Summary of Professor Gibbon's Lecture on Prosody

Fudan University, Shanghai

Summer School on Contemporary Phonetics and Phonology 2018

Kechun Li

1 Overview

The term "prosody" covers a wide range of phenomena in linguistics. In the *Encyclopedia of Language and Linguistics* and *International Encyclopedia of Linguistics*, many linguistic phenomena are referred to as “prosody”. For example, prosodic units and prosodic structure (See “prosodic phonology”), stress rules (See “metrical phonology”), tone rules (See “autosegmental phonology”, “sandhi tone”), intonation, syntactic-phonological interface (See “Syntax-Phonology interface”) and so on. Professor Gibbon's lecture is related to many of the above-mentioned fields, but the approach is mainly from phonetics. Instead of the three traditional phonetic measurements of prosody, pitch, intensity and duration, Professor Gibbon emphasized the concept of “time”, and highlighted empirical evidence from all levels of linguistic structure, from discourse to phoneme and allophones.

Professor Gibbon gave us four lectures with the following titles: (1) The Phonology of Prosody (not the same as prosodic phonology). (2) The Phonetics of Prosody 1: Rhythm. (3) The Phonetics of Prosody 2: Melody (4) Social Phonetics of Prosody. Along with the lecture some practical tools for investigating rhythms such as TGA and OSCAR are also introduced.

1 Lecture 1: Introduction: time, rhythm, duration

Rhythm is an important component of prosody. People usually consider rhythm in a musical context, but it is much more difficult to identify in speech. Rhythms are generally defined as sequences of alternating values of some feature or features of speech at approximately equal time intervals, such as alternating stress/no stress, and alternating long/short syllables. The classification of English as a “stress-timed” language and Chinese as a “syllable-timed" language is also based on the concept of rhythm.

Isochrony of alternating units is a key point for rhythm. Therefore, the concept of time, or timing, is critical to the study of rhythm. In phonology, time is not treated as concrete minutes or seconds. Rather, it uses abstract features, and the distinction of long and short contrastive durations in paradigmatic relations, and the aggregation of shorter units into larger units in syntagmatic relations. Actual measurable durations are not part of traditional phonological models. For example, studying the duration of syllables, researchers found that there are phonetic differences depending on the stress of syllables, but the difference may not be significant in the phonology of the language, similar to the situation in Chinese where there are insignificant differences in duration of syllables bearing different tones. The classification of stress-timed/syllable-timed has not been supported by consistent acoustic measurement, and if there is a physiological basis, it is more likely to be on the perception side.

2 Lecture 2: Annotation mining and timing measures

Despite the difficulty in finding empirical correlates of rhythm, Professor Gibbon introduced three technical methods, based on annotation mining, which attempt to incorporate phonetic duration into the study of rhythm:

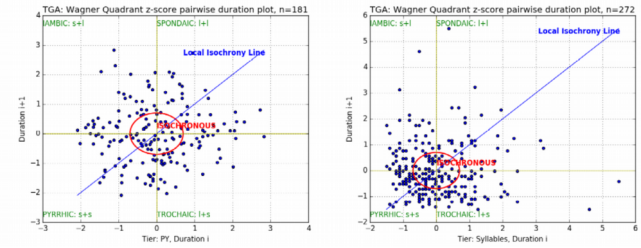
2.1 “One-dimensional” dispersion index models

The so-called ‘rhythm metrics’ are one-dimensional, being based on a single parameter, for example the variance of the duration distribution. This also applies to the slightly more complex metrics PIM, PFD, nPVI, etc. The nPVI (Normalized Pairwise Variability Index) is the most widely used. It considers the variation of the duration of adjacent time intervals in order to rule out influence of

speaking rate. The larger the index, the less stable the duration is. In the sample displayed in class, the nPVI of English is above 50, Chinese is around 40, and some African languages are is below 30, suggesting the duration of the syllables is extremely stable.

2.2 "Two-dimensional" relational model

The two dimensional models demonstrate relations rather than single parameters. These relations can be visualised by means of scatter plots. We can use the four quadrants of a two-dimensional scatter plot to show the relationship of adjacent syllable durations (see the figure below, Mandarin on the left and English on the right). Syllable durations are plotted on the x-axis, and the durations of the following syllables are arranged on the y-axis. The four quadrants show relations the relations short-long (top left), long-short (bottom right), long-long (top right) and short-short (bottom left). The Chinese syllable durations of Chinese are evenly distributed around the average, indicating that the syllable durations are not very different from each other; this shows that Chinese tends to be a syllable-timed language. English syllable durations are also around the average, but they are more centered at the lower left corner (the successive two short-short durations are lower than average) than at the upper right corner (the successive two long-long durations are longer than average). This means that "short + short" syllable combinations are more frequent than the "long + long" combinations.



2.3 "Three-dimensional" time tree model

In generative linguistics, a tree diagram can be generated top-down, based on a series of rules including the Nuclear Stress Rule to determine the prominence of each component. Thinking from another perspective, if we take the actual duration as representative of prominence, then the tree can be inversely induced in a bottom-up procedure, to derive the stress distribution according to certain algorithm. The time-tree method is three-dimensional because it contains a linear syntagmatic relation between linguistic units, such as words or syllables, a hierarchical relation defining groups of linguistic units, and a relation to phonetically measurable time.

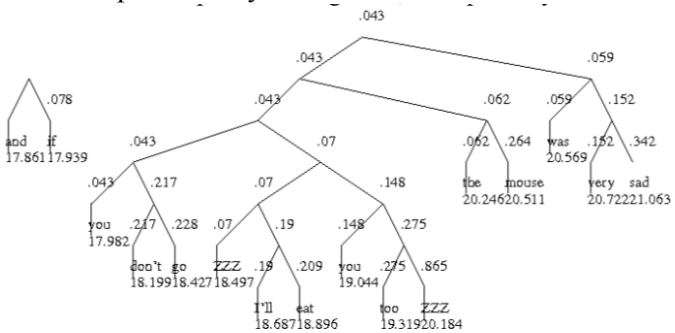


Figure 3: Time Tree spanning 3 seconds of a story ("The Tiger and the Mouse"), iambic time relations.

3 Lecture 3: Oscillator models: modulation and demodulation

3.1 Oscillation

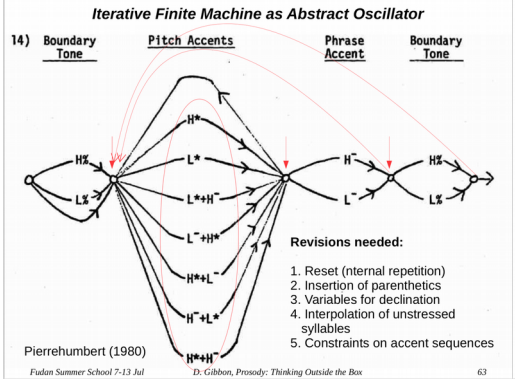
In either phonetics or phonology, the discussion of the rhythm usually concerns time. But annotation mining of timing relations does not fully capture the characteristics of rhythm. An important aspect is ignored in these approaches, which is the alternation of strong and weak features. Any "alternating"

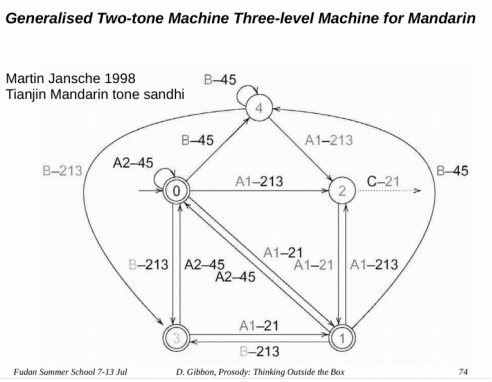
features, such as higher and lower pitch, and stronger and weaker intensity, as well as shorter and longer durations, are all related to rhythm. To resolve this gap, Professor Gibbon proposed an "oscillator model" to explain the perception of rhythm. Similar to the transmission of a broadcast signal, the vocal cords produce a basic carrier frequency, which is subsequently modulated so that the amplitude envelope changes and carries information about rhythm. Several phoneticians have proposed oscillator models for speech production. On the side of perception it is necessary to demodulate the signal, separating the carrier frequency from the modulating frequency. The core steps in demodulation as a formal model of perception can be simplified as follows:

1. Rectify the wave form so that it has only positive values.
2. Extract the amplitude envelope from this half.
3. Treat the envelope as a modulation signal which is independent from the original carrier waveform, and apply Fast Fourier Transform to obtain the frequency of the envelope, that is, the modulation frequency.

The rhythm may be reflected in multiple modulation frequency domains. For example, the rhythm of the phrase level is around 0.5 Hz (corresponding to a duration of 2s), and the rhythm of the syllable level is around 4 Hz, corresponding to a duration of about 250 ms.

In fact, in the study of phonology, oscillator modelling also exists. It appears as an abstract oscillation – iteration. For example, the description of English intonation and Tianjin tone sandhi can be modeled by a finite state machine which not only defines tonal sequences, but also defines iterative patterns by means of loops over alternating features. With the help of the finite state machine, we can incorporate a concept of computable rhythm into the study of prosody.





Although the intonation and tone models were originally designed to describe the melody pattern rather than the rhythm, they became part of the rhythm once they form a repetitive pattern, because rhythm is an "emergent" phenomenon, accompanied by many other characteristics such as intensity, duration and pitch. Melodic alternations can also be part of the rhythm of speech.

2 Melody

Melody also has independent functions. The most direct way of understanding melody is through visual representations - the pitch curve attached to linguistic units. Tone, stress, pitch accent and intonation are all manifestations of melody. In the third lecture on melody, Professor Gibbon first introduced several methods to smooth the fundamental frequency curve (F0 stylisation). We may think that F0 curve showed in *Praat* seems good enough, why should we further smooth it? This is because F0 curve in *Praat* is only numeral measurements. If we want to have a deeper understanding of the melody, we need to construct a mathematical model. The algorithm we use to smooth the fundamental frequency curve is such a model. A good model can also be used in speech synthesis. Professor Gibbon mainly explained linear regression and polynomial regression models. After the smoothing, we also need to evaluate whether the model is suitable. One way is to calculate the difference between the smoothed F0 curve and the original curve. The other is to use perceptual experiments, applying the smoothed F0 curve to the original materials, and see if participants could tell the difference.

The smoothing methods can not only be applied to words and sentences, but also to texts and discourse, so that we get melody of higher levels. A typical melody is the "call contour" in English, such as calling the name "Jenny" at the beginning of the conversation, or saying "thank you" "bye-bye" at the end of the conversation. Interestingly, the call contour in English is limited to start and end a conversation. Professor Gibbon mentioned that when he first heard a German student using the same call contour in saying "Lauter" (meaning “louder”) in classroom, he was surprised because according to English conventions, using such a tone in the middle of conversation was impolite. But later on, he found out that it is a common practice in German. This is an interesting functional difference between the intonation systems of German and English. "Call contour" can only be used at the beginning and end of conversations in English, which corresponds to the function of marking the opening and ending of dialogue in speech act theory. In German, the same melody also has this function, but it can also indicate a problem in communication, such as a misundertanding.

(demonstration of call contour)

“Johnny” “thank you” “bye-bye”

“Lauter”

3 Lecture 4: Sociophonetics

In the fourth lecture on social phonetics of prosody, Professor Gibbon first introduced Labov's classical research paradigm, the "correlation" approach, which links the phonetic variables with social economic variables. Professor Gibbon then introduced another sociophonetic method the online survey to collect public opinions on phonetic data. In the standard experiments, the reliability of the participants’ judgment is often under discussion, but the variability in the judgment data is exactly the target under this paradigm. Professor Gibbon invited us to participate in an OSCAR survey to investigate the assignment of descriptors of perceived pitch of Mandarin tones. We rated the pitch descriptors “high, mid, low, flat, rising, falling, falling-rising, rising-falling” on a five-point scale. In addition to analyzing the relevance of these factors for tone perception, we can also group the participants by hierarchical clustering based on similaritiers in their perception data. It is found that the classification leads to a roughly similar result as traditional dialect classification. This indicates that people from a similar dialect background tend to have similar perception of Mandarin tones. To some extent the results of perceptual experiments can be used as evidence supporting dialect classification.

4 Summary and Outlook

The four lectures closely link to each other, starting from the physical and physiological basis of prosody to model the existing rhythm phenomena, then correlation research through questionnaires and other methods, the results of which could help with further verifying or improving the prosodic model. Professor Gibbon's lectures are clearly presented and full of humorous comments. The emphasis on data visualization has the effect of “one picture says more than a thousand words”. Professor Gibbon also introduced two practical tools for prosody analysis, which he encouraged everyone to make use of.

Time Group Analyzer (TGA) is a practical rhythm analysis tool. With the input of annotated Textgrid files, it can automatically extract the duration information of each interval, and calculate relevant parameters like the variance, nPVI for each subsection as well as the whole passage, along with some visualizations. TGA can be used online.

The Online Survey Collation and Reporting tool (OSCAR) is an online platform for opinion survey collection and reporting. The survey input uses the Likert format to investigate participants’ opinions on phonetic data as well as to collect background information such as age, region, gender, and language. The software will automatically analyze the results. Experimenters who are interested in using OSCAR should contact Professor Gibbon.

Both tools can be found on Professor Gibbon homepage.

http://wwwhomes.uni-bielefeld.de/gibbon/

4 Further reading

Gibbon, Dafydd. "Time types and time trees: Prosodic mining and alignment of temporally annotated data." Methods in Empirical Prosody Research (2006): 281-209.

Gibbon, Dafydd. "Prosody: The Rhythms and Melodies of Speech." arXiv preprint arXiv:1704.02565 (2017).

Gibbon, Dafydd. "The Future of Prosody: It's about Time." arXiv preprint arXiv:1804.09543 (2018).

Gibbon, Dafydd, and Huangmei Liu. "Variability in Mandarin Tone Perception: a multidialectal approach." Proc. 9th International Conference on Speech Prosody 2018. 2018.

Liberman, Mark, and Alan Prince. "On stress and linguistic rhythm." Linguistic inquiry 8.2 (1977): 249-336.

Pierrehumbert, Janet Breckenridge. The phonology and phonetics of English intonation. Diss. Massachusetts Institute of Technology, 1980.

Wagner, Petra (2007). “Visualizing levels of rhythmic organisation.” Proc. International Congress of Phonetic Sciences, Saarbrücken 2007, pp. 1113-1116, 2007.