

FORMAL IS NATURAL: TOWARD AN ECOLOGICAL PHONOLOGY

Dafydd Gibbon

Universität Bielefeld, Germany

gibbon@uni-bielefeld.de

ABSTRACT

Naturalism in linguistics has a history of opposition: to abstractness, to generative linguistics, to formalist approaches. The present approach concentrates on a key feature of Natural Linguistics and Natural Phonology in particular, namely the empirical centrality of external evidence. A step further than traditional naturalism is taken, in replacing the Viennese schools of philosophy of science, specifically Logical Empiricism and Critical Rationalism, which typically underlie the metatheory of Natural Phonology, by an ecological view that science is evaluated - intellectually and by the taxpayer - in terms of its operational functionality in personal, social, political and economic environments. The ecological perspective on science is related to applications of linguistics in speech technology.

Keywords: Natural Phonology, finite state technologies, syllables, tone

1. INTRODUCTION

The aim of this contribution is to stimulate discussion and modernisation of the concept of naturalism as a universal principle in post-generative linguistics: requirements for theory evaluation are extended beyond traditional notions of internal and external evidence, and assigned instrumental roles in a utilitarian ecology of science.

The perspective of “science for science’s sake”, represented most notably by the Vienna schools of positivism, logical empiricism and critical rationalism, is a popular one, but in many ways unrealistic. Even Kuhnian paradigm theory and more recent approaches to the philosophy of science are limited in this respect. In linguistics, the view is often coupled with a restriction to “internal evidence”, i.e. evidence from judgments on the structure of language. In contrast, the empirical foundation of Natural Phonology (in fact Natural Linguistics in general) is distinguished by its emphasis on “external evidence”, i.e. by evidence from language use in context. However,

the ecological circle does not close: “products” are rarely if ever put back into these environments in operational applications and evaluations.

A key factor in understanding the development of science, which has been determinedly ignored in many quarters, including linguistics but less so in physics, with its telescopes and cyclotrons, to confirm this. The birth of modern phonetics in the second half of the 19th century, for example, can arguably be assigned to two applications oriented pushes: in technology, the invention and marketing of the telephone by the Scottish engineer and phonetician Alexander Graham Bell, and in foreign language teaching the founding of the International Phonetic Association and its journal *Le Maître Phonétique*, later the *Journal of the IPA*. Linguistics and explorative taxonomic dialectology contributed further intellectual foundations.

Likewise, many recent insights in phonetics owe a great deal to the technology push which took computers out of laboratories and put them on every phonetician’s desk, and to personal and community initiatives to provide sophisticated free software for these computers. The combined effect has been a creative cycle of scientific ideas which go back into the technology and from there push intellectual developments forward in an ecological cycle. Technology is not *causal* in this development: the cause lies in the inventiveness of the human mind. Technology is instrumental, utilitarian, a *necessary*, though not a *sufficient* condition for science; computation is likewise a necessary, but not a sufficient condition for technology; neither technology nor computational models need be close mirrors of the workings of the mind.

This claim is intended to provoke discussion by taking up the central feature of external evidence, which is closely associated with Natural Phonology. Exploiting external evidence is not sufficient: theories have full explanatory status only with explicit procedural and computational interpretations within the ecological niche of science itself, and within broader technologies.

2. NATURALISM AND TECHNOLOGY

The following theses are intended to be part of a specific approach to ecological explanation in linguistics, in which technology has a central role to play:

Utility Thesis: Only that subset of scientific endeavour which has technological consequences survives.

Formalisation Thesis: In order to have technological consequences, theories and descriptions must be formalised according to complete and sound criteria.

Operationality Thesis: In addition to formalising theories and descriptions, the procedures of derivation and constraint resolution must be provided with an explicit, computable operationalisation.

In short: no “naturalness” without computation.

Discussions of scientific methodology in linguistics have often appealed to scientific revolutions. However, technological revolutions are equally important.

We do not need to go back to Gutenberg and printing technology, or even to Archimedes, or Newtonian and Einsteinian mechanics, in order to illustrate the Utility Thesis. A few examples will illustrate the point:

1. The relationship between what has been called “symbol phonetics” and “signal phonetics” has been (and is still being) convincingly clarified by work with time-aligned annotation of transcription labels with speech signals, using a wide variety of free and commercial software.
2. Phonological and phonetic theories and descriptions are increasingly being operationally tested by operationalisation in speech synthesis systems (Hertz 1999; Dirksen & al. 1997; Bachan & al. 2006). Operational tests are necessary though not sufficient criteria for truth.
3. The success of statistically based speech recognition has clearly demonstrated that phonology and impressionistic phonetics have severe limits in relating phonological rules to the reality of speech on their own.
4. It is increasingly being recognised that the linguistic and phonetic sciences cannot be seriously practised without formally well-defined resource creation in the developing *Language Documentation* paradigm.
5. The practical requirements of creating reusable linguistic and phonetic resources (defined as signal and annotation data, tools and descriptions) depend on concepts which go back

to the text grammars of 1970s Text Linguistics and are embodied in XML text and document technology implementations for the internet and for libraries and archives.

The second of these five areas, technological operationalisation of theories and descriptions, forms the substance of the following discussion. Two examples will be discussed: the formalisation of syllable structure, with reference to English, and the formalisation of prosodic structure with reference to tone languages.

3. SYLLABLE STRUCTURE

3.1. Modelling

Syllable structure is conventionally represented in various ways. One common representation lists constraints on segment class sequences:

V, CV, CVC, VC, CCCVCCC, ...

Another defines sets of constraints on segment properties, e.g. for English (informally):

in the sequence #x[stop][liquid], *x = /s/*

A third defines syllable structure in terms of sonority differentials:

$\text{sonority}_{\text{margin}} < \text{sonority}_{\text{peak}} > \text{sonority}_{\text{margin}}$

These techniques, in their more sophisticated versions, are well-suited to describing properties of syllable structure. But they have in common that they are sets of disconnected constraints of various kinds, and they each leave two questions from the formalisation thesis open:

1. How do we know that the descriptions are complete, i.e. that nothing is missing?
2. How do we know that the descriptions are sound and nothing unwanted is included?

The obvious answer to these questions is to combine each separate constraints into a coherent single system; this is, in general, what is done in morphology and syntax: a formal grammar is designed with the task of enumerate all and only the expressions of a language, thereby fulfilling formal criteria of *completeness* and *soundness*.

To write a syllable grammar would seem straightforward and uninteresting because syllables have a clear, small finite limit on their length (unlike words and sentences): this length may range from 2 (in *CV* languages) to 8 (depending on the analysis, in *CⁿVC^m* languages). In the trivial case, syllables may therefore simply be listed.

In a *CV* language this is unproblematic. With, say, 15 consonants and 10 vowels, the maximal number of *potential* syllables is $15 \cdot 10 = 150$; there

may be constraints on combining *Cs* and *Vs*, limiting this number further.

In a C^mVC^m language, the situation can be very different; based on a detailed formalisation, the number of *potential* syllables in English can be shown to be at least 30,000; the number of *actual* syllables is usually estimated at less than 10,000.

Whether all remaining potential syllables are used as *actual* syllables in the lexicon (or as monosyllabic words) is a different issue: limiting the set to actual syllables precludes explanation of how new syllables such as “skype” are invented, i.e. the re-combination of existing sub-sequences (onsets, nuclei and codas) in different ways.

There are two main options for defining a grammar for capturing syllable structure in a more interesting way which captures generalisations. One is to use a phrase structure grammar, as in many approaches to syntax and morphology. This kind of grammar is generally used for two reasons. First, it can define arbitrary kinds of recursion as well as constituent groups; however, for a finite language, recursion is not needed. Second, this kind of grammar has the advantage of showing syntagmatic clusters: groups which share mutual constraints, such as Noun Phrases in syntax, or Onsets, Rhymes, Nuclei and Codas in phonology.

to advantage by visualising the topology of the grammar in detail. There are several equivalent way of formulating this kind of grammar: regular expressions, right or left branching rewriting rules, Finite State Automaton (FSA) or network (FSN).

An example of a Finite State Transition Network (one possible representation of a Finite State Automaton) is shown in Figure 1. This kind of representation has several advantages, as with many other kinds of data visualisation:

1. The network shows the topology of the syllable structure of English directly.
2. The onset structures and substructures (mutual constraints) are very clear.
3. The lack of constraints between onsets and rhymes is also clear: these two structures meet at a single node; the one exception is also clear: /u:/ after certain consonants in many English dialects.
4. The intricate constraints between nucleus and coda are clearly distinguished. The anomalous pattern with /u:/ after certain consonants is also clear.
5. The network is easily shown to be incomplete: there is no path describing sequences such as /spju:/, /stju:/, /skju:/; insertion of these paths is unproblematic.

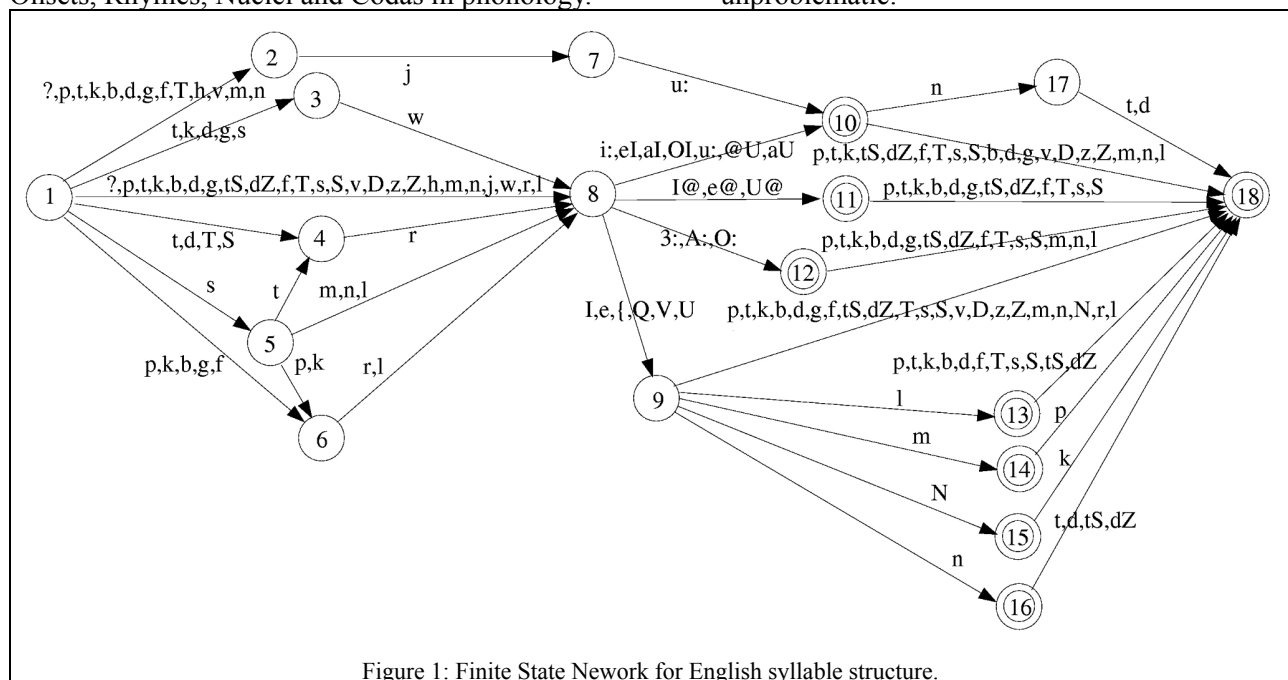


Figure 1: Finite State Network for English syllable structure.

The second main option is to start from basics and to use a regular (linear, finite state) grammar, with which restricted kinds of recursion (left-branching or right-branching but not centre-embedding) can be defined, and which can be used

6. The network can be split, if required, into the individual constraints of which it is constituted, to conform to the conventional expectations of phonologists.

7. If non-local constraints are required, such as **/sCiVCi/* where *Ci* means consonants with identical properties, then these can be added simply by including conditions on finite registers (variables), which do not affect the FS status of the grammar.
8. The contexts for allophones of a given phoneme are completely defined by the position of the phoneme in the FSN.
9. The linear left-to-right structure of the FSN appropriately represents the purely linear, non-hierarchical contexts required for identifying allophonic variation.
10. The subnetworks in the FSN provide contexts for defining dissimilation (in the Onset) and assimilation (in the Coda) between neighbouring segments.
11. The ambisyllabicity problem with tree-based grammars disappears, formally: an intervocalic consonant simply appears linearly just where it is, with no artefactual coercion into one of the neighbouring syllables or the other.
12. The mapping to tree-shaped Syllable, Onset, Rhyme, Nucleus, Coda notations is straightforward, each of these categories corresponding to an area of the FSN.
13. It is very straightforward to represent this kind of network in such a way that it can be interpreted by a well-defined Finite State Automaton interpreter, implementable in any programming language. This interpreter can be used either to accept or reject hypotheses about specific potential syllables, or to generate a complete set of potential syllables for checking.

3.2. Application

A test of whether a system of rules or constraints is *complete* and *sound* is to link the individual constraints into a connected system such as the one illustrated here, and generate appropriate output. Although the Phrase Structure Grammar can also be formalised in a similar way, in practice it is necessary to enhance such grammars with constraints across sub-trees, which do not lend themselves to perspicuous visualisation.

From the perspective of the *Operationality Thesis* presented above, the FSA is in fact standard technology in computational linguistics and in the human language technologies, and is embodied in many kinds of applications, from speech synthesis and speech recognition to morphological analysers and the spell-checkers provided in well-known

word processors. A widely used and linguist-friendly FS software tool is the Xerox *xfst* package (Beesley et al. 2003). Detailed operational accounts of FS phonology, with applications in speech recognition and machine learning, are given by Carson-Berndsen (1998) and Belz (2001).

4. Tone

4.1. Description: Niger-Congo languages

Finite State modelling has frequently been applied to prosody, and operational naturalness has been approximated, for instance, in intonation generation in speech synthesis systems. The intonation literature cannot be reviewed here; suffice it to say that it has long been known, from work by Fujisaki in the 1960s and later, the IPO group in the 1970s, Pierrehumbert in the 1980s, and work by the present author from the 1980s to the present, that intonation systems can be represented by finite state models of contour or boundary tone patterns: from initial boundary tones through iterated accentual tones to final boundary tones.

It is less well-known that similar considerations apply to the pitch patterns of tone languages (Gibbon 1987, 2001, 2004). Unlike standard practice in syntax, the influential study of right-branching metrical trees for tone terracing by Clements (1981) did not include grammatical rules to generalise over specific trees. It is not an accident that a formal property of FS grammars is that they can be used to construct right or left branching tree structures but not centre-embedded or mixed right and left branching tree structures. This insight was used to develop a FS model of tone terracing for Niger-Congo languages.

Ibibio (Niger-Congo, specifically Delta Cross; ISO 693-2: *nic*; Ethnologue: *IBB*), simplifying slightly, has two lexical tones, H and L; nouns have lexical toneme contrast, verbs have templatic patterns and inflectional tone, and grammatical and word formation contexts provide structural templates to which tones are assigned. Because of these complex factors, tone sequences are more or less arbitrary from a tonological perspective alone, and could be represented by a 1-node network and two tonal transitions, one for H and one for L. However, tonal terracing requires a phonetic interpretation which takes four different possible tone-to-tone transitions into account: HH, HL, LH, LL, for which the following phonetic conditions

hold, in which lower case “h” and “l” stand for the phonetic correlates of “H and L”:

HH: $h_1 = h_2$ LH: $h > l$

HL: $h \gg l$ LL: $l_1 = l_2$

Consequently, a more complex network is required in order to capture these constraints. A FSN which will generate or recognise such sequences is shown in Figure 2.

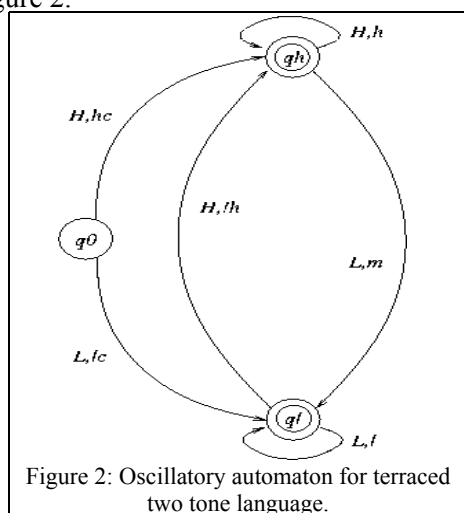


Figure 2: Oscillatory automaton for terraced two tone language.

These relations can be quantified in terms of pitch ratios (Gibbon 2001); in some languages, the “=” relation is rising for H contexts (upsweep) and falling for L contexts (downdrift), and there are languages with more complex but still FS terracing contexts.

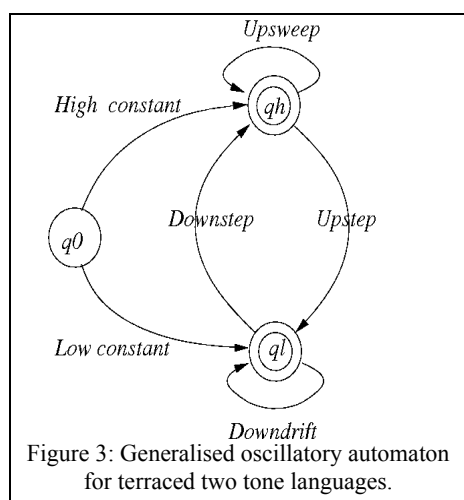


Figure 3: Generalised oscillatory automaton for terraced two tone languages.

From the start node on the left, either a H node or a L node can be reached. Different allotones are associated with the HH, HL, LL, LH context types, which can be generalised using common terms from the tone language literature as shown in Figure 3. A complete terrace is then represented formally by one transition cycle from either a H node or a L node back to the same node.

4.2. Description: Mandarin

The question immediately arises of how far the FS tone model may be generalised to other language types. Jansche (1998) applied a FS formalisation to a Mandarin dialect by unifying tone sandhi rules into a connected system which can be used to recognise or to accept potential tone sandhi sequences in this dialect. This unification procedure reveals marked typological differences between Niger-Congo and Mandarin tonal language types (cf. Figure 4); detailed comment on this is not called for in the present context.

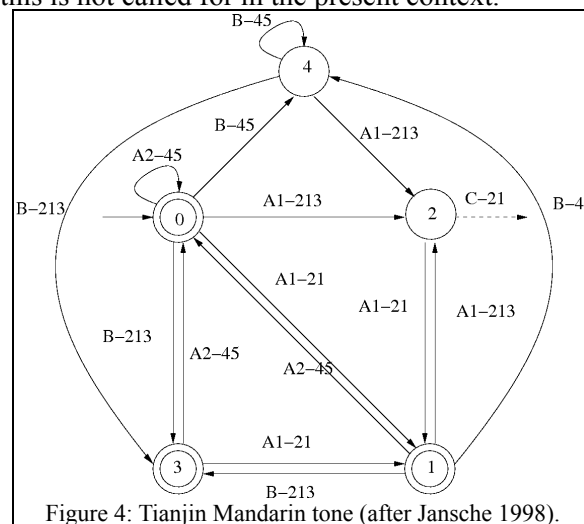


Figure 4: Tianjin Mandarin tone (after Jansche 1998).

4.3. Application

An operational use of the FS model for empirical modelling of tonally annotated speech recordings, with mapping of phonetic tone categories to the quantitative operations of Pierrehumbert et al. (1984), is presented by Urua et al. (2003).

The tonal FSA has also been implemented in a specific speech synthesis prototype. One of the most useful HLT systems for use with non-literate communities, with typical applications in health, market and educational information dissemination, and a good environment for the fulfilment of the Utility Thesis is Text-To-Speech synthesis. However, current TTS development environments are not language independent, and voice creation is not easy, but adaptation of a voice in one language to another language or a different style or gender in the same language can provide an adequate platform for some purposes (Bachan et al. 2006).

Adaptation is plausible when languages are prosodically and phonemically similar but severe problems arise when languages are very dissimilar (e.g. ‘intonation languages’, for which TTS

systems are typically developed, vs. ‘tone languages’, e.g. Ibibio). In the LLSTI project (Tucker et al. 2005), the adaptation procedure was applied to Ibibio. In the prototype for this project, the two-tone FS model was used (Gibbon et al. 2006). The generalisability of the approach is being investigated by Govender et al. (2005).

Ongoing work also shows that tones can be modelled in relation to complex morphosyntactic structures in Ibibio using FS transducer models (Gibbon et al. 2006). Initial studies show that these problems can also be generalised to other African tone languages, which have lexical (phonemic) tone but, unlike Asian tone languages, also morphemic tone; it is planned to use these FS models in TTS language models.

5. CONCLUSION

In this paper, an operationally extended notion of linguistic naturalism was discussed, and illustrated with models using Finite State technologies in phonology and some of their applications. A precondition for seeing beyond the current state of the art in the field surely lies in being prepared for apparent paradoxes - “formal is natural” - and accepting naturalism in linguistic descriptions not just as *receiving criteria from language use as empirical evidence* but also as *testing hypotheses with applications in performance contexts using operational computational models*.

So a genuinely naturalistic approach to syllable structure (and of any language structure) is one which is able to use appropriate technology to put its approach back into the context of language use from which it took its original evidence, and to enhance this use and its future analysis in line with the Operationality Thesis. “Formal is natural”: explicit formalisation is a prerequisite for computation and, in turn, for technology; technology then provides the impetus for further social and scientific development.

The universal role claimed here for technology in science appears in the contexts discussed in this paper not as a one-way theory-to-application flow, but as a cyclical interaction in which intellectual potential and technological instrumentality are linked within the overall ecological space within which science - phonology and phonetics no less than other sciences - operates and within which scientists exercise or fail to exercise their involvement with and responsibilities to their intellectual and social environments.

6. REFERENCES

- [1] Bachan, Jolanta & Dafydd Gibbon (2006). Close copy speech synthesis for speech perception testing. *Investigationes Linguisticae* (13), Poznań, 9-24.
- [2] Beesley, Kenneth R. & Lauri Karttunen (2003). *Finite State Morphology*. Stanford: CSLI Publications.
- [3] Belz, Anja (2000). Computational Learning of Finite-State Models for Natural Language Processing. Doctoral thesis, University of Sussex.
- [4] Carson-Berndsen, Julie (1998). *Time Map Phonology: Finite State Models and Event Logics in Speech Recognition*. Dordrecht, Holland: Kluwer Academic Publishers.
- [5] Clements, G.N. (1981). The hierarchical representation of tone features. *Harvard Studies in Phonology* Vol. II. G. Clements, ed., 50-107. Reproduced and distributed by Indiana University Linguistics Club.
- [6] Dirksen, Arthur & John S. Coleman (1997). All-Prosodic Synthesis Architecture. J. P. H. van Santen, R. W. Sproat, J. P. Olive and J. Hirschberg, eds. *Progress in Speech Synthesis*. New York: Springer-Verlag, 91-108.
- [7] Gibbon, Dafydd (1987). Finite state processing of tone systems. *Proc. 3rd European Association of Computational Linguistics Conference 1987*, 291-297.
- [8] Gibbon, Dafydd (2001). Finite state prosodic analysis of African corpus resources. *EUROSPEECH-2001*, 83-86.
- [9] Gibbon, Dafydd (2004). Two problems and two methods for prosodic typology. *Proc. Int. Symposium on Tonal Aspects of Languages with Emphasis on Tone Languages*, Beijing.
- [10] Gibbon, Dafydd, Eno-Abasi Urua, Moses Ekpenyong (2006). Problems and Solutions in African Tone Language Text-To-Speech. In: *Proc. ISCA Tutorial and Research Workshop on Multilingual Speech and Language processing*, Stellenbosch, South Africa.
- [11] Govender, Natasha, Etienne Barnard, Marelise Davel (2005). Fundamental Frequency and Tone in isiZulu: Initial Experiments. *Proc. Interspeech2005 - Eurospeech*, Lisbon, 1417-1420.
- [12] Hertz, Susan R., Rebecca J. Younes & Nina Zinovieva (1999). Language universal and language specific components in the multi-language ETI Eloquence Text-To-Speech System. *Proc. 14th ICPhS, San Francisco*, 2283-2286.
- [13] Jansche, Martin (1968). A two-level take on Tianjin Tone. In *Proc. 10th European Summer School in Logic, Language and Information, Student Session*. Saarbrücken, 162-174.
- [14] Liberman, Mark & Janet B. Pierrehumbert (1984). Intonational invariance under changes in pitch range and length. Aronoff, Mark & Richard T. Oehrle, eds., *Language, sound, structure*. Cambridge: MIT Press, 157-233.
- [15] Tucker, Roger & Ksenia Shalnova (2005). Supporting the Creation of TTS for Local Language Voice Information Systems. In: *Proc. Interspeech'2005 - Eurospeech*, Lisbon, 453-456.
- [16] Urua, Eno-Abasi, Gibbon, Dafydd & Gut, Ulrike (2003). A Computational Model of Low Tones in Ibibio. *Proceedings of the ICPhS conference 2003*, Barcelona, 623-626.