A computational model of low tones in Ibibio

Eno-Abasi Urua†, Dafydd Gibbon‡ and Ulrike Gut‡

† University of Uyo, Nigeria

‡ Bielefeld University, Germany

E-mail: enourua@skannet.com, gibbon@spectrum.uni-bielefeld.de, gut@spectrum.uni-bielefeld.de

ABSTRACT

This study reports on a production experiment and a new computational model in an attempt to resolve the question of the status of floating Low tones in Ibibio. The pitch height of high tones (H), low tones (L) and downstepped high (!H) tones in a set of utterances spoken by an Ibibio speaker were first measured acoustically and secondly modelled computationally. Results show that there is a difference between a floating L and an overt L. Furthermore, the overt Low tone causes a higher degree of lowering than the floating Low tone.

1. INTRODUCTION

Ibibio is a Lower Cross (Delta Cross) language of the (New) Benue-Congo family of languages and is spoken by over four million people in Akwa Ibom State and to a lesser extent in Cross River State of Nigeria. Ibibio is a classic example of a terraced level tone language [12] and has been shown to have two level tones, High and Low and a downstepped High tone. The downstepped High tone is usually preceded by a High tone. Other tonal realisations are the High-Low and Low-High contour tones, which are synchronically phonetic combinations of the level tones and may diachronically be from lowering and raising of the level tones, High and Low, respectively by the oral stop consonants, syllable and segment loss, tonal assimilation, melody levels, etc.

The *tonal downstep* pitch relation in tone languages has been of interest to a number of researchers for many years. [13] and [10] have argued that tonal downstep occurs as a result of the presence of a floating or lost Low tone. A distinction between *automatic downstep* and *non-automatic downstep* has been made in the literature. Automatic downstep refers to the situation where non-initial High tones are lowered by intervening overt Low tones, and has been computationally modelled by [2] using finite state automata, while non-automatic downstep refers to the situation where a sequence of (generally non-initial) High tones is lowered without the intervention of any overt Low tone. There is both diachronic and synchronic evidence to support the postulation of 'floating' ('diacritic', 'covert', 'lost') Low tones. In Ibibio itself, there is abundant synchronic evidence from lexical and phrasal constructions supporting the postulation of a floating Low tone as the source of non-automatic lowering.

The instrumental acoustic support to determine one way or the other the degree of lowering of an overt Low tone and a floating Low is sparse. [8], in a study of Igbo (a Benue-Congo language spoken in south-east Nigeria) tones, conclude that the degree of lowering by automatic lowering, (i.e. with overt Low tone) is greater than that of non-automatic lowering, (i.e. with floating Low tone). A similar, though tentative result is also reported for Ibibio in [11]. On the other hand, [6] reports that both automatic and non-automatic lowering exhibit the same degree of lowering phonetically. [9] reports that "… in Bimoba a High tone that is downstepped by a floating Low tone *is* phonetically indistinguishable from one that is downstepped by a non-floating Low tone." Obviously, although they support the fact of lowering, these instrumental studies differ widely in the degree of lowering caused by overt and floating L tones. It would be tempting to argue that different languages map the degree of lowering differently. However, [6] and [7] both investigated the same language, Igbo, and arrived at different conclusions, so this argument may not hold. Other factors, including experimental factors and individual variation, may be involved in the different results obtained.

In this study, we ask the following questions:

1. Is an overt Low tone in Ibibio similar to a floating Low tone?
2. Do they cause the same degree of lowering?
3. What is the value and what is the duration of the floating Low tone?

2. METHOD

This study is based on a production experiment and uses a method which is both empirical and simulative, in that the F_0 measurement sequences assigned to productions are simulated by a computational model.

The data used for this study were produced by an adult female Ibibio speaker, who has lived the better part of her life in Ibibioland. The variety is that spoken in the Uyo/Uruan areas. The prompts consisted of forty-five utterances with various tonal patterns. The relevant utterances were those with a sequence of alternating High and Low tones and those with downstepped High tones. High tone sequences and Low tone sequences were also added for control. The forty-five utterances were shuffled to produce five different random patterns, making a total of one hundred and thirty-five tokens per tone pattern.

The recording was made with a Sony TCD-D7 digital audio tape-corder, with Sennheiser condenser microphones, in the recording studio of Bielefeld University. The F_0 measurements were made with 'Praat'. F_0 measurements were taken at three points in the vowel, the beginning, the mid point and the end point, and were averaged. We use L, H and !H to represent phonetic Low, phonetic High, and Downstepped High tones, respectively.

The computational model was designed in several stages, starting on the lines of the model introduced by [7]. We investigated several models, including overlay and sequential types, and came to the conclusion that overlay and sequential models are formally equivalent (cf. [3]). The final model contains a *reference line* of variable slope (exponential or asymptotic), and a multiplicative *phonetic tone factor* for each phonetic tone, which defines the distance between the tone and the reference line, the factor being in general >1 for phonetic H tones and <1 for phonetic L tones. The phonetic tone factor and the reference line are the relevant parameters of the model; for each of them, a range of possible values, with variable granularity, is defined. The parameters define a search space for the 'best simulation'. The best fit with the data is determined by the Average Magnitude Difference (AMD) function $(\Sigma_i \ |X_i \ | \ - \ |Y_i|) \ / \ n$, where X_i and Y_i are corresponding measurements and model values, and n is the length of the sequence. The AMD was chosen mainly because the usual correlation functions are misleadingly high for the few data points concerned, but also because of its efficiency: the search space is large, the complexity of the model being exponential in the number of tone factors plus the reference line factor. In order to restrict the search space, initial investigations were made with a rather coarse granularity of model values.

In order to apply the model, each utterance was assigned a tone sequence template containing various combinations of phonetic tone factors, e.g. t_1, t_2, t_3, t_4, t_5 and t_6 with t_1 being an initial H, t_2 a H, t_3 a !H, t_4 a raised L between two Hs, t_5 a L and t_6 a final L. Search for the best fit produced results of the following type:

Data:	246.3	266.85	226.91	222.84	199.3	198.57
Model:	246	266	227	225	197	196
Pattern:	t1	t2	t3	t2	t3	t2

3. RESULTS

The relevant tonal relations are between H and following !H, and between !H and following H as well as between H and L and a following H.

First we modelled 18 of the 46 utterances containing only t_1, t_2, and t_3. Search results were very promising, with an average overall AMD of 1.449 Hz (range from 0.003 Hz to 4.089 Hz). Next, we modelled eight utterances containing a raised L (t_4). Modelling was satisfactory with an average AMD of 5.45 Hz (range from 0.444 Hz to 11.844 Hz). In order to determine the similarity of the raised L and the !H, we then compared the values of t_3 in the first modelling run with t_4 in the second run. The raised L was modelled at an average reference line factor of 0.825 (range from 0.75 Hz to 0.85 Hz), the !H at 1.042 (range 0.94 Hz to 1.17 Hz). The raised L is therefore relatively lower than the !H in their respective utterance environments.

We will illustrate the pitch relations informally by referring to the F_0 measurements, and to pitch differences. For utterance bɨ́p ɔ́!bɔ́ɔ́ŋ ɪ́!nɔ́ sé with a sequence of HH!H!HH we observe that the pitch difference in the first H and !H pair is 39.94 Hz and that the pitch difference in the second H and !H pair is 23.54 Hz, indicating a decrease in range as the utterance progresses (see Table 1). Note that in the shorter utterance ké ú!ké →kú!ké the pitch difference in the H and !H pair, 41.75 Hz, is close to that between the initial H and the initial !H. The third utterance also shows a similar pitch difference in a H and !H pair, which in this case is 38.16 Hz.

Table 1. Fo measurements of three Ibibio utterances containing H !H sequences

utt. 1) bɨ́p ɔ́!bɔ́ɔ́ŋ ɪ́!nɔ́ sé → bɨ́β ɔ́!bɔ́ɔ́ŋ ɪ́!nɔ́ se				
Tone:	H	H	!H	H
F0:	246.30	266.85	226.91	222.84
Diff:			39.94	4.07
utt. 2) ké ú!ké →kú!ké				
Tones:	H	!H		
F0:	282.84	241.09		
Diff:		41.75		
utt. 3) kpé ḿ!bɔ́k dàkká				
Tone:	H	H	!H	H
F0:	287,02	279,30	241,14	191,86
Diff:			38.16	49.28

In utterances with overt Low tones alternating between High tones, we observe that the difference between a H and a following L is higher than that between a H and a !H (62.27 Hz in Table 2, utterance 4). This may be as a result of the prevocalic consonant /k/, which is voiceless and may have the effect of raising the F_0 value of the pitch. However, it clearly shows that there is a difference in F_0 value between a floating Low tone and an overt Low tone. Where for instance the difference between a H and a !H in utterance 3 (Table 1) is 38.16 Hz, here we find that the difference between a H and the following L is 62.27 Hz. In addition, we observe that subsequent H´s are also progressively lowered.

The Fo measurements of utterance 6 are approximately what we found in utterance 4 with a difference of 67.46 Hz. It shows that the difference in F_0 values for a !H is less than that of an overt Low tone. A comparison with the disyllabic utterance 2 (Table 1) confirms this difference There the difference between a H and a !H is 41.75 Hz. Although the final H is higher in pitch than the preceding L at 214.10 Hz, it does not come anywhere near the pitch of the initial H of 275.70 Hz.

Table 2. Fo measurements of three Ibibio utterances containing H L sequences

utt. 4) ké ùkɔ̀bɔ́ ìtá → kúù kɔ̀ bîì tá

Tone:	H	L	HL	H
Fo:	283,83	221,56	223,73	207,49
Diff:		62.27	-2.17	16.24

utt. 5) dép ùbóm ké ákpá → dé βù bóm ké kpá

Tone:	H	L	H	H
Fo:	276,44	265,13	229,24	233,27
Diff:		11.31	35.89	-4.03

utt. 6) ké ŋ̀ké

Tone:	H	L	H
Fo:	275,70	208,24	214,10
Diff:		67.46	-5.86

In sequences of identical tones, a gradual but progressive lowering of successive Low tones was found. In sequences of High tones, the overall trend is rather level. However, in one utterance, the initial H was lower than subsequent occurrences of H. Further, the fourth H was considerably higher than the others, perhaps because of the voiceless obstruent /k/. The final H was somewhat lower, though not as low as the initial H, indicating a general downtrend, most likely.

The modelling technique reported in the previous literature (so far as descriptions of method are explicit enough to permit this kind of conclusion to be drawn) is to average paradigmatically over all the frequencies for one particular tone and then fit these syntagmatically to an overall pattern. This is not our approach. Rather, our technique is closer to the pattern recognition techniques used in speech technology, in that each individual utterance is optimally modelled syntagmatically, and paradigmatic generalisations are made subsequently. It is conceivable that the two approaches would yield the same or similar results, but this is by no means a necessary conclusion.

The tone factor mappings are based on the initial observations as follows:

T_1	#H, i.e. initial H	T_4	^L, i.e. L → ^L / H__H
T_2	H	T_5	L
T_3	!H, i.e. downstepped H	T_6	L#, i.e. final L

Using patterns containing only the H tones (#H, H, !H), eighteen models were generated. The best fitting models for the first three cases are shown in detail (decimal places in the models have been truncated):

Data:	246.3	266.85	226.91	222.84	199.3	198.57
Model:	246	266	227	225	197	196
Pattern:	t1	t2	t3	t2	t3	t2

Data:	246.74	253.2	229.52	221.58	203.04	209.5
Model:	246	253	229	223	204	200
Pattern:	t1	t2	t3	t2	t3	t2

Data:	260.13	256.13	236.37	222.76	207.47	209.34
Model:	259	256	236	226	210	202
Pattern:	t1	t2	t3	t2	t3	t2

The following generalisations were made on the basis of the eighteen models: The averaged optimal baseline (an additive component in the model) is 105 Hz. Whether this baseline has a physiological correlate (e.g. as lowest frequency attainable by the speaker) was not investigated. The mean range covered by the reference line is 163 Hz. The mean slope factor for the reference line is 0.87, and the phonetic tone factors are: #H = 0.95, H = 1.13, !H = 1.04. The average pitch difference between the data and the models was 1.45 Hz.

Using patterns containing mixtures of H and L tones eighteen models were generated. The following generalizations were made on the basis of the eight models: The averaged optimal baseline is 110 Hz. The mean range covered by the reference line is 155 Hz. The mean slope factor for the reference line is 0.869, and the phonetic tone factors are: #H = 0.99, H = 1.083, raised L = 0.838, L = 0.617. The average pitch difference between the data and the models was 4.753 Hz.

4. DISCUSSION

The following general tendencies were noted:

- The F_0 values show that there is a difference between a floating Low tone and an overt Low tone. Floating Low tones seem to have lower F_0 values than overt Low tones.

- Initial Highs in a H sequence generally have a lower F_0 value than a second and sometimes third High, except in a short utterance, e.g. disyllabic utterance, where the initial H has a higher F_0 value than the second H.

- The claim that automatic downstep exists in Ibibio, i.e. that in a sequence of H´s and L´s, H´s following an L do not go back to the level of earlier H´s is confirmed by the data. The subsequent H´s may be slightly higher than preceding L´s but not as high as preceding H´s.

- The claim that non-automatic downstep exists in Ibibio, i.e. that after a !H tone, a subsequent High does not go back to the level of a preceding H, is also substantiated.

- It does appear that consonants have some effect on Ibibio tones. The pattern shows that voiced oral stops seem to have a depressing effect on pitch, while voiceless oral stop consonants appear to raise the pitch. Of course, Ibibio has no mid tones to actually test whether the voiceless stops actually have a raising effect on the pitch of tones.

- A sequence of H´s show a gradual downdrift/declination. Obviously utterance length is likely to be critical in the rate of downdrift. Shorter utterances exhibit more dramatic differences while longer utterances show a more gradual pattern.

- Final L´s show a falling pattern. This pattern is different from those of L´s in initial and non-final positions.

An additional factor in determining the realisation of tones, which has been noted particularly by those involved in experimental work on tone, is the effect of consonants on pitch in both tone and non-tone languages. [5] documents various experiments to show that certain consonants have some effect on tone realisation. Usually it is the prevocalic consonants that appear to have a more pronounced effect on tone than postvocalic ones. The studies have shown that voiced and unvoiced consonants, especially the obstruents, may have a depressing and raising effect on tones respectively. Implosives are reported not to have a depressing effect on F_0. [1] associates a Low tone with the presence of a voiced obstruent although the summary of data she presented seems to suggest that voiceless obstruents do not have any effect on tones.

The results of this study show that voiced stops in Ibibio depress the F_0 values of neighbouring pitches. Let us consider some phonological data in addition from Ibibio:

a)	ú!ké	‘where’	d)	ḿ!bɔ́k	‘please’
b)	ú!dó	‘that way’	e)	ú!mí	‘this way’
c)	ɔ́!bɔ́ɔ́ŋ	‘chief’	f)	î!núén	‘bird’

In these data, it appears that phonological downstep is generally present where there are voiced obstruents and nasal consonants, apart from the lone example with a voiceless stop consonant in (a). This seems to confirm, for Ibibio, the depressing effect of voiced stops on F_0 values of pitches.

The modelling of the raised L and the !H did not yield conclusive evidence for the assumption that a floating low causes a !H to be as low as a L. However, modelling of the utterance containing L was slightly less accurate than the modelling of the utterance containing !H tones. Further modelling is required.

In conclusion:

- *Is an overt Low tone in Ibibio similar to a floating Low tone?* From the values that we have at the moment, the results seem to show that there is a difference between a floating Low tone and an overt Low tone. However, we must bear in mind that we cannot reach a firm conclusion on the basis of the data. This is a pilot study and more data needs to be collected from more speakers before a firm conclusion can be reached.

- *Do they cause the same degree of lowering*? No, they do not seem to cause the same degree of lowering. This result seems to support the conclusions reached by Liberman, Schultz, Hong and Okeke (1993) and Urua (1996/97), where they show that the degree of lowering is greater by automatic lowering (downdrift triggered by an overt Low tone) than for non-automatic lowering (downstep triggered by a floating Low tone). Obviously the overt Low tone causes a higher degree of lowering than the floating Low tone.

- *What is the value and what is the duration of the floating Low tone?* The overall F0 values for floating Low tones seem to be in the region of 39.00 Hz while that for the overt Low tone appears to be in the average of 60.00 Hz.

- There is a further need to explore the impact of consonants, especially obstruents, on tone realisation in Ibibio.

REFERENCES

[1] M. Bradshaw, Tone and its phonological interaction with consonants. *Paper presented at the Tone Workshop, University of Tromso,* Norway, June 2000.

[2] D. Gibbon, Finite State modelling of tone languages. *COLING 87. Proceedings of the Conference on Computational Linguistics.* Copenhagen, 1987.

[3] U. Gut and D. Gibbon, “Is prosody linear?” *Proceedings of the conference Prosody 2000*, Cracow, pp. 83-88, 2000.

[4] J.-M. Hombert, *Towards a theory of tonogenesis: an empirical, physiologically, and perceptually-based account of the development of tonal contrasts in language*. Unpublished Ph.D dissertation, University of California, Berkeley, 1975.

[5] J.-M. Hombert, Consonant types, vowel quality and tone. V. Fromkin (ed.) *Tone: A Linguistic Survey.* New York: Academic Press, 1978.

[6] Y. Laniran, Phonetic aspects of tone realisation in Igbo. *Progress Reports from Oxford Phonetics* 5: 35-51, 1992.

[7] M. Liberman and J. Pierrehumbert, Intonational invariance under changes in pitch range and length. Aronoff M. and R. Oehrle (eds.) *Language Sound Structure*. 157-233. Cambridge, MA: MIT Press, 1984.

[8] M. Liberman, J. Michael Schultz, S. Hong and V. Okeke, “The phonetic interpretation of Igbo tones.” *Phonetica* 50: 147-160, 1993.

[9] K. Snider, “Phonetic realisation of downstep in Bimoba.” *Phonology* 15: 77-101, 1998.

[10] J. Stewart, “Downstep and floating low tones in Adioukrou.” *Journal of African Languages and Linguistics* 5: 57-78, 1983.

[11] E. E. Urua, “A phonetic analysis of Ibibio tones: a preliminary investigation.” *Journal of West African Languages* XXVI.1: 15-26, 1996/7.

[12] W. E. Welmers, *African Language Structures*. Berkeley: University of California Press, 1973.

[13] F. D. D. Winston, “The ‘mid-tone’ in Efik.” *African Language Studies* 1: 185-192, 1960.