) and (lu4 shang5) correspond to words; (ge4 zou3) is not a grammatical constituent. Also, factoring out the effect of the pause, (lai2 le5 ge4 zou3 daor4 de5) corresponds to a grammatical constituent. The correspondences between the syllable groupings and words are shown in Figure 3 (cf. also [31]).

Below a duration difference threshold of about 50 ms, correspondences between syllable groups and words are low, and are comparable among speakers. Correspondences gradually increase, and begin to diverge until about 100 ms, where they rapidly increase and interesting patterns emerge. Correspondences for Beijing Mandarin remain similar as thresholds move beyond 50 ms, while for the Hangzhou variety they are more diverse, as would be expected in a comparison between a standard accent (Beijing Mandarin) and a non-standard regional accent (Hangzhou Mandarin).

Whether the threshold limit of 50 ms is related to the limit found for English at a similar threshold order of magnitude (Section 3.1) needs further study.

3.3. Case study 3: Polish speech styles

In order to analyse syllable durations in Polish, recordings of read speech and dialogues from the Paralingua corpus [32] were used. The aim of the analysis was to look at timing patterns in speech recordings of 20 speakers in three stages of a recording session: (A) read speech produced at the very beginning of the session; (B) telephone conversations (task-oriented dialogues over the telephone); (C) read speech produced at the very end of the session after participating in a dialogue with time constraints imposed.

The time constraints were imposed only in the dialogue part of the experiment while for the final reading there were no time limits and the instruction was exactly the same as with session-initial reading (i.e. in both cases the speakers were requested to read the text in their normal, habitual way). Segmentation into interpausal time groups assumed minimal significant pause duration to be about 100 ms (cf. e.g. [5] for German). However, in some cases the minimal value of only ca. 50 ms was used, based on auditory perception and visual inspection of spectrograms by two experienced annotators. The study investigates the question whether consistent influence of the recording procedure on syllable timing could be associated with durational variability between time groups (see also [33]).

The most significant differences between the three types of speech were observed for slope, as with the English genres (Section 3.1). The overall mean values are included in Table 3. The overall mean of slopes for the dialogue recordings was significantly higher than for read speech of both types. Also, the overall means of intercepts and *nPVI* measures appeared to be highest for conversational speech.

Table 3: *Overall means of duration difference slope, intercept and nPVI for recording stages A, B and C.*

	Slope	Intercept	nPVI
A. Read 1	0,0925	145,05	43,83
B. Dialogue	0,2121	177,00	48,28
C. Read 2	0,0829	145,94	42,50

More detailed information on the individual differences between speakers in the three recording session stages can be found in Figure 4: the plots show the variability of mean slopes, intercepts and *nPVI* values for each of the 20 speakers. As shown in the figure, only in case of three speakers (23, 24, 29) was the mean slope close to the values for read speech while all the remaining speakers differentiated their slopes between read speech and dialogue.

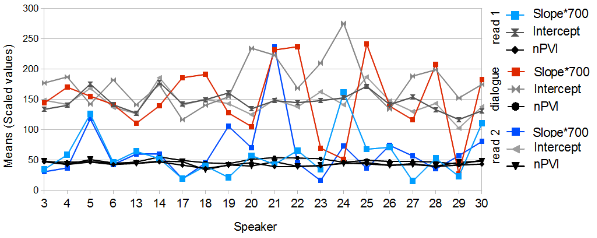

Figure 4: *The variability of selected timing properties in three speaking styles for 20 speakers of Polish.*

The two read speech tasks were very similar as regards the overall values. However, when individual results for particular speakers were investigated it was observed that the values for the initial reading tended to vary more among speakers than those for the final reading. This might suggest that after participating in the preceding tasks (20-30 minutes altogether including the inital reading and three dialogue tasks), inter-speaker differences in speech acceleration or deceleration tended to be less significant than in the earlier stages of the recording session, especially in the dialogues.

Also, speech rate (syll/sec) was measured for each recording stage. The observed overall mean value was slightly higher for the final reading (5.5. syll/sec) than for the two preceding parts (5.3, 5.4 syll/sec, respectively), but the differences were not statistically significant. Individual rate differences between the two reading tasks were mostly negligible, being exactly or almost the same for most speakers. The majority of speakers (except 13 and 17) clearly differentiated read and spontaneous tempi, but using faster or slower rate for a particular type of speech appeared to be individual rather than style-dependent. Overall rate means were in line with values observed for normal reading rate [33] and for dialogues [34] in Polish; [35] reported higher means around 6.9 syll/sec for Polish dialogues, possibly due to the different data type (fluent and coherent utterances with no unintelligible parts, false starts or hesitation sounds).

When comparing these observations with the results of slope variability measurements it was found that the two speakers whose mean rates were the same in both read and spontaneous speech still exhibited different patterns for acceleration-deceleration, as represented by the mean slope values for the two speech styles.

4. Conclusions

We have shown how new computational annotation-mining procedures can be deployed to examine a variety of interesting speech duration parameters in the sociophonetic context of speech genre, style and regional accent variation in typologically different languages. Despite the typological differences of phonology and morphology, the languages showed similarities: in a duration difference (Δt) threshold transition around 50 ms emerged (in different English and Mandarin contexts; not investigated for Polish), and in duration difference slope (English and Polish; not investigated for Mandarin). While the corpora used in the present studies were much larger than the small corpora used in previous manual annotation mining studies, in order to exploit computational annotation mining techniques fully and to move to machine learning techniques, larger annotated corpora for more languages are needed.

An interesting topic for future work is the minimal duration difference of 50 ms found in our production data for English and Mandarin. Present results do not yet permit the formulation or confirmation of relevant difference limen models of the $\Delta I/I$ type, where I is the average time-stamp difference based duration interval Δt, and ΔI is the average interval difference $\Delta\Delta t$.

We foresee applications of our computational annotation mining techniques in foreign language learning and testing studies, in modelling interfaces between phonetics in studies of phonology, prosody, grammar and discourse structure, and in evaluating naturalness in speech synthesis [36].

5. Acknowledgments

Creation of the Paralingua corpus was supported from the Polish financial resources for science in the years 2010-2012 as a development project (O R00 0170 12). The Mandarin corpus is used by kind permission of the Chinese Academy for Social Sciences. The TGA online tool is currently hosted by the Bielefeld University.

6. References

[1] Gibbon, D. and Fernandes, F. R., "Annotation-Mining for Rhythm Model Comparison in Brazilian Portuguese", Proc. Interspeech 2005, 3289-3292, 2005.

[2] Trippel, T., Gibbon, D. and Fernandes, F. R., "A BLARK extension for temporal annotation mining", Proc. LREC 2006, Genoa, 2006.

[3] Gut, U., "Rhythm in L2 speech", in Gibbon, D., Hirst, D., Campbell, N. [Eds], Rhythm, Melody and Harmony in Speech. Speech and Language Technology: Studies in Honour of Wiktor Jassem, 14/15:83–94, 2011/2012.

[4] Arvaniti, A., "The usefulness of metrics in the quantification of speech rhythm", Phonetica 66:46-63, 2009.

[5] Butcher, A., "Aspects of the speech pause: phonetic correlates and communicative function", Arbeitsberichte 15, Institut für Phonetik, Universität Kiel, 1981.

[6] Cruttenden, A., Intonation, Cambridge University Press, 1986.

[7] Dankovicova, J., "The minimum pause duration in spontaneous speech", PROPH – Progress Reports from Oxford Phonetics 5, 17–24, 1992.

[8] Dankovicova, J., Pigott, K., Wells, B., Pepp, S., "Temporal markers of prosodic boundaries in childrens speech production", Journal of the International Phonetic Association, 34 (1), 17-36, 2004.

[9] Duez, D., "Perception of silent pauses in continuous speech", Language and Speech 28.4 (377-389), 1985.

[10] Heldner, M., and Edlund, J., "Pauses, gaps and overlaps in conversations" Journal of Phonetics 38.4: 555-568, 2010.

[11] Hieke, A. E., Kowal, S and O'Connell, D. C., "The trouble with 'articulatory' pauses", Language and Speech 26.3: 203-214, 1983.

[12] Klatt, D. H. & Cooper, W. E., "Perception of segments duration in sentence contexts", in Cohen, A. & Nooteboom, S. [Eds], Structure and Process in Speech Perception, 69–89, Heidelberg: SpringerVerlag, 1975.

[13] Lehiste, I., Suprasegmentals, M.I.T. Press, Cambridge MA, 1970.

[14] Makashay, M. J., Individual differences in speech and non-speech perception of frequency and duration. PhD dissertation, Ohio State University, 2003.

[15] Megyesi, B. and Gustafson-Capkova, S., "Production and perception of pauses and their linguistic context in read and spontaneous speech in Swedish", Proc. Interspeech, 2002.

[16] Zvonik, E. and Cummins, F., "The effect of surrounding phrase lengths on pause duration", Proc. Interspeech, 2003.

[17] Demenko, G., Klessa, K., Szymański, M., Breuer, S. and Hess, W., "Polish unit selection speech synthesis with BOSS: extensions and speech corpora", in International Journal of Speech Technology, Volume 13 (2), 85-99, 2010.

[18] Gibbon, D., "TGA: a web tool for Time Group Analysis". Proc. Tools and Resources for the Analysis of Speech Prosody (TRASP) Workshop, Aix en Provence, 2013. (Cf. ref. there to online TGA tool: http://wwwhomes.uni-bielefeld.de/gibbon/TGA/)

[19] Klessa, K. and Gibbon, D., "Annotation Pro + TGA: automation of speech timing analysis". Proceedings of Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland, 26-31 May 2014.

[20] Klessa, K., Karpiński, M. and Wagner, A., "Annotation Pro - a new software tool for annotation of linguistic and paralinguistic features", Proc. Tools and Resources for the Analysis of Speech Prosody (TRASP) Workshop, Aix en Provence, 2013.

[21] Gibbon, D., "Time Types and Time Trees: Prosodic Mining and Alignment of Temporally Annotated Data." in Sudhoff, S., Lenertová, D., Meyer, R., Pappert, S., Augurzky, P., Mleinek, I., Richter, N. and Schließer, J. [Eds], Methods in Empirical Prosody Research. Walter de Gruyter, 281–209, 2006.

[22] Bird, S. and Liberman, M., A Formal Framework for Linguistic Annotation. Technical Report MS-CIS-99-01, Linguistic Data Consortium, University of Pennsylvania, 1999.

[23] Bird, S. and Klein, E., "Phonological Events", Journal of Linguistics 26, 33–56, 1990 .

[24] Carson-Berndsen, J., Time Map Phonology: Finite State Models and Event Logics in Speech Recognition, Springer, 1997.

[25] Allen, J. F., "Maintaining knowledge about temporal intervals", in Communications of the ACM, 26 November 1983.

[26] Inden, B., Malisz, Z., Wagner, P., and Wachsmuth, I., "Rapid entrainment to spontaneous speech: A comparison of oscillator models", in Miyake, N., Peebles, D. and Cooper, R. P. [Eds], Proceedings of the 34th Annual Conference of the Cognitive Science Society, Austin, TX, Cognitive Science Society, 2012.

[27] Auran, C., Bouzon, C. & Hirst, D. J., "The Aix-MARSEC project: an evolutive database of spoken English", in Bel, B. and Marlien, I. [Eds], Proceedings of the Second International Conference on Speech Prosody, Nara, Japan, 561-564, 2004.

[28] Li A., Zheng, F., Byrne, W., Fung, P., Kamm, T., Liu, Y., Song, Z., Ruhi, U., Venkataramani, V. and Chen, X., "CASS: A phonetically transcribed corpus of Mandarin spontaneous speech", in Proc. Interspeech 2000, 485-488, Beijing, 2000.

[29] Yu, J., "Timing analysis with the help of SPPAS and TGA tools". Proc. Tools and Resources for the Analysis of Speech Prosody (TRASP) Workshop, Aix en Provence, 2013.

[30] Gibbon, D., "Corpus-based syntax-prosody tree matching". in Proc. Eurospeech, Geneva, 2003.

[31] Yu, J. and Gibbon, D., "Criteria for database and tool design for speech timing analysis with special reference to Mandarin", in Proc. O-COCOSDA 2012, 41-46, Macau, 2012.

[32] Klessa, K., Wagner, A., Oleśkowicz-Popiel, M., Karpiński, M., "*Paralingua* – a new speech corpus for the studies of paralinguistic features", in Vargas-Sierra, Ch. (Ed), Corpus Resources for Descriptive and Applied Studies. Current Challenges and Future Directions: Selected Papers from the 5th Int. Conf. on Corpus Linguistics (CILC2013), Procedia – Social and Behavioral Science 95. (48-58), 2013.

[33] Gibbon, D., Klessa, K., and Bachan, J., "Duration and speed in speech events", in Mikołajczak-Matyja, N., Karpiński, M. [Eds], Studies in Phonetics and Psycholinguistics. A Festschrift for Prof. Piotra Łobacz, Poznań, to appear 2013.

[34] Karpiński, M., Klessa, K., Czoska, A., "Local and global alignment in the temporal domain in Polish task-oriented dialogue". Proceedings of 7th Speech Prosody Conference, Dublin, Ireland, 20-23 May 2014.

[35] Malisz, Z., Speech rhythm variability in Polish and English: a study of interaction between rhythmic levels. PhD Thesis, Faculty of English, Adam Mickiewicz University, 2013.

[36] Gibbon, D., Moore, R. and Winski, R., [Eds], Handbook of Standards and Resources for Spoken Language Systems, Mouton de Gruyter, 1997.