TIME GROUP ANALYZER: METHODOLOGY AND IMPLEMENTATION

Dafydd Gibbon[1], and Jue Yu[2]
[1]Fakultät für Linguistik und Literaturwissenschaft, Universität Bielefeld
Bielefeld, Germany
[2]School of Foreign Languages, Tongji University
Shanghai, China
e-mail: gibbon@uni-bielefeld.de, erinyu@126.com

Abstract

The *TGA* (*Time Group Analyser*) tool provides efficient ubiquitous web-based time-saving computational support for phoneticians without computational skills or facilities who are interested in selected linguistic phonetic aspects of speech timing. The input module extracts a specified tier (e.g. phone, syllable, foot) from a single annotation file in the common Praat TextGrid and CSV formats; user-defined settings permit selection of sub-sequences such as inter-pausal groups, and thresholds for minimum duration differences. Several types of output are provided: (1) Tabular outputs with descriptive statistics (including dispersion models like *standard deviation*, *PIM*, *PFD*, *nPVI*, *rPVI*), *linear regression*; (2) novel visual information about duration patterns, including difference *n-grams* and *Time Trees* (temporal parse trees); (3) graphs of duration relations, including Wagner Quadrant graphs. Examples of applications in phonetics are taken from published studies of varieties of Mandarin and English as a form of functional field evaluation of the tool. Other disciplines in which duration analysis has practical uses, such as forensic phonetics, clinical linguistics, dialectometry, speech genre stylometry and language acquisition, will also benefit from the efficient methodology provided by the TGA.

Keywords: online tools, speech timing, speech prosody, annotation processing, duration, time trees

1 Problems, methods, tools, solutions

1.1 Background and overview

Scientific methods are recipes for creating solutions to problems, and the tools used within these methods are the utensils which are used to implement these recipes. The tools themselves embody further methods: for example, in phonetics and speech technology, a descriptive and modelling methodology use annotations, i.e. the pairing of sections of a transcription with sections of a speech signal by means of time-stamps. The annotation procedure requires further methodological assumptions: first, on the categorial perception of speech (in creating the transcription), and second, on the physical parameters of digitized speech signals (in assigning time stamps which point to boundaries or peak points in the recorded signal).

In the history of phonetics, annotation methods have progressed from the traditional 'impressionistic' transcription of perceived sounds in an observed utterance, through recordings using various techniques and manipulation of these recordings. Before the advent of PCs such methods were common outside specialized phonetics labs, and until recently were common among phoneticians in less affluent regions. The concept of speech signal annotation arose with the speech technologies in the 1970s and software such as esps/Waves appeared in the 1980s, for the purpose of searching for, identifying and classifying portions of the speech signal in order to develop speech recognition and synthesis models. The technique of annotation was largely unknown outside of this field until free and public domain software with graphical user interfaces, such as Praat, Transcriber, Wavesurfer became available, starting in the 1990s. Newer annotation software designs with additional analysis facilities are still appearing in the interests of increased functionality and efficiency (e.g. in this volume: Annotation Pro, with facilities for perception experiments, and SPPAS, with automatic annotation based on dictionaries and statistical segment models). Annotation software supports the annotation process (1) by providing measurements and visualizations of various models of the speech signal, such as amplitude and energy envelopes, spectrum, fundamental frequency, and (2) visualizations of the mapping of arbitrarily many layers (tiers) of transcriptions and linguistic categories to segments of the speech signal. Some of these annotation tools such as Praat provide scripting languages which support the automation of particular measurement and analysis procedures. Some of the tools contain functions for exporting data and results in formats suitable for further analysis by means of other software such as spreadsheets or using modern programming languages such as Python or the statistical programming language R. An intermediate stage is represented by tools with scripting languages (e.g. the Praat scripting language) which can capture typical 'recipes' of analysis sequences, record them as scripts, and execute the scripts to analyse speech recordings automatically.

Although programming techniques are well known and widely used in specialized phonetics labs and research departments, there are still many phoneticians world-wide who use the phonetic tools such as Praat for manual numerical analysis, but lack programming skills or helpers, and are not familiar with the technique of annotation and annotation based analysis. Consequently, 'low tech' methods, for example copying on-screen values of signal parameters, such as temporal information, to spreadsheets for further calculation, are still very widespread.

The TGA online[1] 'multitool' described in this contribution is a little different, and is intended to fill a gap for the 'ordinary working phonetician' who is interested in aspects of speech timing and has no or little experience of programming. The TGA tool is is a 'multitool' in the sense that it puts together a broad set of procedures for analysing speech timing, some well known and some new, and produces a variety of

[1] Current URL: http://wwwhomes.uni-bielefeld.de/gibbon/TGA/

analyses of timing relations in a single 'one-click' tool. An offline prototype of the TGA tool exists for handling larger amounts of data.[2]

The TGA tool itself, combining previous separate tools, was originally developed for small projects and for phonetics teaching, in a cooperation between the authors of the present contribution for the description of timing in Mandarin, in dialects of Chinese, and in Chinese English (Yu & Gibbon, 2012, Yu et al., 2014, 2015; Gibbon, 2013; Yu, 2013). The TGA tool in its current online form is designed for the analysis of timing relations in single annotation tiers from single annotation files. Timing analyses across more than one tier are not incorporated in the present version; if such analyses are required the separate results must be exported, combined and further analysed with a spreadsheet or other software application.

The online TGA user interface design is kept very simple: an annotation file is opened in a text editor, copied and pasted into an HTML form on a web page. Parameter settings permit the selection of the relevant tier name and values of parameters for the analysis, and a 'one-click' timing analysis takes place, using a range of analysis procedures and based on the time-stamps in the data, and producing a wide variety of outputs (see Section 3). In addition to common measures such as speech rate and variability, similarity or dispersion of duration values (e.g. *standard deviation*, *nPVI*, described in Section 2), novel measures and displays of acceleration, visualization of regularities by bar charts, time function plots and scatter plots are included, as well as chracter separated value (CSV) outputs for further analysis with statistical tools. A preliminary version of the TGA has been previously described (Gibbon, 2013). Components of the TGA tool have been incorporated in software by other developers (AnnotationPro and SPPAS, this volume). Typically, TGA applications have been applied to the syllable tier, but the duration of intervals on any tier in an annotation can be selected and analysed.

The objective of this contribution is to provide an account of TGA tool development strategy from problem domain through specifications to implementation. It is not primarily a manual for how to use the tool for a specific phonetic analysis purpose, though application examples are given in Section 3.4.

The organization of the contribution follows a general scheme covering problems, methods, tools, solutions, roughly according to a traditional software development procedure of requirements specification, design, implementation and evaluation. The following subsection 1.2 delimits and characterizes a selection of problems in syllable duration analysis. Section 2 deals with a set of linguistic phonetic methods which have been proposed for solving the problems, and with new methods for new aspects of the domain. In Section 3 specification, design, implementation and phonetic applications of the TGA tool are described, as well as its application in selected publications on

[2] There is an offline development prototype capable of handling larger amounts of annotation data and with additional functions which are not available online owing to server limitations. The offline prototype is not yet available for general distribution.

timing problems. Section 4 concludes the description by outlining areas which have been addressed with the TGA, and by addressing planned extensions and noting practical application potential in neighbouring disciplines.

1.2 Aspects of speech timing: delimiting the TGA domain

The domain of issues handled by the TGA tool is characterized in Section 2 The aim of this subsection is simply to delimit this domain very briefly in the context of a broader range of issues in subsegmental, segmental and suprasegmental or prosodic speech timing, ranging from voice onset time and stop closure-opening time through vowel, consonant and syllable reduction, speech rate and rhythm through pause patterning to timing in discourse. Figure 1 compactly summarizes the rank-interpretation hierarchy of the language structures, functions and phonetic correlates involved. The TGA tool focuses on sequential and hierarchical relations within sequences of units such as phones, syllables, words (depending on the annotation tier selected). The TGA tool is in principle suited to analysis of units at any level of the rank-interpretation hierarchy shown in Figure 1, but has so far been mainly restricted to analyzing temporal relations between syllables in Time Groups of two kinds: (1) interpausal time groups, and (2) time groups based on acceleration and deceleration of speech rate (e.g. syllable rate).

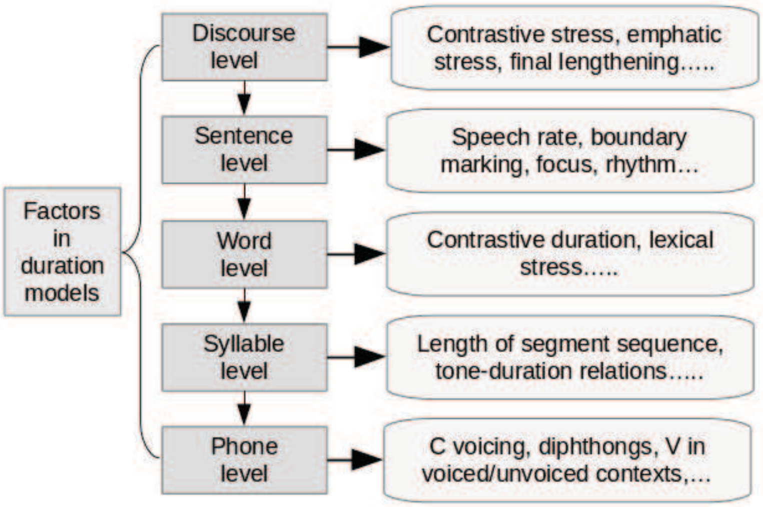

Factors in duration models
Discourse level
Sentence level
Word level
Syllable level
Phone level
Contrastive stress, emphatic stress, final lengthening.....
Speech rate, boundary marking, focus, rhythm...
Contrastive duration, lexical stress.....
Length of segment sequence, tone-duration relations.....
C voicing, diphthongs, V in voiced/unvoiced contexts,...

Figure 1: Domains of speech timing patterns

One of the areas of deployment of the TGA tool is in the study of aspects of speech rhythm, an area which has been conspicuous in the phonetic literature since the study of Pike (1945) on the intonation of American English. One of the questions involved has been whether and how the perception of rhythm in different languages tends towards two poles of *syllable timing* on the one hand, and *foot timing* (with related concepts such as *stress timing*, *interstress timing*) on the other. Searches for correlates of rhythm in the speech signal have been somewhat inconclusive (Arvaniti, 2009), motivating a view that rhythm is an epiphenomenon which cannot be simply induced from the temporal patterning of physical speech signals and which results from the interplay of many factors, including those outlined in Figure 1: discourse and grammatical structure, word familiarity and frequency, morphological structure and phonotactic patterns (Gibbon, 2006). The physical correlates in turn involve several parameters: the timing of units of speech, as well as pitch and intensity patterns. Nevertheless, the search is not over, and the function of the TGA tool is to support research specifically in relation to speech timing in matters including but not limited to rhythm.

There is currently no comprehensive theory of speech rhythm production and perception, and no model of rhythm patterns. An earlier pretheoretical clarification of the term ‘rhythm’ was summarized by Gibbon et al. (2001) as an iteration of alternations of strong and weak values of some parameter or parameter set, whose alternations which have a tendency to isochrony. The model may be termed a ‘Three Constraint Model’ of rhythm:

> Rhythm is the recurrence of a perceivable temporal patterning of strongly marked (focal) values and weakly marked (non-focal) values of some parameter as constituents of a tendentially constant temporal domain (environment).

This Three Constraint Model has turned out to be inadequate in a number of respects as it is missing the similarity and hierarchy properties of speech rhythm (Gibbon, 2003), and of rhythm in music and other domains. A ‘Five Constraint Model’ is more adequate, requiring fulfilment of the following criteria, which will figure in the description of the TGA tool:

1. a dynamic *Alternation Constraint* on patterns of stronger and weaker elements of some parameter or parameter set;
2. an oscillatory *Iteration Constraint* on repetition of adjacent patterns;
3. a qualitative *Similarity Constraint* on elements of the iterated adjacent patterns;
4. a quantitative *Isochrony Constraint* on the iterated adjacent patterns;
5. a structural *Hierarchy Constraint* on rhythm, which specifies temporal domains in a relation of temporal inclusion, to each of which the previous constraints apply (the temporally shortest alternation being the lowest and sometimes the only level in the hierarchy).

The basic strong-weak Alternation Constraint applies at different structural levels in different languages. Typical of tendencies to the ‘ideal type’ of syllable timing is the alternation consonants and vowels (CV, CVC patterns), and in the ‘ideal-type’ of stress timing is alternations between strong syllables and one or more short syllables. The ‘ideal-types’ are in practice only approximative tendencies; so-called stress-timed languages may also have fortuitous syllable timing: *Jim swam fast past Jane’s boat*, and vice versa.

Recent approaches (Cummins, 2009; Inden et al., 2013; Włodarczak, 2012) have addressed more complex issues of modelling rhythm by means of oscillators and of the mutual adaptation or entrainment of rhythms by interlocutors in discourse; this domain is outside the immediate scope of the present study, and inter-tier duration relations are currently not included in the specification of the TGA tool.

2 TGA prerequisites: approaches to prosodic timing

2.1 Phonological and linguistic phonetic approaches

The present section concentrates on the aspects of speech timing analysis for which support by the TGA tool is designed. Overviews of relevant methods of timing analysis at the level of syllable patterning are given by Gibbon (2006) and the contributions to Gibbon et al. (2012). These methods presuppose some prior identification of linguistic and phonetic categories in the form of segmentations and labellings of speech recordings, whether by annotation or direct measurement of signal visualizations. Many analysis methods have been applied to the problem of examining duration relations between consonantal and vocalic syllable constituents, or between syllables, or between stress-based feet. However, most have concentrated implicitly or explicitly solely on the *Iteration Constraint* and the *Isochrony Constraint* outlined in the Section 1.2, to the exclusion of the *Alternation Constraint*, the *Similarity Constraint* and the *Hierarchy Constraint*.

Timing hierarchies have been discussed in several different theoretical and methodological contexts: in post-generative phonologies such as Metrical Phonology (Goldsmith, 1990); as prosodic structure (Jassem, 1952; Abercrombie, 1967); as oscillation (Barbosa, 2002; Inden et al., 2012). In the present contribution, novel methods for modelling the *Alternation Constraint* and the *Iteration Constraint* and the *Hierarchy Constraint* as *Duration Difference Token* (*DDT*) sequences is presented in the present contribution, and the *Time Tree* (*TT*) method of timing hierarchy induction (Gibbon, 2003, 2006) is also discussed.

The comprehensive structural rhythm model which has been most extensively investigated phonetically is that of Jassem (1952) and Jassem et al. (1984), which invokes *alternation* (of stressed syllable and sequences of unstressed syllables), *iteration* (of stressed-unstressed alternations), *similarity* and *near-isochrony* (of stressed-unstressed sequences) and *hierarchy* (of broad and narrow rhythm units). The Abercrombie model addresses the same constraints but with a simpler structure and without the hierarchy constraint. The Jassem model and to some extent the Abercrombie model (1964:219) also take morphological structure (word boundaries) into account.

In Jassem's model, the Broad Rhythm Unit (BRU) has two constituents: an optional Anacrusis (ANA), consisting of unstressed syllables from a grammatical boundary (e.g. utterance, phrase, word boundary) up to but not including the next stressed syllable, and an obligatory Narrow Rhythm Unit (NRU), consisting of a stressed syllable followed optionally by a sequence of unstressed syllables, extending to the next relevant grammatical boundary. Thus, a neutral pronunciation of the sequence *it's stressful today* may yield the following parse:

(BRU: (ANA: it's) (NRU: <u>stress</u> ful)) (BRU: (ANA: to) (NRU: <u>day</u>))

However, the better known model is the simpler and flatter model of Abercrombie (1967), who analyses sequences feet, each consisting of an 'ictus' (a phonetically stressed syllable, which may be phonemically long, medium or short) and a 'remiss' (an optional sequence of unstressed syllables). Initial sequences of unstressed syllables are treated as having an empty ictus or null beat:

|| - it's | stress ful to | day ||

Jassem et al. (1984) have shown that the more complex Jassem model fits the English facts better than the simpler Abercrombie model: no empty beat is needed, and they showed experimentally that unstressed syllables in the Anacrusis have different timing properties from those in the Narrow Rhythm Unit (cf. also contributions to Gibbon et al., 2012 for extensive discussion).

The Jassem and Abercrombie models are both very close to the present Five Constraint model of rhythm in that they incorporate the Alternation, Iteration, Similarity, and Isochrony Constraints and (in the case of the Jassem model) also the Hierarchy Constraint. These Jassem and Abercrombie models and the Five Constraint model are not explicitly included in the domain of the TGA, but need to be borne in mind when using the TGA tool for analysing the relation between phonological and phonetic determinants of speech timing, particularly rhythm.

2.2 Linear quantitative models of duration dispersion

The inclusion of a selection of linear quantitative models of duration in the domain of the TGA tool requires explicit justification. Several studies of speech timing have concentrated on subsyllabic or syllabic properties, looking at the dispersion and percentatages of consonantal and vocalic stretches of the speech signal, for example variance or standard deviation of the durations of consonantal intervals (ΔC) and percentage of vowel durations (%V). Measurements based on the ΔC–%V model introduced by Ramus et al. (1999) yielded interesting results about the differentiation of different languages by means of the relation between these parameters and between these parameters and other dispersion measures such as vocalic normalized pairwise variation (*nPVI*) and consonantal raw pairwise variation (*rPVI*); cf. Low et al. (2000) and very many studies using these two measures. A selection of approaches of this type is shown in Table 1, including two already mentioned.

The top two models in Table 1 are the *Pairwise Irregularity Model* (*PIM*) of Scott et al. (1986) which sums all pairwise log ratios of each interval duration in the whole utterance, and the *Pairwise Foot Deviation* (*PIM*) model of Roach (1982), which takes adjacent pairwise differences rather than all pairwise differences, and is rather like standard deviation, except that the absolute magnitude of differences is taken, rather than the square and the square root. Although the Roach model refers to the foot as a unit, formally speaking the models are agnostic in regard to the units to which they apply.

The bottom two models, which have already been referred to, are variants of an *Average Magnitude Difference Function* (*AMDF*), in which differences in a moving window over pairs of adjacent intervals are averaged. This results in factoring out variations in speech rate, a useful innovation. In the context of speech timing analysis, the binary window AMDF is known as the *raw Pairwise Variability Index* (*rPVI*). The *rPVI* takes iteration and isochrony into account, but not alternation and hierarchy. The *rPVI* is normally applied to consonantal intervals in a speech recording, and is distinguished from the *normalized PVI*, *nPVI*, normally applied to vocalic intervals. The duration differences in the *nPVI* are normalized by dividing each difference by the average of the durations. The overall average is multiplied by 100 (as in the *PFD*), resulting in a scale for the *nPVI* from 0 (complete isochrony) to an asymptote of 200 (completely random).

Table 1: Four dispersion models of speech segment patterning.

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

The PVI variants have become very popular since their introduction by Low et al. (2000), have been used in very many studies and have yielded very interesting results about the dispersion of duration relations between different languages. In the literature there has been plenty of folklore and various simple misunderstandings about the nPVI formula[3]: (1) the component ‘n-1’ has been said to mean that the last syllable is not considered in order to factor out final lengthening, but this is false since the formula is about differences between adjacent items in a sequence, and there is always one difference less than the total number of items, and final lengthening is not factored out; (2) the factor ‘100’ has been said to convert the result to a percentage, but this is false since the nPVI scale is 0...200, because for normalization each duration difference is divided by the average duration of the pair (sum/2) and not the by sum, which would indeed have yielded 0…100.

Critics have also pointed out that (1) essentially the same results may be obtained from phonotactic patterns without phonetic measurements (Hirst 2009); (2) similar dispersions may occur between stylistic and dialectal varieties of the same language (Gut, 2012; Arvaniti, 2009); (3) in the PVI and PFD models the pairwise differences between adjacent syllables imply that rhythms are purely binary, for example with alternations of long and short syllables. This is not necessarily the case in stress-accent

[3] These will not be cited here in order to avoid embarrassment.

timed languages, however, where several unstressed syllables may intervene between stresses.

These measures have been (and often still are) called ‘rhythm metrics’, but this is a misnomer since, like plain standard deviation, none of these four measures fulfils either the Alternation, or the Iteration, or the Hierarchy Constraint; each concentrates only on a dispersion measure for relative isochrony. This is a fundamental formal criticism. With the first two models, ordering the values in any order, whether by in order of occurrence, or in increasing or decreasing or random order, yields the same dispersion values. With the PVI models it is also possible for different patterns to yield the same value, e.g. an alternating pattern like 2 4 2 4 yields the same value as 2 4 8 16, or 2 4 8 4, namely 66.6'. The reason for this oddity is the use of absolute magnitudes (the ‘|...|’ notation) with the result is that the direction of differences or ratios becomes irrelevant and therefore the Alternation Constraint is factored out.

Another fundamental criticism which applies to all four models, is that that the nearer the index is to zero, the more similar the timing pattern is to syllable (or foot, etc., depending on the unit being measured) timing. The further away from zero the index is, the less is known about what units are actually being measured, and the less one can be certain about whether it is a rhythm which is being measured (Gibbon, 2003). It is thus impossible to know what these results actually mean without combining them with further studies of units of different size, and taking the alternation, iteration and hierarchy constraints into accounts. The models account for a subset of the necessary conditions for rhythm, but do not provide a sufficient condition.

However, as measures of smoothness, regularity or relative isochrony relative to a unit such as a consonantal, vocalic, syllabic or foot interval the measures yield consistently useful results in demonstrating differences between languages. Examples of such analyses obtained with the TGA tool will be given in the case studies of applications in Section 3.4.

2.3 Dynamic timing factors: speed and acceleration

Values such as the minimum and maximum values of interval durations are subject to large fluctuations determined by the wide range of determining factors shown in Figure 1. However, useful further notions are connected with the speed of speech, usually measured in terms of the rate of phonemes, syllables, feet, stresses, phrases, etc. per second. The rate is the inverses of the mean duration; this, if the mean syllable length is 125 msec, the syllable rate is 8 syll/sec.

Another interesting parameter is the rate of change of speed, i.e. the overall acceleration or deceleration of a sequence of units such as the syllable, whether very locally with long-short syllable pairs or over an entire utterance. If measured over a long sequence, a useful measure is provided by linear regression models: the resulting slope indicates acceleration (if negative, i.e. with decreasing interval durations) and deceleration (if positive, i.e. with increasing interval durations).

These measures of speed and acceleration-deceleration are included in the TGA tool, and examples of the use of these measures are discussed in the application case studies in Section 3..4.

2.4 Patterns and relations: data visualization

An important part of scientific methodology, both with experimentation on small data sets and with inductive analytics applied to ‘big data’ is the visualization of data structures and distributions as a source of insights for explanations. Particularly useful visualizations of speech timing data have been forthcoming from use of the ΔC–%V model and the *PVI* models, sometimes in combination, in illustrating similarity clusterings and differences among languages, as previously discussed.

There are other forms of visualization which can be very helpful. Even a straightforward plot of durations as a function of time enables an instant intuitive assessment of temporal evenness or variability (see the *Implementation* section of this contribution). Even more useful is the Wagner Quadrant visualization method (Wagner, 2007) for showing the relations between adjacent interval durations without using the absolute magnitude, a method which was developed as part of a criticism of the methods shown in Table 1 and discussed above, and which, unlike those measures, does not factor out the directionality of differences.

Sections 3.3.2 and 3.4 provide examples of the different kinds of visualizations which are provided in the TGA output.

3 The Time Group Analyzer (TGA) tool

3.1 TGA Requirements specification

As noted previously, methods are recipes for creating solutions to problems, tools are the utensils which are used to implement these recipes, and each utensil is itself based on other theories and models. Practical recipes for the analysis of speech sounds have been around for a long time, and software timing analysis tools may be seen as the utensils for these recipes. There are many other kinds of tools. For example, teachers of English as a foreign language know about ‘gesture tools’ such as the dodge of isochronous tapping on the table and clapping or drumming rhythmically in time with stress beats (though these rhythms may be far from the properties of natural live English speech). A variant of the same isochronous tapping has been the use of a metronome tool in experimental work on timing entrainment (Cummins, 2009).

The TGA tool exploits each of the four main steps involved in creating the input annotations:

1. Extraction of the relevant annotation tier, representing an attribute (i.e. feature type)
2. Extraction of the text of the tier, i.e. the values of the attribute represented in the tier (e.g. phonemes, syllables, feet, phrases, tones stresses, boundaries); currently a subset of UTF-8 encoding is handled, but X-SAMPA encodings, rather than IPA glyph codes are preferred.
3. Extraction of the time-stamps representing association of the sequence values represented in the tier with segments of the signal.
4. Analysis and visualisation of information derived from the time-stamps.

Thus the annotation process essentially follows the segmentation and classification procedures of structuralist phonetics and phonology, and the TGA tool picks up the thread at this point by analyzing temporal relations between time-stamped segments in the selected tier. Input formats for annotations are Praat TextGrids in long or short format, or Character Separated Value (CSV) tables. Other annotation tools than Praat such as Elan and Annotation Pro or SPPAS have import and export functions for these formats as well as their own formats. None of these formats is particularly complex and it is fairly simple to convert one into another. TGA analyses identify the speech rate of the segments on the selected tier such as phones or syllables, duration dispersion by standard deviation and previously mentioned similar functions which yield measures of relative, 'sloppy' or 'fuzzy' near-isochrony, either relative to adjacent units (e.g. *rPVI*, *nPVI*), or relative to the whole sequence, as with standard deviation, the *PIM*, and the *PFD*.

3.2 TGA design

The literature reveals several common methods for processing time-stamped data, in order of increasing sophistication:

1. copying into spreadsheets, sometimes using templates available on the internet for semi-manual processing: a traditional procedure, still common outside well-equipped labs and phonetics departments;
2. use of online tools for specific purposes, such as *nPVI* or speech rate calculation, and further processing with spreadsheets or specialised statistics software;
3. use of prefabricated or *ad hoc* Praat scripts to create numerical output for further processing;
4. implementation of applications in appropriate scripting languages such as *Perl*, *Tcl*, *Ruby*, *R* or *Python*;
5. implementation in languages such as *C*, *C++*, mainly in specialised speech technology applications), independently of time-stamping visualization software.

The TGA online tool falls into the second of these classes, thus filling a gap between non-programming and programming approaches, within a circumscribed functionality for duration analysis, and side-stepping the need for the 'ordinary working phonetician' to use programming techniques. For those with programming abilities, libraries of analysis tools are available, e.g. those in *Perl* in the Aix-MARSEC repository (Auran et al., 2004), or parsing functions programmed in *Python*, such as the *Natural Language Took Kit*, *NLTK* (Bird et al., 2009), or the *TextGrid tools* (Buschmeier et al., 2013).

The architecture of the TGA online tool is shown in Figure 2. Input from an HTML form is passed to a server on the internet (or a localhost server on a standalone machine) and processed by a number of TGA modules, with a variety of output types. The basic design is heavily dependent on the theoretical assumptions outlined in Section 2.

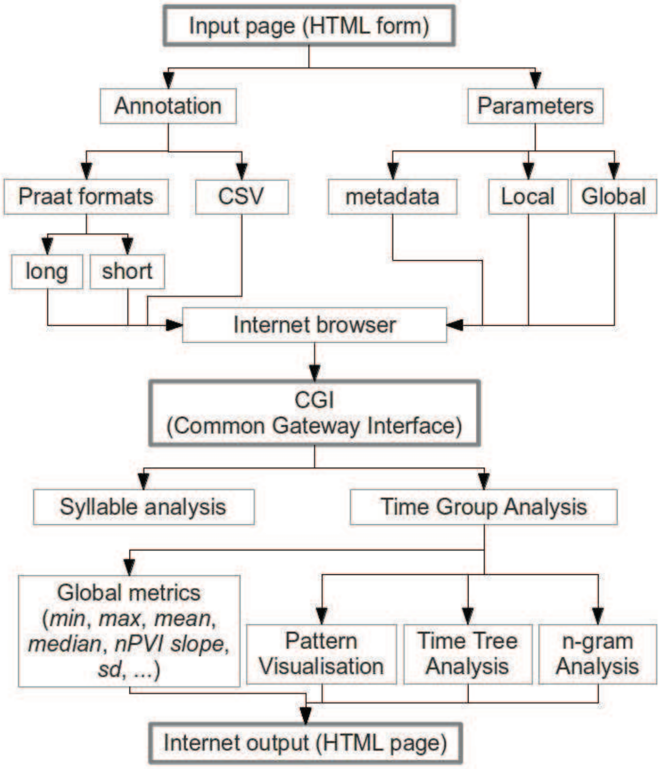


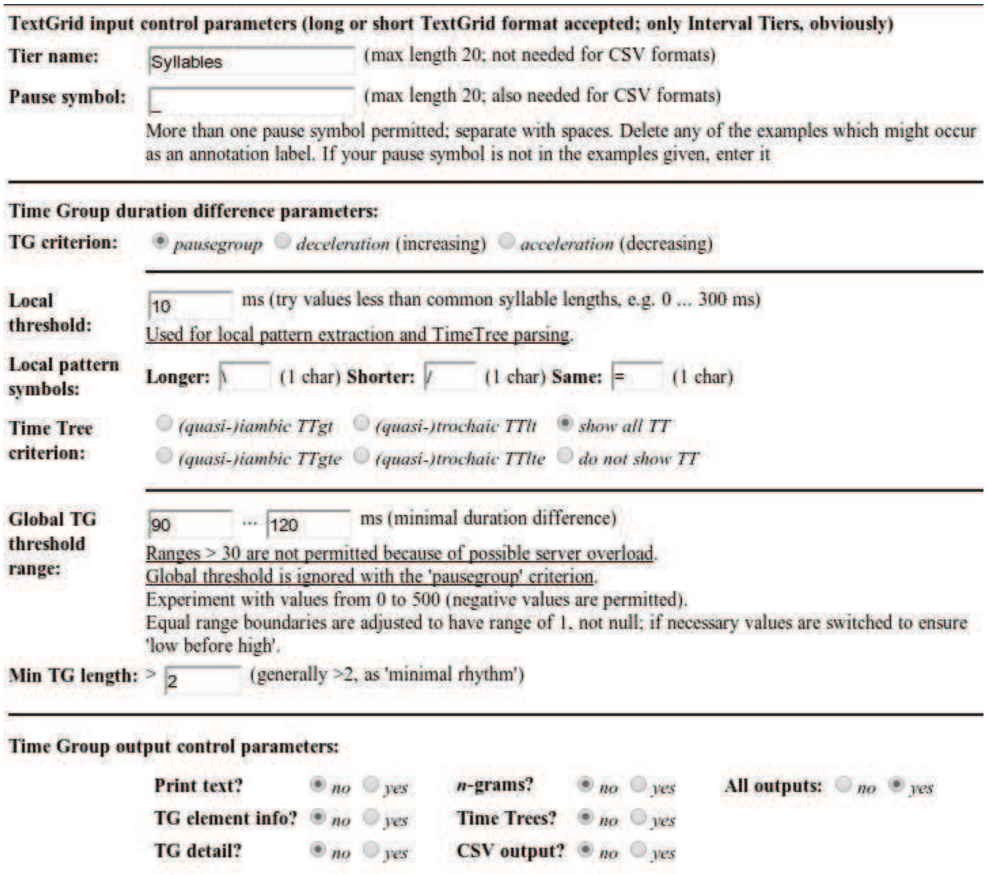
Figure 2: Online TGA architecture.

3.3 TGA Implementation

3.3.1 Input format and parameter setting. The TGA tool is currently implemented in Python 2.7 as a server-side application in the CGI internet environment. The choice of an online environment has many advantages: operation in a standard browser; consistent (because identical) environment at any given time. A disadvantage which is sometimes mentioned is that data input into online tools may be collected on the server by the tool provider. This does not happen with the TGA; user data are neither inspected nor ollected, and user anonymity is preserved.

Input identification and parameter setting in the TGA tool are shown in Figure 3. The parameters are organized into three functionally related groups: input identification, processing parameters, and output selections.

The TGA input module extracts a specified tier (e.g. phone, syllable, foot) from inputs in long or short TextGrid format, or as character separated value (CSV) tables with any common separator. The example specifies a tier ‘Syllables’ and a set of pause symbols which may be used. The pause symbols may be freely selected as long as they do not clash with names of other text labels. The underscore ‘_’ shown in the figure is a very common pause symbol.



TextGrid input control parameters (long or short TextGrid format accepted; only Interval Tiers, obviously)

Tier name: Syllables (max length 20; not needed for CSV formats)

Pause symbol: _ (max length 20; also needed for CSV formats)

More than one pause symbol permitted; separate with spaces. Delete any of the examples which might occur as an annotation label. If your pause symbol is not in the examples given, enter it

Time Group duration difference parameters:

TG criterion: ● *pausegroup* ○ *deceleration* (increasing) ○ *acceleration* (decreasing)

Local threshold: 10 ms (try values less than common syllable lengths, e.g. 0 ... 300 ms)

Used for local pattern extraction and TimeTree parsing.

Local pattern symbols: **Longer:** \ (1 char) **Shorter:** / (1 char) **Same:** = (1 char)

Time Tree criterion: ○ *(quasi-)iambic TTgt* ○ *(quasi-)trochaic TTlt* ● *show all TT*
○ *(quasi-)iambic TTgte* ○ *(quasi-)trochaic TTlte* ○ *do not show TT*

Global TG threshold range: 90 ··· 120 ms (minimal duration difference)

Ranges > 30 are not permitted because of possible server overload.
Global threshold is ignored with the 'pausegroup' criterion.
Experiment with values from 0 to 500 (negative values are permitted).
Equal range boundaries are adjusted to have range of 1, not null; if necessary values are switched to ensure 'low before high'.

Min TG length: > 2 (generally >2, as 'minimal rhythm')

Time Group output control parameters:

Print text? ● *no* ○ *yes* **_n_-grams?** ● *no* ○ *yes* **All outputs:** ○ *no* ● *yes*
TG element info? ● *no* ○ *yes* **Time Trees?** ● *no* ○ *yes*
TG detail? ● *no* ○ *yes* **CSV output?** ● *no* ○ *yes*

Figure 3: Screenshot of parameter input options.

In the processing parameter section, the local threshold permits specification of the minimal difference in milliseconds between durations which determines which durations count as different and which count as equal. The local threshold is relevant for constructing the Duration Difference Tokens (DDTs) described in Section 3.3.2.3 and the Time Trees (TTs) described in Section 3.3.2.4: the larger the threshold, the more duration pairs count as equal, removing random ‘duration difference noise’. The DDT symbols can be freely defined. Four TT types are defined, two based on short-long pairs (quasi-iambic, pairwise deceleration), two based on strong-weak pairs (quasi-trochaic, pairwise acceleration).

The global threshold range is a tentative experimental feature for identifying Time Groups by means of accelerating or decelerating sequences within the specified range.

The minimum Time Group length permits restriction of analysis to Time Groups with a length which promises useful numerical results.

Finally, the output parameter section specifies output of selected results from the modules (see Section 3.3.2) or of all possible outputs.

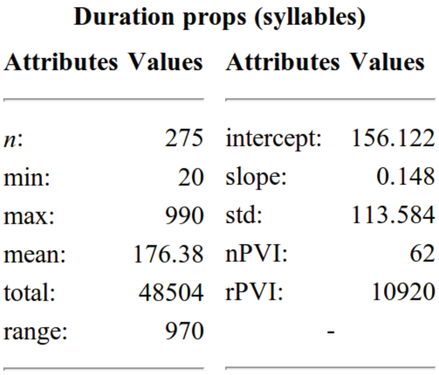
3.3.2 TGA solutions: the main modules. Currently there are three main TGA modules besides I/O and format conversion: (1) text extraction; (2) global basic descriptive statistics for all elements of the specified tier; (3) segmentation of the tier into *Time Groups* with statistics for individual *Time Groups*, and (4) three new visualization techniques for Δ*dur* duration patterns: duration difference tokens, duration column charts, and *Time Trees*.

3.3.2.1 Text extraction. When the annotation has been made directly with annotation software, without prior transcription, there may be a need for transcription text extraction, as documented by a number of web pages providing this functionality, for various purposes such as discourse analysis, natural language processing, archive search, re-use as prompts in new recordings. This facility is provided by extracting labels from annotation elements as running text, separated into sequences by the boundary criteria, e.g. pause, specified in the input. The following example of interpausal groups is extracted from an annotated recording in the CASS corpus of Mandarin (Li et al., 2000):

bei3 feng1 gen1 tai4 yang2 p
you3 yi4 hui2 p
bei3 feng1 gen1 tai4 yang2 zai4 nar4 zheng1 lun4 shui2 de5 ben3 shi5 da4 p
zheng1 lai2 zheng1 qu4 jiu4 shi4 fen1 bu4 chu1 gao1 di1 lai2 p
zhe4 shi2 hou5 lu4 shang5 lai2 le5 ge4 zou3 daor4 de5 p
ta1 shen1 shang5 chuan1 zhe5 jian4 hou4 da4 yi1 p
ta1 men5 lia3 jiu4 shuo1 hao3 le5 p
shui2 neng2 xian1 jiao4 zhe4 ge5 zou3 daor4 de5 tuo1 xia4 ta1 de5 hou4 da4 yi1 p
jiu4 suan4 shui2 de5 ben3 shi5 da4 p
bei3 feng1 jiu4 shi3 jinr4 de5 gua1 qi3 lai2 le5 p
bu2 guo4 p
ta1 yue4 shi4 gua1 de5 li4 hai5 p
na4 ge5 zou3 daor4 de5 p
ba3 da4 yi1 guo3 de5 yue4 jin3 p
hou4 lai2 bei3 feng1 mei2 far3 le5 p
zhi3 hao3 jiu4 suan4 le5 p
guo4 le5 yi2 huir4 p
tai4 yang2 chu1 lai5 le5 p
ta1 huo3 la4 la4 de5 yi2 shai4 p
na4 ge5 zou3 daor4 de5 ma3 shang4 jiu4 ba3 na4 jian4 hou4 da4 yi1 tuo1 xia4 lai2 le5 p
zhe4 xiar4 bei3 feng1 zhi3 hao3 cheng2 ren4 p
ta1 men5 lia3 dang1 zhong1 hai2 shi5 tai4 yang2 de5 ben3 shi5 da4 p

Further analysis of the text output (frequency lists of items, concordance) is planned in future versions of the TGA tool.

3.3.2.2 Global and local descriptive statistics for all Time Groups in the annotation. For calculating global descriptive statistics, three versions of the data are prepared: (1) with all annotation elements on the tier, including boundary elements (e.g. pauses); (2) with only non-boundary elements; (3) with only boundary elements; cf. Figure 4.



Duration props (syllables)

Attributes	Values	Attributes	Values
n:	275	intercept:	156.122
min:	20	slope:	0.148
max:	990	std:	113.584
mean:	176.38	nPVI:	62
total:	48504	rPVI:	10920
range:	970	-	

Figure 4: Screenshot of summary of collated *Time Group* properties and correlations.

Basic statistics, and additionally linear regression (slope and intercept) to show acceleration/deceleration, are also tabulated for each Time Group separately (cf. Table 2, with a selection). The full table output contains not only descriptive statistics for each Time Group row, as shown in Table 2, but also additional information on each row (for this cf. Figure 5, Figure 6). Some of this additional information is dependent on the setting of the minimal difference threshold parameter, which defines degrees of approximate (i.e. ‘fuzzy’ isochrony), rather than strict time-stamp differences. In addition to the numerical output, three novel structural Δdur pattern visualizations are defined (cf. also Figure 5):

1. tokenization of duration differences Δdur into ‘longer’, ‘shorter’ and ‘equal’ duration difference tokens, represented by character symbols (cf. Figure 5), to support prediction of whether specific properties such as rhythmic alternation are likely to make sense (threshold dependent);
2. top-suspended bar chart illustrating the duration Δt of elements in the Time Group (Figure 5), the Duration Bar Sequence (DBS);
3. duration parse tree (*Time Tree*, *TT*) for each Time Group (Figure 6), based on signed duration differences Δdur^{+} and Δdur^{-} (Gibbon, 2003, 2006) to facilitate study of correspondences between duration hierarchies and grammatical hierarchies (threshold dependent).

3.3.2.3 Duration Bar Sequences (DBS) and Duration Difference Tokens (DDT). In Figure 5 two of the novel visualizations are displayed. The hanging Duration Bar Sequence (DBS) provides an iconic representation of syllable (or other selected unit) durations both in width and in height. The row of slashes above the DBS shows the directionality – i.e. alternation – of syllable duration differences as Duration Difference

Tokens (DDT). Comparison with the DBS shows that '\' represents a short-long relation, or deceleration (rallentando, iambic), and '/' represents a long-short relation, i.e. acceleration (accelerando, trochaic), while '=' represents equality of duration (depending on the currently defined local duration difference threshold). In the Mandarin example (top) the DBS shows no obvious alternation of syllables into larger structures such as feet, while the English example (bottom) shows a conspicuous tendency to alternation between long and short syllables. The DDTs show an effect of the local difference threshold: differences <= 10 ms are shown as equal. A selective distributional analysis of bigram DDT sequences is shown in Table 3, providing an indication of the degree of (binary) alternations vs. non-alternations.

Table 2: Selection of output table of local measures for each interpausal Time Group. The full table contains additional columns on the right with the transcription of the TG and visualizations on each row (cf. Figure 5). Number of Time Groups: 23 ; Total duration (without pauses): 31771 ms.

#	**n**	**dur(ms)**	**rate**	**mean**	**median**	**stdev**	**nPVI**	**mednPVI**	**PIM**	**PFD**	**intercept**	**slope**
01	00	0000	0.00	000.00	000.00	00.00	00	00	000	00	000.00	-000.00
02	05	1199	4.17	239.80	250.00	42.29	33	36	005	15	245.60	00-2.89
03	03	0531	5.65	177.00	110.00	94.75	48	48	004	50	076.50	-100.50
04	14	2516	5.56	179.71	186.00	50.48	42	39	070	22	196.11	00-2.51
05	12	1991	6.03	165.92	163.00	58.28	50	46	063	28	166.63	00-0.12
06	11	1834	6.00	166.73	161.00	54.55	34	27	049	27	154.95	-002.35
07	09	1572	5.73	174.67	173.00	52.75	26	22	026	20	135.93	-009.68
08	07	1185	5.91	169.29	181.00	50.69	55	55	018	25	143.46	-008.61
09	16	2470	6.48	154.38	153.00	53.59	40	34	108	27	138.49	-002.12
10	07	1143	6.12	163.29	181.00	50.41	54	55	019	26	167.14	00-1.28
11	10	1752	5.71	175.20	172.50	55.19	39	32	037	24	227.40	0-11.59
12	02	0371	5.39	185.50	185.50	60.50	65	65	001	33	125.00	-121.00
13	07	1149	6.09	164.14	182.00	70.25	58	56	024	36	112.50	-017.21
14	05	0876	5.71	175.20	168.00	55.76	49	52	009	24	130.00	-022.60
15	07	1218	5.75	174.00	162.00	48.33	38	38	014	22	146.89	-009.04
16	07	1332	5.26	190.29	213.00	43.60	27	32	013	21	149.57	-013.57
17	05	0935	5.35	187.00	186.00	65.08	53	54	010	28	207.40	0-10.19
18	04	0641	6.24	160.25	127.00	85.34	56	55	008	44	099.20	-040.70
19	05	0872	5.73	174.40	166.00	16.18	14	16	002	09	185.20	00-5.39
20	07	1344	5.21	192.00	169.00	81.79	42	34	022	36	191.14	00-0.29
21	18	3051	5.90	169.50	167.50	47.11	25	17	109	22	176.53	00-0.82
22	08	1557	5.14	194.63	173.50	41.86	24	19	014	19	167.92	00-7.63
23	13	2232	5.82	171.69	171.00	76.06	63	68	094	35	179.80	00-1.34

In this instance of 'educated Southern British' pronunciation, i.e. slightly modified Received Pronunciation (RP), alternations figure at the top two ranks, totalling 42% of the digrams, and therefore have potential for identification as satisfying the rhythmic Alternation Constraint; deceleration patterns (short-long relations) occupy rank 3. Analyses with thresholds higher than 10ms are necessary for more information about the Alternation Constraint (see Section 3.4.2).

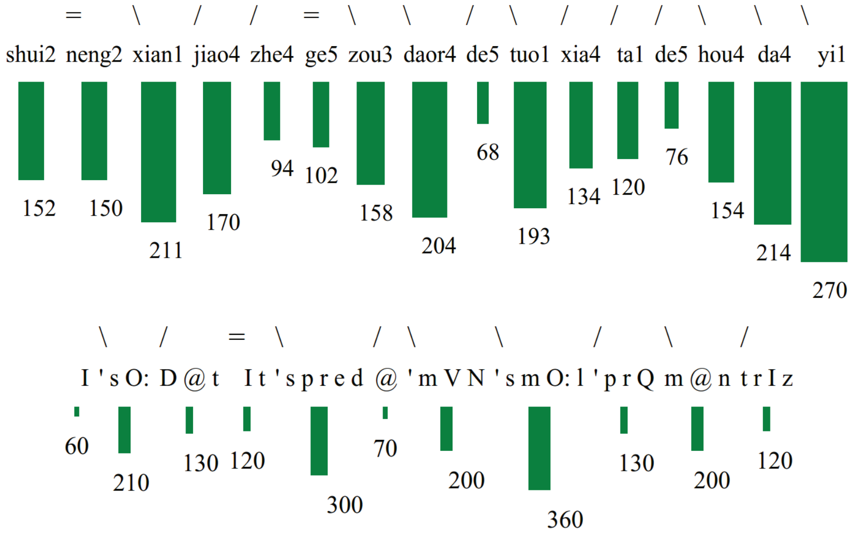


Figure 5: Top: Mandarin. Bottom: English. Duration Difference Token sequence (above) and top-suspended Duration Bars (below); duration is represented by both width and length; scaling is dependent on length of syllables in the transcription.

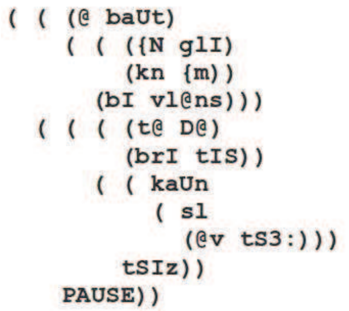
*Table 3: Δ*dur *t*oken rank and frequency analysis.

Rank	Percent	Count	Token digram
1	22%	60	/ \
2	20%	55	\ /
3	11%	31	\ \

3.3.2.4 Time Trees. A further non-traditional visualisation is the *Time Tree* (Gibbon 2003), which groups items in *Time Groups* into binary trees based on the alternation properties of syllables. The *Time Tree* induction algorithm follows a deterministic context-free bottom-up left-right shift-reduce parser schedule. The grammars use Δdur^{+} and Δdur^{-} tests on annotation events in order to induce two types of *Time Tree*, with ‘quasi-iambic’ (decelerating, rallentando) constituents, and ‘quasi-trochaic’ (accelerating, accelerando) constituents, whereby larger constituents inherit the longest duration of their smaller constituents. In Figure 6, a Time Tree constructed over the interpausal group ‘about Anglican ambivalence to the British Council of Churches’ is shown in nested parenthesis notation. The example is taken from the Aix-MarSec English corpus (Auran et al., 2004).

The purpose of generating Time Tree output is to support study of the relation between temporal hierarchical structures and grammatical constituents in a systematic *a posteriori* manner, rather than simply looking for timing correlates of higher level units such as feet or other event types in an *a priori* prosodic hierarchy framework. The example in Figure 6 shows a number of correspondences with grammatical units at different depths of embedding, e.g. ‘about’, ‘British’, ‘Anglican ambivalence’, ‘about

Anglican ambivalence', 'Council of Churches', 'to the British Council of Churches', including foot sequences of Jassem's 'Anacrusis + Narrow Rhythm Unit' type.



```
( ( (@ baUt)
    ( ( ({N glI)
        (kn {m))
      (bI vl@ns)))
  ( ( ( (t@ D@)
        (brI tIS))
      ( ( kaUn
          ( sl
            (@v tS3:)))
        tSIz))
    PAUSE))
```

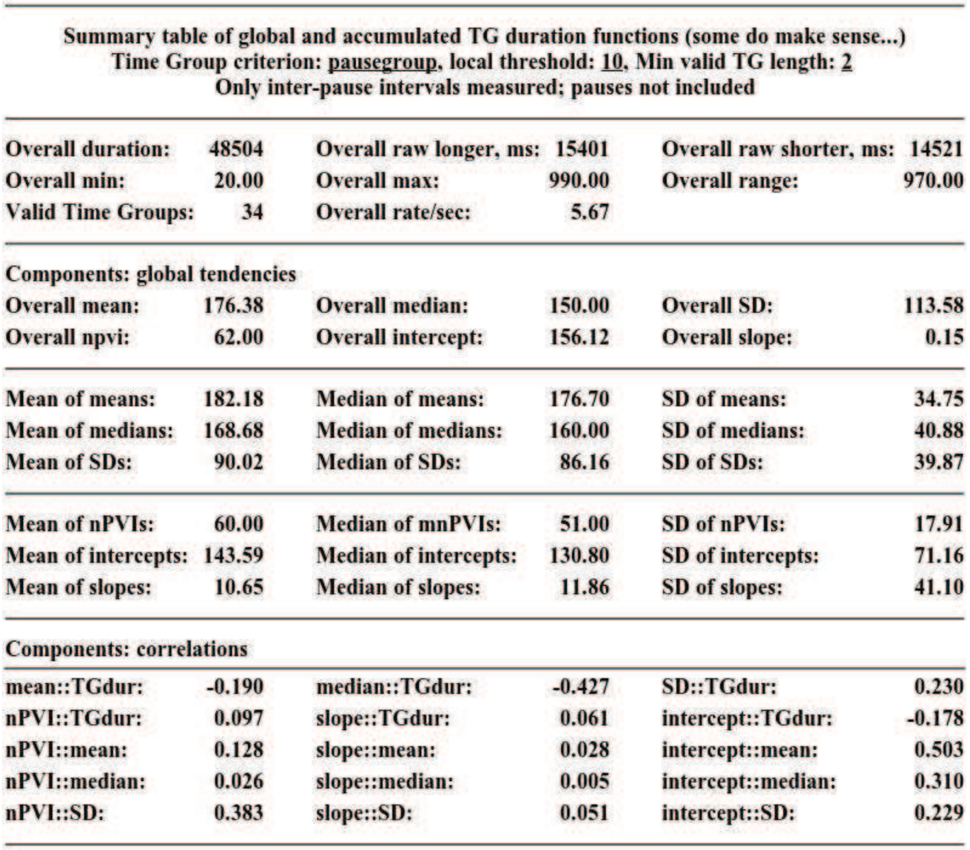
Figure 6: Automatic prettyprint of a quasi-iambic *Time Tree* in nested parenthesis notation.

Crucially, Δ*dur* token patterns and *Time Trees*, (unlike *standard deviation*, *PIM*, *PFD*, *rPVI*, *nPVI*) use signed, not unsigned duration differences, and may therefore claim to represent true rhythm properties. In each case, the minimal local difference threshold setting applies, determining the degree of 'fuzziness' in the distance measurement used in representing duration relations.

A detailed summary chart of the overall statistics is given in Figure 7. The numerical informtion in the chart contains averages over the individual Time Groups, and also provides correlations between the different measures.

3.3.2.5 Wagner Quadrant Graphs. The main further visualization provided by the TGA is the Wagner Quadrant Graph (Wagner, 2007), a scatter plot which reflects the signed z-scores of duration differences rather than the absolute magnitude of differences. The signed differences and z-scores, i.e. (*meanduration – duration*) / *standard deviation*, were used in order to preserve comparability of data, in the context of a critique of the PVI model, which uses absolute magnitudes and raw data. The scatter plots show the duration z-scores of adjacent syllables on the X and Y axes (cf. Figure 8).

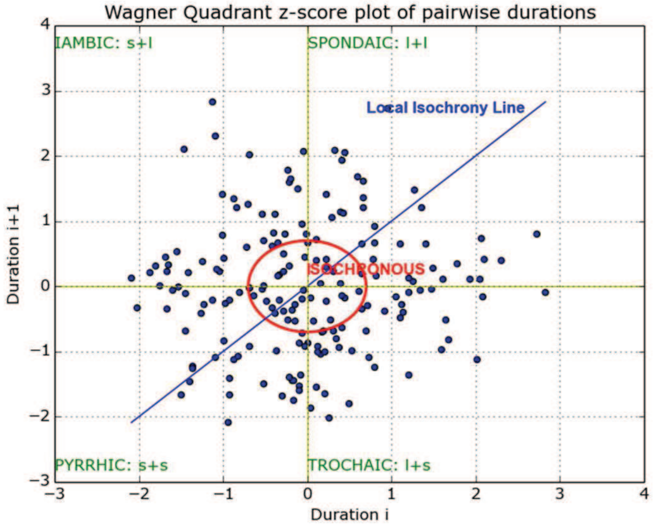
The differences between Mandarin and English syllable duration dispersions are shown very clearly. Mandarin syllable durations are relatively randomly dispersed around a range of durations in an area limited by approximately two z-scores, reflecting a lack of structuring into larger units such as feet. English syllable durations are distributed in an L-shaped formation, with a much larger dispersion and a large cluster of relations between shorter neighbouring syllables in the bottom left quadrant, presumably correlating with sequences of unstressed syllables, as well as a fair number of long-short and short-long syllable pairs, indicating a higher level of structuring, e.g. into feet. There are very few long-long syllable pairs.



Summary table of global and accumulated TG duration functions (some do make sense...)
Time Group criterion: pausegroup, local threshold: 10, Min valid TG length: 2
Only inter-pause intervals measured; pauses not included

Overall duration:	48504	Overall raw longer, ms:	15401	Overall raw shorter, ms:	14521
Overall min:	20.00	Overall max:	990.00	Overall range:	970.00
Valid Time Groups:	34	Overall rate/sec:	5.67		
Components: global tendencies					
Overall mean:	176.38	Overall median:	150.00	Overall SD:	113.58
Overall npvi:	62.00	Overall intercept:	156.12	Overall slope:	0.15
Mean of means:	182.18	Median of means:	176.70	SD of means:	34.75
Mean of medians:	168.68	Median of medians:	160.00	SD of medians:	40.88
Mean of SDs:	90.02	Median of SDs:	86.16	SD of SDs:	39.87
Mean of nPVIs:	60.00	Median of mnPVIs:	51.00	SD of nPVIs:	17.91
Mean of intercepts:	143.59	Median of intercepts:	130.80	SD of intercepts:	71.16
Mean of slopes:	10.65	Median of slopes:	11.86	SD of slopes:	41.10
Components: correlations					
mean::TGdur:	-0.190	median::TGdur:	-0.427	SD::TGdur:	0.230
nPVI::TGdur:	0.097	slope::TGdur:	0.061	intercept::TGdur:	-0.178
nPVI::mean:	0.128	slope::mean:	0.028	intercept::mean:	0.503
nPVI::median:	0.026	slope::median:	0.005	intercept::median:	0.310
nPVI::SD:	0.383	slope::SD:	0.051	intercept::SD:	0.229

Figure 7: Screenshot of global statistics over a sequence of interpausal units.



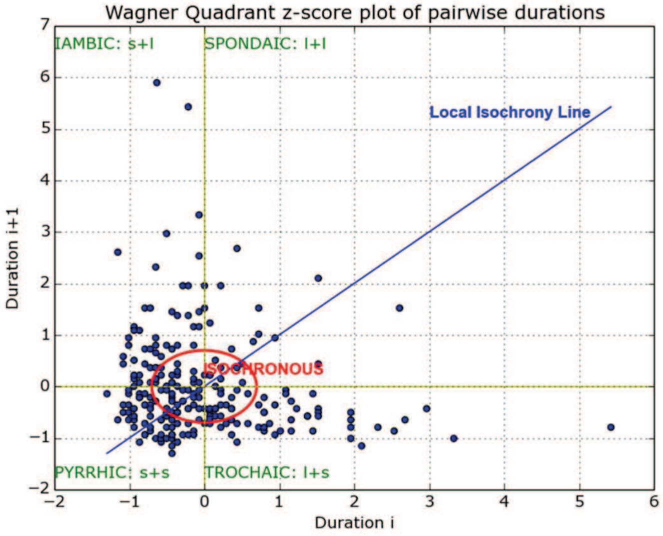


Figure 8: Wagner Quadrant Graphs for Mandarin and English syllable durations in similar reading genres.

3.3.2.6 Reformatted data and analysis outputs. A number of additional options are provided for converting the input data and calculated values (e.g. duration differences, z-scores, DDTs, statistics) into character separated value (CSV) formats, which are convenient for further processing with spreadsheets and other statistical tools. One of the CSV outputs, whether derived from a Praat or CSV input format, has a format identical to a CSV input format, tested by 'recycling' as input to the TGA, leading to identical outputs for all analyses.

3.4 Application in phonetic studies as TGA evaluation

3.4.1 Overview. The TGA online tool has been used in a number of published studies, which may count as a form of functional evaluation of the tool. The most interesting applications have been in studies of native and non-native varieties of Mandarin Chinese, but other applications have been made to genres in English, to Polish and to the Niger-Congo language Tem (ISO 639-3 kdh), a language of Togo (Klessa et al., 2014; Gibbon et al., 2014; Yu, 2013; Yu & Gibbon, 2012, Yu et al., 2014, 2015).

In the following subsections, two constrastive studies are outlined, on native vs. dialect-accented Mandarin, and on the proficiency levels of Mandarin L2 non-native vs. native L1 English pronunciation.

3.4.2 Dialect-accented Mandarin vs. Standard Beijing Mandarin. 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', taken from the CASS corpus.

Time Tree (TT) relations (Gibbon 2006) over interpausal groups were investigated. The following brief example shows a quasi-iambic TT (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 TT (the numbers represent tones):

(((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)))

A comparison of the TT bracketing and the grammatical bracketing (shi2 hou5) and (lu4 shang5) in the TT correspond to the words (shi hou) and (lu shang) in the grammatical bracketing.

Different trees were constructed based on different local thresholds for syllable duration differences, from 10ms to 220ms. Relations between the different trees and words of one or more characters/syllables were investigated. The percentage of agreement between tree constituents and words is shown in Figure 9 as a function of duration difference thresholds (DDTs), for three Hangzhou dialect speakers (HD) and three Mandarin (MD) speakers.

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: (1) correspondences for Beijing Mandarin remain similar as thresholds move beyond 50 ms; (2) 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).

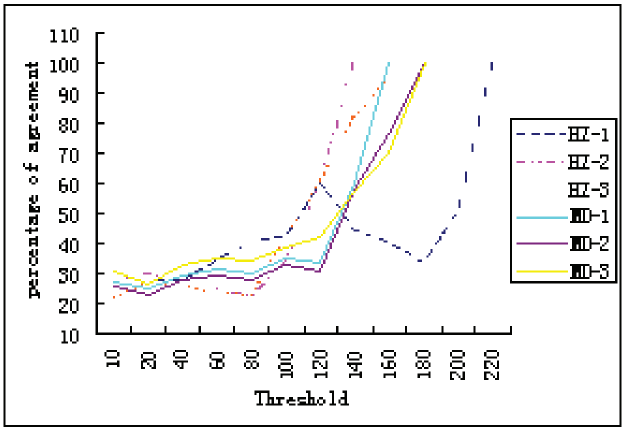


Figure 9: Relations between duration-based syllable groupings and words for speakers of Beijing and Hangzhou varieties of Mandarin Chinese.

3.4.3 Chinese EFL learners vs. English native speakers. Speech recordings of 20 Chinese L2 speakers and 10 English native speakers were used. First the proficiency of the non-native speakers was graded by expert native and non-native English teachers into *poor*, *medium* and *advanced* groups. Using the TGA, the data time-stamps in the annotation files were then further investigated for temporal properties nPVI and syllable rate and temporal patterns. The results are shown in Table 4. The variability of both male and female Chinese learner groups are clearly functions of the proficiency level, while the proficiency of the female learners is somewhat higher by these measures.

Table 4: Summary of mean variability and mean syllable rate for female (F) and male (M) reader groups.

		F: *nPVI*	**F: syll rate**	**M: *nPVI***	**M: syll rate**
Ch L2	**poor**	56	4.2	59	4.3
	medium	62	4.7	65	4.9
	advanced	73	6.3	-	-
Eng	**native**	73	5.3	73	4.8

Wagner Quadrant graphs were constructed with the same data, and show interesting differences in the distribution of adjacent syllable durations (Figure 10). The important feature of the figures is the overall distribution shape, not the details. The low proficiency speaker shows a random distribution of values through the four quadrants. The English native speaker, on the other hand, tends to cluster values in the shorter-shorter, shorter-longer and longer-shorter quadrants; the overall pattern is L-shaped, with larger dispersion range. The advanced Chinese speaker also shows an approximate L-shaped distribution, but small dispersion range. The L-shaped distributions reflect anisochronous syllable timing in English, and the clustering in the shorter-shorter quadrant could be interpreted as sequences of unstressed syllables, indicating non-binary foot structures. Further research is needed to investigate this claim.

Additionally, duration difference token (DDT) *n*-grams were investigated. Percentages for purely alternating quadgrams and quingrams were calculated for each speaker (Table 5). The number of strict quadgram alternations appears as a function of proficiency. Quingrams show no obvious tendency. The non-natives have far fewer strictly alternating sequences than the English native speakers.

Finally, percentages of time-tree/grammar matching between Chinese L2 learners and native speakers were compared in respect of matching and proficiency. Results are shown in Table 6; matchings and proficiency correlate, $r^2 = 0.955$, $p < 0.01$.

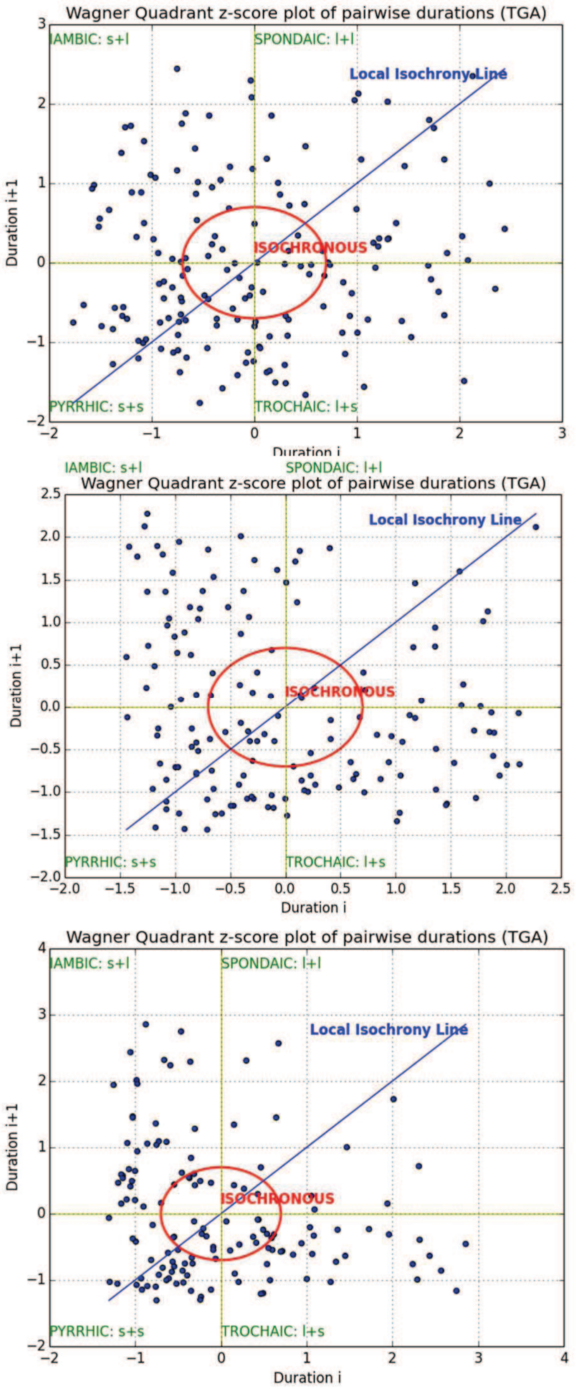


Figure 10: Automatically generated WQ graphs for Chinese L2 English, poor, female; Chinese L2 English, advanced, female; native speaker (USA), female (dispersion shapes are important in the figures, not details).

Table 5: Temporal quadgram and quingram alternation.

		F: 4-gram	F: 5-gram	M: 4-gram	M: 5-gram
Chinese	**poor**	4.5	1.7	5.1	02.6
	medium	8.5	4.3	2.3	08.5
	advanced	5.1	5.8	-	12.2
English	**native**	2.6	2.4	-	09.8

Table 6: Average Time Tree - grammar correspondences.

		female	male
Chinese	**poor**	65.80	67.08
Chinese	**medium**	72.40	69.20
Chinese	**advanced**	75.40	-
English	**native**	77.00	76.95

4 Summary and outlook

The present contribution provides an overview of relevant methodologies for analyzing temporal structures by means of annotation mining with annotated speech data leading to the specification, design, implementation and application of an online tool, Time Group Analyzer (TGA), for the support of linguistic phonetic analysis of speech timing, using time-stamped data, are described. The online tool provides extensive basic statistical information, including linear regression (for duration slope, i.e. acceleration and deceleration) and correlations between the different statistics over sets of Time Groups defined as interpausal units or dynamic (accelerating or decelerating) units. Three innovative visualizations are introduced: Δ*dur* duration difference tokens; top-suspension column charts for Δ*t* and Δ*dur* visualization, and Δ*dur* based *Time Trees* , which are represented as nested parentheses.

Informal evaluation of usability by four trained phoneticians and field evaluation is demonstrated by successful use in published studies, as well as adoption of modules of the TGA tool in software by other developers (AnnotationPro, SPPAS, this volume). The TGA tool reduces previous analysis times for mining time-stamped annotations by several orders of magnitude and supports the achievement of insightful results.

An offline version of TGA for processing large annotation corpora rather than single files is undergoing testing, and further functions such as box plots for timing distributions are in progress.

We anticipate further applications in the L2 teaching field for materials design and proficiency testing, and for the development of models for speech technology. Other disciplines which use duration metrics, such as forensic phonetics, clinical linguistics, dialectometry, stylometry and language acquisition, are also expected to benefit from the efficient methodology provided by the TGA.

References

Abercrombie, D. (1964). Syllable quality and enclitics in English. In: Abercrombie, D., Fry, D. B., McCarthy, P. A. D., Scott, N.C., & Trim, J.L.M. (Eds.): *In Honour of Daniel Jones.*

London: Longmans, 216-222.

Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh University.

Arvaniti, A. 2009. Rhythm, Timing and the Timing of Rhythm. *Phonetica* 2009, 66, 46–63.

Auran, C., Bouzon, C., & Hirst, D. J. (2004). The Aix-MARSEC project: an evolutive database of spoken British English. *Speech Prosody* 2004, Nara, 561-564.

Barbosa, P. A. 2002. Explaining Brazilian Portuguese resistance to stress shift with a coupled-oscillator model of speech rhythm production. In *Cadernos de Estudos Lingüísticos* 43, 71-92. Campinas.

Bird, S., Klein E., & Loper, E.. (2009). E. *Natural Language Processing with Python*. Beijing, etc.: O'Reilly.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International* 5:9/10, 341-345.

Buschmeier, H., & Wlodarczak, M.. (2013). TextGridTools: A TextGrid Processing and Analysis Toolkit for Python. *Tagungsband der 24. Konferenz zur Elektronischen Sprachsignalverarbeitung (ESSV 2013),* Bielefeld, Germany, 152–15.

Carson-Berndsen, J. (1998). *Time Map Phonology: Finite State Models and Event Logics in Speech Recognition*. Dordrecht: Kluwer Academic Publishers.

Cummins, F. (2009). Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics,* 37(1), 16-28.

Gibbon, D. (2003). Computational modelling of rhythm as alternation, iteration and hierarchy. In *Proceedings of ICPhS 15, Barcelona, 2003.*

Gibbon, D. (2003). Corpus-based syntax-prosody tree matching. In *Proceedings of Eurospeech 2003, Geneva 2003*. 761-764.

Gibbon, D. (2006). Time Types and Time Trees: Prosodic Mining and Alignment of Temporally Annotated Data. In: Sudhoff, S., D. Lenertova, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter and J. Schließer, Eds.*Methods in Empirical Prosody Research*. Berlin: Walter de Gruyter, 281-209.

Gibbon, D. (2013). TGA: a web tool for Time Group Analysis. In Hirst, D. J., & Bigi, B. (Eds.) *Proceedings of the Tools and Resources for the Analysis of Speech Prosody (TRASP) Workshop, Aix en Provence, 2013*. 66-69.

Gibbon, D., & Gut, U. (2001). Measuring speech rhythm. *Proceedings of Eurospeech 2001*, Aalborg, Denmark, 91-94.

Gibbon D., Hirst, D., & Campbell, N. (Eds.). (2012). *Rhythm, Melody and Harmony in Speech: Studies in Honour of Wiktor Jassem, Special edition of Speech and Language Technology* 14/15. Poznań: Polish Phonetics Society.

Gibbon, D., Klessa, K., & Bachan, J. (2014). Duration and speed in speech events. In: Mikołajczak- Matyja, N., & Karpiński, M., (Eds.). (2013). *Studies in Phonetics and Psycholinguistics. In honour of Prof. Piotra Łobacz*. Poznań: Adam Mickiewicz Press. 59-83.

Goldsmith, J. (1990). *Autosegmental and metrical phonology*. Oxford: Basil Blackwell.

Gut, U. (2012). *Rhythm in L2 speech*. In Gibbon D., Hirst, D., & Campbell, N. (Eds.): *Rhythm, Melody and Harmony in Speech: Studies in Honour of Wiktor Jassem, Special edition of Speech and Language Technology* 14/15. Poznań: Polish Phonetics Society. 83-94.

Hirst, D. (2009). The rhythm of text and the rhythm of utterances: From metrics to models. *Proceedings of Interspeech 2009, Brighton*. 1519-1523.

Jassem, W. (1952). *Intonation of conversational English (Educated Southern British)*.Wrocław: Wrocławskie Towarzystwo Naukowe.

Jassem, W., Hill, D. R., & Witten, I. H. (1984). Isochrony in English speech: Its statistical validity and linguistic relevance. In: Gibbon, D., & Richter, H. (Eds.): *Intonation, accent and rhythm: studies in discourse phonology*. Berlin: Walter de Gruyter, 203–225.

Inden, B., Malisz, Z., Wagner, P., & Wachsmuth, I. (2012). Rapid entrainment to spontaneous speech: A comparison of oscillator models. In Miyake, N., Peebles, D. & Cooper, R. P. (Eds.): *Proceedings of the 34th Annual Conference of the Cognitive Science Society, Austin, TX*: Cognitive Science Society.

Klessa, K., & Gibbon, D. (2014). Annotation Pro + TGA: automation of speech timing analysis. *Proceedings of LREC 2014, Reykjavik*. Paris: ELDA.

Li, A., Zheng, F., Byrne, W., Fung, P., Kamm, T., Liu, Y., Song, Z., Ruhi, U., Venkataramani, V., & Chen, X.. (2000). CASS: A phonetically transcribed corpus of Mandarin spontaneous speech. In *Proc. Interspeech 2000*, 485-488, Beijing.

Low, E. L., Grabe, E., & Nolan, F.. (2000). Quantitative characterisations of speech rhythm: Syllable-timing in Singapore English. Language and Speech, 43(4), 377–401.

Pike, K. L. (1945). The Intonation of American English. Ann Arbor.

Ramus, F., Nespor, M., & Mehler, J.. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition,* 73(3), 265-292.

Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In Crystal, D., (Ed.): *Linguistic Controversies: Essays in Linguistic Theory and Practice*. London: Edward Arnold, 73–79.

Scott, D. R., Isard, S. D., & de Boysson-Bardies, B. (1986). On the measurement of rhythmic irregularity: a reply to Benguerel, *Journal of Phonetics,* 14, 327–330.

Wagner, P. (2007). Visualizing levels of rhythmic organisation. *Proceedings of the International Congress of Phonetic Sciences, Saarbrücken 2007*, 1113-1116.

Włodarczak, M., Simko, J., & Wagner, P. (2012). Temporal entrainment in overlapped speech: *Cross-linguistic study*. *Proceedings of Interspeech 2012*. 615-618.

Yu, J. (2013). Timing analysis with the help of SPPAS and TGA tools, *Proceedings of TRASP 2013, Aix-en-Provence*. 70-73.

Yu, J., & Gibbon, D.. (2012). Criteria for database and tool design for speech timing analysis with special reference to Mandarin. *Proceedings of Oriental COCOSDA 2012, Macau* (IEEEexplore Conf ID 21048). 41-46.

Yu, J., Gibbon D., & Klessa, K.. (2014). Computational annotation-mining of syllable durations in speech varieties. *Proceedings of 7th Speech Prosody Conference, 20-23 May 2014. Dublin*. 443-447.

Yu, J., & Gibbon, D. (2015). How natural is Chinese L2 English prosody? *Proceedings of ICPhS 2015, Glasgow.* https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0304.pdf