

# Morphotonology for TTS in Niger-Congo languages

Dafydd Gibbon & Eno-Abasi Urua

Universität Bielefeld, Germany; University of Uyo, Akwa Ibom State, Nigeria

gibbon@uni-bielefeld.de; amandinyene@yahoo.com

## Abstract

It is well-known that many East Asian languages have lexical (i.e. phonemic) prosody, and languages such as Mandarin are very well described. African languages are also frequently mentioned in the literature as tone languages, and phonetic interface patterns such as *downstep* are well-documented. It is less well-known that the functionality of tone patterning in African tone languages is fundamentally morphosyntactic rather than phonemic, in that (a) tonal patterning is specific to particular parts of speech, (b) tones may have inflectional function and play a role in both (c) derivational and (d) compounding word formation patterns and (e) in marking syntactic phrasal templates. The aim of this paper is both to document the morphosyntactic functionality of tones in African languages within a typological context as compared to East Asian tone languages such as Mandarin, and to develop finite state architectures for tone handling in practical Text-To-Speech synthesis in health and agriculture information projects in Ivory Coast and Nigeria. Morphosyntactic tonal functionality is illustrated for Ibibio (Lower Cross Niger-Congo, South-Eastern Nigeria), but also applies to other Western and Central African languages.

## 1. Forms and functions in pitch systems

The theoretical objective of this study is to place the typological characteristics of the tonal systems of Niger-Congo languages, and possibly also of Bantu languages, within a clear framework for pitch system typology. The term ‘pitch’ is used deliberately in order to ground the analysis empirically in a clear phonetic domain; this is not possible with generally rather ill-defined notions such as ‘intonation’ and ‘tone’. The operational objective is to provide a basis for incorporating the complex pitch systems of Western and Central African languages, which differ greatly from the better known intonation-accent languages of Europe and the lexical tone languages of South-East Asia, into speech synthesis systems for practical deployment in health and agricultural information and education projects. Attention is restricted for reasons of space to Ibibio (Lower Cross, Niger-Congo, Nigeria) and Mandarin, as typical representatives of two major ‘lexical tone’ language types.

### 1.1. Descriptive preliminaries

In a cross-linguistic perspective, both the organisation of pitch forms and their functionality are highly versatile, and can be summarised in a straightforward manner with a Jespersenian functionalistic rank-based language architecture [Jespersen 1924]; cf. Table 1.

Pitch functionality at the phonological and morphological ranks (which determine word prosody) is usually referred to as *tone*, and at the syntactic, text and dialogue levels (which determine discourse prosody) as *intonation*.

Table 1: Pitch-functional rank levels.

<i>Phonology</i>	phonemic ‘lexical tone’
<i>Morphology</i>	morphemic sub-rank: - morphophonemic ‘lexical tone’ derived word sub-rank: - morphological templatic tone: compound word sub-rank: - tonal interfixation inflectional sub-rank: - morphosyntax
<i>Syntax</i>	phrase sub-rank: - templatic phrasal intonation sentence sub-rank: - sentence intonation: phrasing, accentuation, nucleus
<i>Text</i>	textual ‘paragraph’ intonation: cohesive pitch contours, focal and contrastive accentuation
<i>Dialogue</i>	dialogue control, emotion

### 1.2. Formal preliminaries

It is known that—at least for the organisation of forms—the basic structure of pitch systems can be modelled by regular (linear) languages and regular (linear) grammars (equivalently: finite state automata, FSAs). Finite State (FS) modelling holds even where hierarchies are involved in the organisation of pitch systems: the hierarchies put forward so far are either of finite depth or are purely right-branching or left-branching (but cf. [Steedman 1991]) and thus formally FS-equivalent.

Explicit applications of FSAs to modelling intonation forms have been available since [Reich 1969], the 1970s IPO model [‘t Hart & Cohen 1973], [Pierrehumbert 1980] and [Gibbon 1981]. Implicitly, many other intonation models, including [Fujisaki 1988], are also FS models. In this paper, FS intonation patterning will not be dealt with further except in the context of interface issues.

FST modelling of the tone-phonetics interface started with a model of two Niger-Congo languages (Baule, Kwa; Tem, Gur) in [Gibbon 1987] (cf. also [Gibbon 2001]). The technique was extended to Mandarin tonal sandhi, mapping lexical tone sequences to other lexical tone sequences, in [Jansche 1998]. These models will be discussed below.

## 2. Finite-State tonal interfaces

### 2.1. FS syntax-prosody interface (intonation)

Basic FS techniques have also been used for modelling the syntax-prosody interface. Perhaps the first contribution to FS syntax-prosody interface modelling was hinted at in [Chomsky 1965] in a discussion of criteria for performance models, where he points out that right-branching sentences such as *I called up the man who wrote the book that you told*

me about are more acceptable than centre-embedded sentences such as *The man who the boy who the students recognised pointed out is a friend of mine* (p. 10f.). In Chomsky's later work prosodic patterns were derived from arbitrarily embedded nested structures, leading to unrealistic numbers of distinctions between 'stress' levels [Chomsky & Halle 1968].

The first explicit discussion of the syntax-prosody relation in FS terms is by Reich [Reich 1969], who criticises Chomsky's use of overly complex grammar types, and proposes:

1. that English syntax tolerates only one level of centre-embedding and otherwise left and right branching,
2. that therefore FS devices are adequate for modelling English syntax, and
3. that the essentially iterative structure is marked by intonation (p. 840):  
 "[...] the taking of a loop [...] is marked by a distinctive intonation pattern, the rising non-terminal contour. [...] When the loop is not taken, the characteristic falling terminal contour results."

Reich thus models intonation patterns of the type *rise\* rise|fall* (right or left branching) for cases such as the following:

1. Right-branching: *The cow tossed the dog that worried the cat that killed the rat that ate the malt*
2. Left-branching: *Henry is Doug's father's second wife's sister's daughter's husband*

The distinction was further developed in the 'readjustment rule discussion' (an older term for the syntax-prosody interface discussion), particularly in [Bierwisch 1966], who first pointed out the need for flattening syntactic structures, and then in [Culicover & Rochemont 1983]. These approaches developed algorithms for flattening syntactic structures. Many applications of this 'flatter prosody principle' to speech synthesis have been made, e.g. [Campbell 1993] and [Wagner 2000], the latter explicitly using FS methods.

### 2.2. Tonemic sandhi interface (tone)

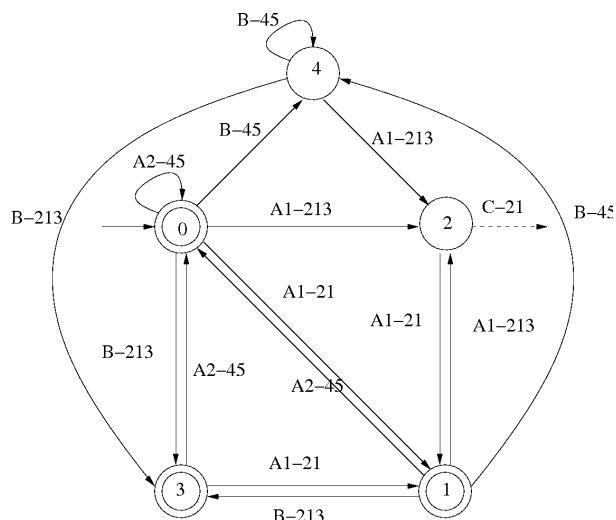


Figure 1: Mandarin tone FST (from [Jansche 1998]).

Taking a new approach to the computational prosody of Mandarin, Jansche presents a Finite State Transducer (FST) model of tone sandhi in the Tianjin variety, shown in Figure 1. The tone sequences of canonical lexical forms are mapped to other tone sequences; both input and output tones are from the lexical tone inventory. Mandarin sandhi mapping is therefore not a phonetic mapping to allotones, but substitutes lexical phonemic tones in the environment of other lexical phonemic tones.

Formally, the Tianjin Mandarin FST can be analysed into a set of relatively unrelated and idiosyncratic FSTs. The overall FST models the union of the regular relations which are modelled by the individual FSTs [Kaplan & Kay 1994].

### 2.3. Morphotonemic-phonetic interface (tone)

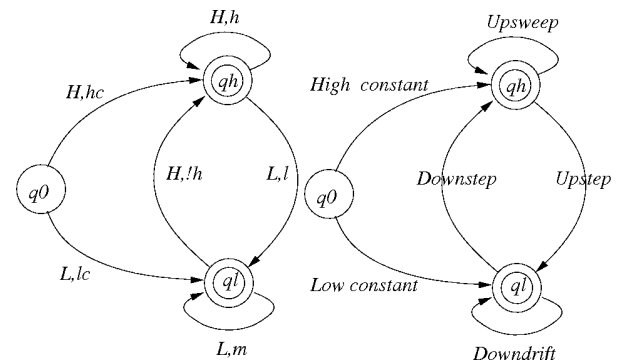


Figure 2: (a) Basic 2-tone Niger-Congo FST; (b) Generalisation of tone FST mapping types ([Gibbon 2001]).

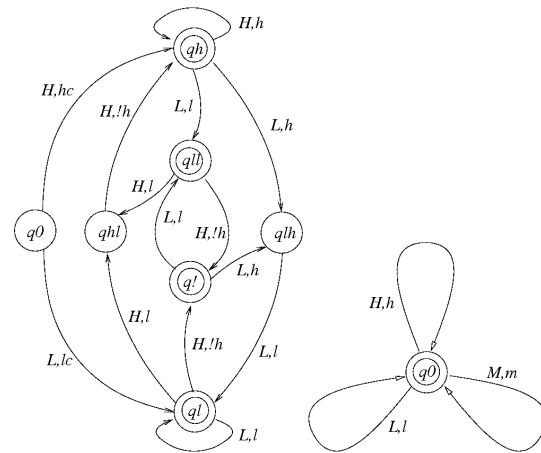


Figure 3: Variants: Baule FST (with lookahead) and 3-tone FST.

The complex tonal structure and functionality of Western and Central African languages has often appeared in the literature: tone terracing vs. discrete level tone patterns, automatic and lexical downstep, upstep, downdrift and upsweep, tonal blocking, tone-depressor consonants. Of specific interest here is tone terracing. Tone terracing is a phonemic-phonetic mapping, and was modelled in Metrical Phonology by right-branching trees [Clements 1981]. Since right-branching trees are accessible to FS modelling, [Gibbon 1987] concluded that

tone terrace mapping is formalisable with FSTs, and provided FST models of terracing in Baule (Kwa, Ivory Coast; cf. Figure 3(a) for a locally non-deterministic FST with 1-place lookahead) and in Tem (Gur, Togo). In [Gibbon 2001] these models were generalised to a schema for any two-tone terraced system; cf. Figure 2(b). Figure 3(b) shows the simple FST model required for 3-tone discrete-level tone systems.

Formally, the basic two-tone FST is the union of just two simpler FSTs, one starting with high tones and one with low tones. The simpler FSTs are isomorphic but for the labelling.

The topology of the African tone FSTs shown in Figures 2 and 3 has very general and symmetrical properties, and is thus quite different from the more idiosyncratic topology of the Mandarin tone FST of Figure 1, in addition to the difference in levels of representation. It should be noted that this kind of explicit tone grammar modelling has not yet found its way into tone description in ‘mainstream phonology’ (cf. the Handbook contributions of Odden and Yip in [Goldsmith 1995], and many later publications).

## 2.4. Phonetic-acoustic interface (tone)

For modelling the pitch time function for high and low tone sequences, an asymptotic function similar to that of [Lieberman & Pierrehumbert 1984] was used (the reference unit for pitch association is the syllable):

$$pitch_{i+1} = tone * (pitch_i - baseline) + baseline$$

with speaker-specific initial, *baseline* and *tone* values, where  $tone < 1$  for low tones,  $tone > 1$  for high tones. The model is linear and local. Using the FSTs shown in Figures 2 and 3 it is straightforward to implement a transducer in which the phonetic output symbols are replaced with the appropriately instantiated numerical functions. A ‘toy instantiation’ can be illustrated informally (not in the detail required for Ibibio) as follows:

High tone factor:	1.1
Low tone factor:	0.8
Downstep factor:	1.3
Upstep factor:	0.6
Baseline component	100 Hz
Initial high component:	80 Hz
Initial low component:	80 Hz
Input (tones):	H L H L
Output (Hz):	180 148 162 137

The ‘real world’ empirical basis for the actual Ibibio model was induced by an exhaustive prosodic data mining algorithm applied to Ibibio data [Gibbon, Urua & Gut 2003].

## 3. Morphophonemic and morphosyntactic tone

The previous section dealt with form-form interfaces; the present section deals with morphotonology at the syntax-morphology interface. Such factors are frequently referred to in passing in the literature, but, like metrical trees, never provided with a grammar model. The morphotonological factors will be illustrated with Ibibio, South-Eastern Nigeria, with the fourth largest Nigerian language population, classified as a Lower Cross language.

The factors involved in Ibibio morphotonology (simplifying for brevity of presentation) are:

1. Part of Speech (POS): there are four tonological categories determined by POS in Ibibio [Essien 1990, Urua 2000]:

- (a) Nouns: lexical tone with phonemic functionality,

comparable with tone in East Asian languages; òbù ‘crayfish’ - óbù ‘dust’.

- (b) Verbs: Fixed tonal templates, modifiable by inflexion and verb subcategorisation.
- (c) Autonomous tonal function morphemes: HL meaning ‘proximate future/past’ and LH meaning ‘non-proximate future/past’, with the tense prefixes *yaa* and *maa*, respectively, e.g. *n-yaa-ka* ‘I will go (sometime)’
- (d) Composition template morphemes: in word-formation, superimposed patterns which function as ‘interfixes’: *èno* ‘gift’, *àbàsi* ‘God’ form a compound: *èno* + high Tone + *àbàsi* → *ènoábàsi* (cf. the interfix function in German: *Liebesbrief*).
- (e) Templatic function words and affixes: determiners, which are NP-initial, tend to have the same pattern, high-low, while quantifiers, which are NP-final, tend to have the opposite pattern, low-high.

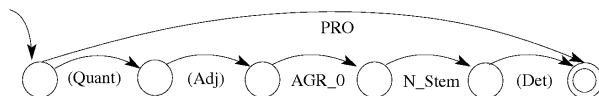


Figure 4: Ibibio Noun Phrase FST.

The autonomous tonal function morphemes and the tones of templatic function words and affixes are particularly interesting from the FS processing point of view: the sub-sentential structures which they mark are non-hierarchical, and can be modelled with FS devices; cf. Figure 4. The symbols in the morphosyntactic FST stand for pairs <POS, TONES> of a part of speech (POS) and the appropriate set of tones for each POS. Consequently, a full description of tonal morphosyntax requires three aligned levels of representation (tiers, tapes). This architecture has not yet been implemented, but two models are being considered: either Kay’s multi-tape FSTs for Arabic templatic morphology [Kay 1987], or a cascade of a morphosyntactic FST as in Figure 4 and a terracing FST as in Figure 2, which can be composed for efficiency into a single FST [Kaplan & Kay 1994].

## 4. Conclusion

Table 2: East Asian / African pitch system comparison.

Level	Mandarin	Ibibio
Phonetic:	+	+
Phonemic:	+	
Morphemic:		+
Morphosyntactic:		+

It was shown that form, structure and function of pitch systems of Western and Central African languages differ fundamentally from those of East Asian languages. Ibibio was selected as a typical Niger-Congo language, with a wide range of morphotonological features which are known to be characteristic of these languages. Jansche’s FST for Mandarin tone sandhi was selected because it is the most explicit and detailed formal

account of this mapping. The differences are shown in Table 2. In addition, both language types have differences in intonation and accent, and Mandarin tones also have characteristic phonetic mappings (not dealt with here).

From this result two relevant conclusions can be drawn. First, on the descriptive level, Mandarin has the expected phonemic lexical tone systems, while the tone system of Ibibio is better classified as morphophonemic and morphotactic than phonemic. Second, because of this, the languages are typologically very different, and it is not possible to apply results of work in accent–intonation languages or phonemic lexical tone languages directly to applications in speech technology for West and Central African languages.

For applications in speech synthesis, the results obtained in this study provide at least some of the lacking specifications for both resource and system development:

1. Resource development: in corpus design and markup for unit selection (whether at diphone or higher unit levels) the tone–relevant morphophonemic and morphosyntactic contexts have to be included.
2. System development: in the linguistic TTS component, appropriate tone–oriented parsing and tagging is required as a basis for unit selection search and costing.

The solutions outlined here are linguistic results. The engineering application of these results in the *Local Languages Speech Technology Initiative (LLSTI)* project and the evaluation of these results are described in [Tucker & Shalanova 2005].

## 5. References

- [Bierwisch 1966] Bierwisch, Manfred (1966): Regeln für die Intonation deutscher Sätze. Untersuchungen über Akzent und Intonation im Deutschen. *Studia Grammatica* VII. Akademie-Verlag, Berlin, pp. 99-199.
- [Campbell 1993] Campbell, Nicholas (1993). Multi-level Timing in Speech. Doctoral thesis, Sussex University, U.K.
- [Clements 1981] Clements, G. N. (1981). The hierarchical representation of tone features. *Harvard Studies in Phonology* 2. Indiana University Linguistics Club.
- [Chomsky 1965] Chomsky, Noam (1965). *Aspects of the Theory of Syntax*. Cambridge MA: MIT Press.
- [Chomsky & Halle 1968] Chomsky, Noam & Morris Halle (1968). *The Sound Pattern of English*. New York: Harper & Row.
- [Culicover & Rochemont 1983] Culicover, Peter W., and Michael Rochemont (1983). Stress and focus in English. *Language* 59(1), pp. 123-165.
- [Essien 1990] Essien, Okon E. (1990). *A Grammar of the Ibibio Language*. Ibadan: University Press Limited.
- [Fujisaki 1988] Hiroya Fujisaki (1988). A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour. In Osamu Fujimura, ed., *Vocal physiology: Voice production, mechanisms and functions*, pp. 347-355. New York: Raven.
- [Gibbon 1981] Gibbon, Dafydd (1981). A new look at intonation syntax and semantics. In A. James & P. Westney, eds., *New Linguistic Impulses in Foreign Language Teaching*. Tbingen, Narr, pp. 71-98.
- [Gibbon 1987] Gibbon, Dafydd (1987). Finite State Processing of Tone Systems. *Proceedings of the European Chapter of the ACL*, Copenhagen, pp. 291-297.
- [Gibbon, Urua & Gut 2003] A computational model of low tones in Ibibio. *Proceedings of the International Congress of Phonetic Sciences*, Barcelona, 2003, pp. 623-626.
- [Gibbon 2001] Gibbon, Dafydd (2001). Finite State Prosodic Analysis of African Corpus Resources. 7th EU-ROSPEECH Conference, Aalborg, Denmark, pp. 83-86.
- [Goldsmith 1995] Goldsmith, John A., ed. (1995). *The Handbook of Phonological Theory*. Oxford: Blackwell.
- [Hart & Cohen 1973] Hart, Johan & Antonie Cohen (1973). Intonation by rule, a perceptual quest. *Journal of Phonetics* 1, p. 309-327.
- [Jansche 1998] Jansche, Martin (1998). A Two-level Take on Tianjin Tone. In: Gert-Jan Kruijff & Ivana Kruijff-Korbayová, eds., *Proceedings of the Third ESSLLI Student Session*, 10th European Summer School on Logic, Language and Information, Saarbrücken, Germany, pp. 162-174.
- [Jespersen 1924] Jespersen, Otto (1924). *The philosophy of grammar*. London: Allen & Unwin.
- [Kay 1987] Kay, Martin (1987). Nonconcatenative Finite-State Morphology. *Proceedings of the European ACL Conference*, Copenhagen, pp. 2-10
- [Kaplan & Kay 1994] Kaplan, Ronald M. and Martin Kay. 1994. Regular Models of Phonological Rule Systems. *Computational Linguistics*, 20:3, pp 331-378.
- [Lieberman & Pierrehumbert 1984] Lieberman, Mark & Janet Pierrehumbert (1984). Intonational invariance under changes in pitch range and length." In: Mark Aronoff and Richard T. Oehrle (eds.), *Language Sound Structure*. Cambridge: MIT Press. pp. 157-233.
- [Pierrehumbert 1980] Pierrehumbert, Janet (1980). The phonology and phonetics of English intonation. Diss. Massachusetts Institut of Technology.
- [Reich 1969] Reich, Peter A. (1969). The finiteness of natural language. *Language*, 45, 831-843.
- [Steedman 1991] Steedman, Mark (1991). Syntax, Intonation and 'Focus', (1991) In Ewan Klein and Frank Veltman, (eds.), *Natural Language and Speech*. Mahwah, N.J.: Lawrence Erlbaum Associates, pp. 331-342.
- [Tucker & Shalanova 2005] Tucker, Roger & Shalanova, Ksenia (2005). Supporting the creation of TTS for local language voice information systems, INTERSPEECH-2005, pp. 453-456.
- [Urua 2000] Urua, Eno-Abasi (2000). *Ibibio Phonetics and Phonology*. Cape Town: Centre for Advanced Studies of African Society.
- [Wagner 2000] Wagner, Petra (2000). Rhythmical Alternations in German Read Speech. *Proceedings of the Prosody 2000 Conference*. Krakow, Poland.

## 6. Acknowledgments

For much discussion and joint work on problems of tonal typology we are also heavily indebted to Shu-Chuan Tseng, Academia Sinica, Taipei, Taiwan, and to Firmin Ahoua, Université de Cocody, Abidjan, Côte d'Ivoire.