

Probabilistically speaking: a quantitative exploration of Yorùbá speech surrogacy*

Túndé Adégbölá

African Languages Technology Initiative (Alt-I), Nigeria

taintransit@hotmail.com

1. Introduction

Speech surrogacy may be defined as the use of instruments other than the human speech apparatus to mimic aspects of speech. Many cultures of West Africa engage in speech surrogacy to varying degrees. Examples include:

- ? Ashanti, Dagomba and Ga in Ghana
- ? Ewe in Ghana and Togo
- ? Fon in Benin
- ? Yorùbá in Nigeria, Benin and Togo

The Ewe phrase *vu gbe* meaning voice of the drum or drum language aptly expresses the concept of speech surrogacy among the Ewe.

Despite the works of La Velle (1974), Bakare (1975), Connel & Ladd (1990) and Laniran (1992), Yorùbá language can not be said to have enjoyed extensive systematic instrumental study, yet it is still one of the most studied languages of West Africa. This, unfortunately, reflects the fact that the various languages of Africa may not be available to developments in modern information communication technologies (ICTs). In anticipation of the impending information society therefore, there is a need to complement the existing knowledge-base in the literature of Human Language Technology (HLT) with mathematical characterization of

tonality and other peculiar qualities of African languages as has been done for various tone languages of Asia.

In surrogate speech, man-made mechanical devices are used to mimic aspects of human speech in order to achieve communication. Speech surrogacy may therefore be viewed as an age-old, technology-mediated human communication and it may provide some insights into the important role of tones in some of these languages. Hence a quantitative exploration of Yorùbá speech surrogacy is bound to provide frameworks for the mathematical characterisation of tonality.

Yorùbá Speech Surrogacy

In Yorùbá culture, drums and horns as well as plucked and bowed stringed instruments are used as speech surrogates; each of these instruments mimicking aspects of human speech to varying degrees of success. The *Bàtá* drum for example, employs tones in combination with some vocables codes that represent consonants and vowels, but the *Dùndún* (the most popular Yorùbá speech surrogate) and most of the other surrogates employ only tones and rhythm in their mimicry of human speech.

In surrogate speech, the absence of information from the consonantal and vocalic dimensions of the syllabic structure of speech results in ambiguity, and this renders, the *iyáàlù* (the lead drum in most Yorùbá drum ensembles) a very powerful tool of equivocation in the hands of a satirist. This fact is supported by Yorùbá idioms such as *Onílù?tó?tó Ì-pá? r-j-r-j-, ?òun ló mō oun tí òun?¿sō* (it is only the drummer that can say for sure what he employs his powerful drum stick to say) and *Ebi ¿pa mí kò seé fi ifé wí, òmō ¿ relé ló máa jō* (you don't employ whistle speech to say 'I'd like some food' because a reluctant host may chose to translate it as 'I'd like to take my leave').

Accordingly, Beier (1970) observed that “a single (drum) phrase can mean a number of different things but in context, the drum talk is clear to any educated Yorùbá”. Euba (1990) further elaborated on this observation thus:

“An *iyáàlù* is unable to reproduce human speech in exact terms. What it does is to simulate enough of the elements of human speech to make itself intelligible. Of the various parameters of human speech, only two – rhythm and intonation – are reproducible on the *iyáàlù*. Words that have the same intonation in ordinary speech are distinguishable only by their syllabic structure. Since *iyáàlù* cannot enunciate syllables,... contexts must be applied to correctly interpret such words. ... the context must be clear either in terms of ... the words preceding or following it in a sentence”.

Despite this limitation however, Euba (1990) further noted that “the practice of talking with drums has been carried to a high degree of perfection among the Yorùbá”. In ancient times, surrogate speech was generously employed for both religious and secular purposes and today it still forms an important aspect of Yorùbá popular culture. Even contemporary Christian liturgy among the Yorùbá has benefited tremendously from the use of ‘talking drums’ in the praise and worship of the Holy Trinity.

The fact that the Yorùbá have carried the practice of talking with drums to a high degree of perfection despite the limitations of surrogate speech warrants two important directions of investigation. The first direction of investigation should seek to identify any unique linguistic features of the Yorùbá language that enables the Yorùbá to achieve this level of perfection, in contrast to their various neighbours with whom they share many important cultural traits. It is particularly curious that various cultures to the north of Yorùbáland through where the versatile hour-glass tension drum (*dùndún*) is supposed to have come to the Yorùbá (Euba, 1990) do not use surrogate speech at levels as advanced as the Yorùbá. Instead, more distant cultures as far away as in Ghana, with no apparent patterns of contiguity to the Yorùbá homeland use surrogate speech to appreciable extents. The second

direction of investigation should seek to understand (from an information theoretical point of view) how the Yorùbá manage to employ surrogate speech for functional communication despite the ambiguities introduced by the loss of the two important syllabic dimensions of consonants and vowels.

3. A Manipulation of probabilities

An examination of Yorùbá speech surrogacy easily reveals a deliberate manipulation of probabilities. This is done in order to resolve the ambiguities brought about by the loss of information from the consonantal and vocalic dimensions of the syllabic structure of Yorùbá phonology. Three main devices for this manipulation of probabilities are identifiable. These are the development of a special drum vocabulary, the design of a unique two-sentence structure for statements in the drum vocabulary and the deliberate use of rhythm to delimit sentences during rendition. Also apparent is the natural use of contexts for the further manipulation of probabilities.

3.1 The drum vocabulary

The Yorùbá system of speech surrogacy is based on a drum vocabulary of pithy sayings drawn from proverbs, idioms and other literary devices of Yorùbá language and culture. (Abimbola, 1997). Apart from tonal and rhythmic dexterity therefore, *iyáàlù* scholarship involves the learning of these pithy sayings. The more a drummer knows the wider his repertoire of drummable statements. In the larger Yorùbá society too, some knowledge of the drum vocabulary is regarded as a social skill. A Yorùbá who is totally ignorant of the drum vocabulary is regarded as inept. In common parlance, *kò gbùlù*.

3.2 The two-sentence structure

Most of the statements in the Yorùbá drum vocabulary consist of two sentences (Oladapo, 1995), the first sentence being usually shorter than the second. In rendition, the first sentence is produced twice, then the second sentence is produced once, after which the first sentence is produced again as a coda, signifying the end of the statement. For example, a common proverb; **Bí il, bá ẹ̀ gbe òfikà, tí ikú ẹ̀ pa olòótù, bó p,, títí, oore a sù ni í ẹ̀** (generosity becomes wearisome if the wicked continue to flourish while the righteous perish) is rendered in drum talk as:

Oore a sù ni í ẹ̀ (2)
Bí il, ẹ̀ gbòfikà tí kú ẹ̀ p'olòótù
Oore a sù ni í ẹ̀

The quest for economy in the two-sentence structure of the drum vocabulary is quite apparent. It regards the **bó p,, títí** as superfluous and therefore dispensable. It also employs vowel elision to achieve economy.

It must be said however, that even though this two-sentence structure is standard in Yorùbá speech surrogacy, it is freely disregarded under various circumstances. For example, when a learned drummer is 'speaking' to a group of equally learned listeners, the two-sentence structure may be disregarded from time to time. In (Oladapo 1995), most of the statements recorded in the book conform strictly to this two-sentence structure while a few feature various exceptions.

3.3 Delimitation of sentences by the use of rhythm

In rendition, statements in the drum vocabulary are delimited by rhythm in order to aid the listener to perceive the statements in sentences rather than words. The following example

shows how a statement is cast in 6/8 timing with ‘|’ indicating the beginning of a metrical bar.

*Mo|lák,,k,, mo mú regbó ifá, mo| b’ábàjà, mo mú rẹdí àgbòn, àgbun| mì tìtì, mo
lẹ bójà ló r̄, òjò| pagì lápá kan o dápá kan sí, òjò |pa mí, o má lẹ p̄r,, mí.|*

Such a passage will be much more freely delivered by a casual speaker or a performance poet, devoid of the strict metricality imposed on the above rendition as a drum statement.

Furthermore, surrogate speech takes due advantage of vowel elision to enhance metricality in the Yorùbá drum vocabulary. For example the use of *ì gbòlìkà* instead of *ì gbe òlìkà* and *ì p’olòótù* instead of *ì pa olòótù* in the example of 3.2 above serve to address both the quest for economy and enhanced metricality.

3.4 Contexts

There are various other context-based constraints that help to manipulate probabilities towards the resolution of ambiguity. These include word-level contexts created by words adjacency and situational contexts created by the prevailing circumstances in the physical environment of a drum mediated communication.

4. Insights from information theory

From a theoretical point of view, the feasibility of speech surrogacy can be examined based on such concepts as ergodicity and entropy to explain how syntactic and semantic redundancies that characterise natural language can be exploited to elicit information from incomplete cues.

Ergodicity is a character of predictability in non-deterministic systems. Simply put, an ergodic source is one that maintains some level of consistency in the statistical distribution of its output. Natural language is approximately ergodic (Rosenfeld, 2000), hence, speech, texts and other linguistic activities normally exhibit some level of consistency in the

statistical distribution of their basic units. Such basic units include phonemes, consonants and vowels, syllables, tones and words as well as phrases, clauses and even sentences. This implies therefore that it is possible to make predictions of the occurrences (of incidences) of these linguistic units based on a prior knowledge of their statistical distributions.

Entropy on the other hand is a quantitative measure of the minimum amount of cues required to identify the underlying implications (or the meaning) of an information event. Redundancy in language can therefore be measured by entropy in order to determine the minimum cues required to deduce the true meaning of a drum statement.

Redundancy in natural language is widely exploited to elicit information in situations of incomplete cues. As observed by Bird (1998), 'analysis shows that tone marking degrades reading fluency' in Dschang, a Grassfields Bantu language of Cameroon. This implies that some level of reading fluency is possible without tone marking and by extension, that context can be employed to resolve the ambiguities introduced by the absence of tone marking, taking advantage of the redundancies in natural language.

Even though certain key dimensions of the syllabic structure are missing in surrogate speech therefore, the statistical predictability of natural language based on its ergodicity and its inherent redundancies which can be measured by entropy offer scope for the application of surrogate speech in functional communication. Since we can quantify ergodicity and measure entropy, we therefore have a framework within which we can examine speech surrogacy quantitatively.

In order to facilitate the examination of speech surrogacy within a mathematical framework suggested by the above concepts, the following notions would be used as defined below.

- ? A base syllable b_x is a valid syllable devoid of its associated tone¹.
- ? A tonal syllable s_x is a valid syllable with its associated tone²
- ? A tonal skeleton t is the string of tones associated with a string of tonal syllables

Euba's (1990) assertion that the contexts of surrogate speech must be clear either in terms of the words preceding or following it in a sentence can be expressed mathematically based on the principle of the Markov chain.

The Markov Chain expresses the probability of a given event conditioned on the events that precede it. Within this framework, we can express the probability of the occurrence of a tonal syllable s_n conditioned on the occurrences of s_{n-1} , s_{n-2} down to s_1 that precede it as $P(s_n/s_{n-1}, s_{n-2}, \dots, s_1)$. Based on this, we can raise the question: Given a tonal skeleton t , what is the probability that it was instigated by the sequence of tonal syllables s_1, s_2, \dots, s_n ?

This is equivalent to evaluating the probability $P(s_1, s_2, \dots, s_n/t)$. We can therefore explore the conditions that guarantee the resolution of whatever ambiguities that are introduced in speech surrogacy, by examining the conditions under which $P(s_1, s_2, \dots, s_n/t) = 1$.

From Baye's theorem:

$$P(s_1, s_2, \dots, s_n/t) = P(t/s_1, s_2, \dots, s_n)P(s_1, s_2, \dots, s_n)/P(t)$$

But we know that a given sequence of tonal syllables can yield only one tonal skeleton.

Therefore,

$$P(t/s_1, s_2, \dots, s_n) = 1$$

hence the simplification

$$P(s_1, s_2, \dots, s_n/t) = P(s_1, s_2, \dots, s_n)/P(t)$$

Both $P(s_1, s_2, \dots, s_n)$ and $p(t)$ can be obtained by undertaking occurrence counting in a collection of drum statements and we can thereby examine the conditions under which $P(s_1, s_2, \dots, s_n/t) = 1$.

5. Tests and results

Drum statements from the book *Àyàn Akéwì (Afilùs-r̀)* (Oladapo, 1995) were analyzed. 185 drum statements that conform strictly to the two-sentence structure were selected from the book as test data and the following analyses were carried out on them.

The tonal skeleton of the first sentence of each drum statement was extracted and a frequency count of the occurrences of these tonal skeletons in the test data was undertaken. We sought to answer three main questions:

1. How many times did each tonal skeleton occur in the test data?
2. How many different strings of tonal syllables in the test data could have instigated each tonal skeleton?
3. What variables affect the resolution of the ambiguity presented by the tonal skeletons?

Table 1 gives a summary of the answers to these questions for 11 out of the 185 drum statements. In the table, the first column (Index) shows the position of the drum statement in the 185 statements analysed while the second column (Tonal Skeleton) provides the speech melody of the first line of each drum statement. L, M and H refer to the Low, Mid and High tones respectively. The third column (Freq) indicates the frequency of occurrence of each tonal skeleton in the 185 drum statements analysed, the fourth column (Length) is the length of the first line of each drum statement in syllables, while the fifth column (Length at Resolution) indicates the length at which ambiguity was resolved. The sixth column (Actual

Statement) gives the actual statement from which the tonal skeleton in the first column was extracted.

5. Observations

Out of the 185 drum statements in the test data, a total of 118 had first sentences whose tonal skeletons occurred only once. This implies that in 64% of the cases, an interpreter of these tonal skeletons would have been supplied with enough information to identify the tonal syllables that instigated them by the end of the first line.

Of the remaining 67 statements, 20 occurred twice while 11 occurred thrice. Various other occurrence frequencies were also observed as shown in Table 1. On close examination however, it was found that in most of the cases of multiple occurrences, the repeated occurrences were merely embedded within other sentences. The situation changed drastically when we took account of sentence delimitation by rhythm to align the tonal skeletons to the beginning of the sentences that they are being compared with. Under this condition, only one tonal skeleton occurred more than once and the ambiguity was easily resolved by the second sentence.

It was observed that the length of the tonal skeleton plays a significant role in the resolution of ambiguity. On the average, ambiguity is resolved by the 7th syllable. Hence, the condition for $P(s_1, s_2, \dots, s_n/t) = 1$ is $n=7$. Also observed is the fact that the mean length of the first sentence of drum statements in the test data is approximately 7 (6.88) syllables, suggesting that the two-sentence structure of the Yorùbá drum vocabulary is deliberately designed to promote resolution of ambiguity by the end of the first sentence.

6. Conclusion

The average length at which ambiguity is resolved and mean length of the first line of the statements in Yorùbá drum vocabulary suggest a balance between two important principles in information theory and cognitive psychology respectively. The principle of the Markov chain suggests that the longer the sentence, the less the likelihood of ambiguity. However, we know from cognitive psychology that the size of the short term memory of the average person is 7 ± 2 . We may conclude therefore that the Yorùbá drum vocabulary was designed based on the two-sentence structure and sentence delimitation by rhythm so as to resolve ambiguity before human cognitive capacity is overwhelmed.

In answering the question of the 'how' of Yorùbá speech surrogacy therefore, we can conclude that the Yorùbá have managed to take speech surrogacy to a high level of perfection by developing a drum vocabulary, designing a two-sentence-structure for the vocabulary and taking advantage of metricality to delimit sentences.

As for the 'why' of a high level of perfection in Yorùbá speech surrogacy to the exclusion of its neighbouring cultures, we propose as follows: Because Yorùbá uses three lexically contrasting tones, the tonal scope of Yorùbá is such that on the average, ambiguity is resolved in most of the first sentences of our test data by the 7th syllable. A language that employs less than three lexical tones will require more syllables for the resolution of ambiguities created in speech surrogacy. It is instructive to note that a language that is based on two lexical tones system can represent only 128 (2^7) distinct notions in seven syllables, whereas, a language that is based on three lexical tones can represent up to 2,187 (3^7) distinct ideas in seven syllables.

Furthermore, Yorùbá language accommodates a system of vowel elision with associated tone retention characteristics that offers great scope for the enhancement of economy and metricality in the rendition of drum statements. With these facilities, more often than not, ambiguity in drum statements is resolved before human cognitive capacity is overwhelmed. However, investigations in the languages of neighbouring cultures would be needed as control for this proposition.

In conclusion therefore, it can be said that Yorùbá speech surrogacy is based on a manipulation of probabilities for the resolution of ambiguities; hence, talking with drums in the Yorùbá tradition is, in a manner of speaking, probabilistically speaking!

Acknowledgement

³The data for this investigation was made possible by funding from Bait Al-Hikmah for the project “Corpus Development of Computer Readable Yorùbá Texts”. We also wish to acknowledge the support of Tiwa Systems Ltd and Open Society Initiative for West Africa (OSIWA). We thank Prof. Akin Euba for clarifications of various issues in his book “Yorùbá Drumming”. Thanks also to Amanda Vincent for pointing out the exceptional capacity of the Bata to enunciate more than tones.

Appendix

Index	Tonal Skeleton	Freq	Length	Length at Resolution	Actual Statement
1	MHMMHLL	1	7	6	Èlprun lològùn
2	LLHMLM	1	6	6	Kò gbàgbé ùnì kan
3	MHMHHLM	1	7	7	Èlprun rínú èmi
26	MMMHL	2	5		èmö ni ká rà
27	HHHLHHM	3	7		Kí ló tún kù tí ó ƣe
28	LHHMM	5	5		•gún ó gun un
29	MMMMMHH	2	7		Baba ni baba á j,,
35	MHM	55	3		Irù ni
36	MMHHMLMHMM	1	10	8	Ohun tó bá sönù, a ó ri i
44	LMH	16	3		Àpagbé
185	MHHHMMLHLLHMM	1	13	9	Ikú tí ó pa ni, à á yàgò fún un ni

Table 1: First sentences of some Yorùbá drum statements showing frequency of occurrence of their tonal skeleton in the test data, length of first line and length at resolution.

Endnotes

¹ The terms base syllable and tonal syllable are borrowed from T. Demeechai, K. Mäkeläinen (2000)

² For quantitative consistency, we take the position that the Yorùbá mid tone is valid tone in its own right, rather than absence of tone.

References

1. Abimbola, W., (1997) *Ifa Will Mend Our Broken World: Thoughts on Yorùbá Religion and Culture in Africa and the Diaspora*. Aim Books, Massachusetts.
2. Akinlabi, A. (in press) "Yorùbá Sound System," in N. S. Lawal and M. Sadiku (ed.s) Handbook of Yorùbá Life and Culture
3. Bakare, C.A., (1975) Discrimination and Identification of Yorùbá Tones: Perception Experiments and Acoustic Analysis, in Kolawole Owolabi (ed) Language in Nigeria.
4. Beier, U., (1970), *Yorùbá Poetry: An anthology of Traditional poems*, Cambridge University Press. p 12
5. Bird, S., (1998) *When Marking Tone Reduces Fluency: An Orthography Experiment in Cameroon*. University of Edinburgh
6. Connell, B., and Ladd, D.R., (1990) Aspects of Pitch Realisation in ? ?????? in *Phonology Yearbook 7*: 1-30.
7. Demeechai, T. & Makelainen, K., (2000). Recognition of syllables in a tone language, in Speech Communication 33 (2001) 241-254, www.elsevier.nl/locate/specom
8. Euba, A., (1990) ? ??????Drumming: The Dundun Tradition, Bayreuth African Studies Series, Bayreuth University.
9. Hanpanich, S., (1993), *The statistical distribution of consonants, vowels and tones within Thai syllables existing as words and word constituents*

10. LaVelle, C.R., (1974) An Experimental Study of Yorùbá Tones, in Studies in African Linguistics. Supplement 5
11. Laniran, Y.O., (1992) Intonation in Tone Languages: The phonetic Implementation of Tones in Yorùbá??PhD Thesis, Graduate School, Cornell University.
12. Oladapo, O., 1995, Ayan Akewi (Afilusoro), Olatunbosun Records Company, Ibadan.
13. Oyelamin, M., (2001) personal communications supported with demonstrations
14. Rosenfeld, R., (2000), Two decades of statistical language modelling: Where do we go from here? CMU, Pittsburgh.