International Journal of Computing & Information Sciences *Vol. 9, No. 1, April 2011*
Unifying Speech Resources for Tone Languages: A Computational Perspective
Moses E. Ekpenyong, Eno-Abasi E. Urua, Victor J. Ekong, Okure U. Obot and Imelda I. Udoh
Pages 47 - 60

Unifying Speech Resources for Tone Languages: A Computational Perspective

[1]Moses E. Ekpenyong, [2]Eno-Abasi E. Urua, [3]Victor J. Ekong, [4]Okure U. Obot and [5]Imelda I. Udoh

[1,3,4]Department of Computer Science
University of Uyo, NIGERIA
([1][ekpenyong_moses@yahoo.com, mosesekpenyong@gmail.com],
[3]vekong@yahoo.com, [4]abatakure@yahoo.com)
[2,5]Department of Linguistics and Nigerian Languages
University of Uyo, NIGERIA
([2]anemandinyene@yahoo.com,[5]ichejiudoh@yahoo.com)

Abstract: *In this paper, we propose a computational approach to unifying speech resources for tone languages. The paper proceeds in three folds: (i) it provides background and tutorial on one of the speech resources (the Ibibio speech synthesizer), discussing the synthesis development phases based on working experience using the Festival Text-to-Speech (TTS) system; (ii) it suggests ways of sustaining the synthesis development with perspective and infrastructural solution that will not only make the synthesizer more intelligible but also easily replicable for other tone languages; (iii) it introduces the archiving of a new language resource (the talking Medefaidrin Web dictionary). These resources require standardization to make them interoperable for a unified documentation. An architectural design for unification of the resources is finally presented to further this innovation. It is hoped that this paper would spin-off interests and further collaborative ties that will expand our network horizons and ensure a dependable information base for archiving tone language resources.*

Keywords: *Natural language processing, human language technology, HTS synthesis, Ibibio speech synthesizer, unified modeling.*

Received: May 30, 2010 | ***Revised:*** December 25, 2010 | ***Accepted:*** January 31, 2011

1. Introduction

Implementing linguistic knowledge on computers has posed numerous challenges to scientists and linguists, due to the dynamic nature of natural languages. To cope with this dynamic nature, a new field of computer science has emerged. This field known as Natural Language Processing (NLP) is concerned with the interactions between computers and human (natural) languages and has evolved numerous algorithms such as part of speech (POS) tagging and word sense disambiguation, which deal with the ambiguous nature of natural languages. The problems associated with NLP relate to generation and understanding of linguistic features of the processed language, i.e., the ability to model morphology (the structure of words).

Speech processing is becoming increasingly common in the world today. Speech resources benefit four categories of experts. These categories are:

(i) Core tools and data: for speech scientists, phoneticians and phonologists.
(ii) Training and test data: for speech recognition and synthesis researchers
(iii) Corpora and corpora processing tools: for researchers working on dialog systems
(iv) Corpora from spoken domain: computational linguists working with text.

The by-products from these categories (for instance e-dictionaries, translation systems, etc.) are useful to everyone else, most especially the language communities. Otherwise, the fairness of such resources is questionable. Language resource builders must therefore address the issue of fairness [1], as this will enhance the efficiency and longetivity of such resource systems.

2. Speech Synthesis: Background Knowledge

The recent commercialization of robust speech recognition systems and the manifestation of the Web, have generally replaced speech and language processing applications and also revealed a plethora of existing possibilities of applications. Speech synthesis and speaker recognition systems have in recent times moved from rule-based systems to data-based systems. In the data-based approach, the design of the system provides knowledge on the problem, defining appropriate models and features; but the specific features or parameter values are derived from data training. In speech synthesis, this strategy has been applied to all tasks common to other speech and language technologies.

A simplified data-based speech processing procedure is illustrated in Figure 1. The advantage of the data-based approach, herein referred to as corpus-based approach is that once the models and algorithms are ready, they can be ported or applied to create new voices in the same or different language with limited effort. Furthermore, debugging and modification of the models are relatively simple. This can be done without heavily depending on experts who developed the models.

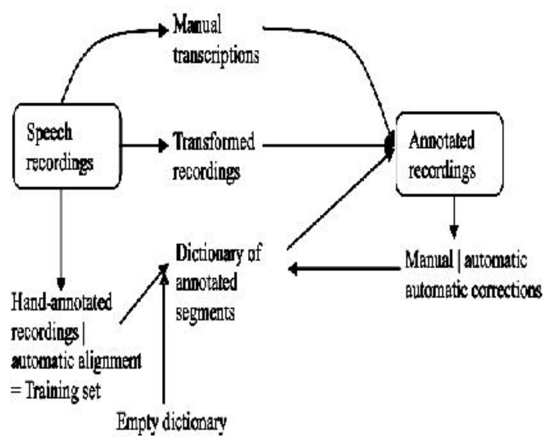


Figure 1. A simplified data-based speech processing

On the other hand, if the corpus does not exist, its development can be very expensive. This makes the portability of the corpus-based approach to languages with limited resources difficult. The lack of available resources has impaired useful research and the use of speech technologies in languages with small or medium data, such as Ibibio. We believe that the production of public resources and making them readily accessible with as few restrictions as possible is the best practice to contribute to the development of language technology.

To achieve a dependable speech resources base, some groups have established collaboration networks. For instance, [2] have introduced the EU-FP7 project CLARIN, a joint effort of 150 institutions in Europe, aimed at the creation of sustainable language resources and technology infrastructure for the humanities and social sciences research community. The project's vision is to turn existing, fragmental language resources and tools into accessible and stable services that any user, not just the technology experts, can access and use for research and development purposes. In [3], they describe the constitution of activities of the International Speech Communication Association (ISCA) Speech Synthesis Special Interest Group SynSIG. The group's goal borders on the provision of collaborative experiments, teaching resources, student exchanges and standardization.

Speech synthesis has received considerable attention, thus attaining a high level of quality naturalness. This achievement is made manifest through effective techniques such as unit selection-based systems [4] or other recent emerging technologies [5], based on the analysis of large speech corpora. The Festival speech system for instance, offers a generic framework for building speech synthesis systems as well as incorporating examples of various modules. It offers full text-to-speech functionalities through a number of Application Interfaces (APIs): from the shell level, through a scheme command interpreter, as a C++ library, from Java and Emacs interface. Festival is multilingual. Though English is the most advanced, other group of new languages including full tools and documentation for building new voices are available at the Carnegie FestVox Project (http://festvox.org).

In [6], a description of the design and production of Catalan database for building synthetic voices is presented. They use two speakers with 10 hours speech recordings per speaker and attempt to improve on the diphone system of festival by extending it using clunit and HTS algorithms. The MultiSyn unit-selection algorithm is implemented in [7] and adapted to Ibibio, an African Tone Language System (ATLS). The authors provide a stepwise approach to adapting festival to a new voice with detailed problems and associated challenges. They also proffer solutions to the challenges with a pilot implementation of the system.

Recently, HMM-based synthesis is becoming popular and has also been integrated into Festival. In HMM synthesis, the database is used to estimate the mathematical models (HMM) used to generate speech parameters which are finally transformed into speech. Though the quality seems slightly degraded when compared with corpus-based synthesis, several advantages of this approach still abound. These advantages are discussed in section 4.2.1.

3. Ibibio Speech synthesis system prototype: a tutorial and development roadmap

3.1 Speech synthesis phases

We present a summary of the Ibibio synthesizer development phases. Ibibio belongs to the Lower Cross subgroup of the (New) Benue-Congo language family widely spoken in the Southeastern region of Nigeria with approximately five million speakers.

The synthesizer is developed using the Festival speech synthesis system. A sample prototype of the synthesizer is available at the local language speech technology initiative consortium web site (http://www.llsti.org) and can be downloaded and tried. We have evaluated the synthesizer and it is found to achieve a fair enough intelligibility. The next development phase is the inclusion of tone to the synthesizer. This will enable us achieve an unrestricted synthesizer with high intelligibility.

3.1.1 Corpus preparation and refinement

Corpus data for a synthesizer could be collated from various sources (news readings, documentaries, textbooks, etc.). In the case of Ibibio, this phase was most challenging because we had no concrete electronic form of an Ibibio corpus and had to develop the corpus from scratch. We obtained a corpus of about 4,500 phrases and also defined a suitable orthographic representation that is compatible with Festival, by formulating the Ibibio SAMPA notations [7]. Table 1 shows the formulated table adopted for implementation.

Table 1. Adopted Ibibio SAMPA

S/no	Graphemes	Sound Type	SAMPA
1	a	Vowel	a
2	aa	Vowel	aa
3	b	Consonant	b
4	d	Consonant	d
5	e	Vowel	e
6	e	Vowel	ee
7	ə	Vowel	@
8	f	Consonant	f
9	gh	Consonant	R
10	h	Consonant	h
11	i	Vowel	i
12	ii	Vowel	ii
13	ị	Vowel	I
14	k	Consonant	k
15	kk	Consonant	kk
16	kp	Consonant	kp
17	m	Consonant	m
18	mm	Consonant	mm
19	n	Consonant	n
20	nn	Consonant	nn
21	ny	Consonant	J
22	ñ	Consonant	N
23	ññ	Consonant	NN
24	ʌ	Vowel	V
25	o	Vowel	oo
26	oo	Vowel	o
27	ọ	Vowel	O
28	ọọ	Vowel	OO
29	p	Consonant	p
30	pp	Consonant	pp
31	s	Consonant	s
32	t	Consonant	t
33	tt	Consonant	tt
34	u	Vowel	u
35	uu	Vowel	uu
36	ụ	Vowel	}
37	w	Consonant	w
38	y	Consonant	j

A phoneme-to-grapheme analyzer [8] was written to strip the corpus of Ibibio diacritic markers (to simplify the front-end) and orthographic/grapheme representations, i.e. replacing the orthographic forms with their SAMPA equivalent.

3.1.2 Selection of phonetically balanced sentences

Should an existing corpus be insufficient to build a synthesizer with all phoneme realizations, one alternative is to resort to a diphone realization process. Consequently, we parsed the corpus through a phonetically balanced sentence generator [9], and obtained 165 diphone rich sentences. These sentences were ported into the Festival utterance file format (utt.scm). The file is the utterance database/lexicon and is required during the unit selection process. Figure 2 shows a sample of the first six sentences. The first segment of the sentence corresponds to the sound file (ibibio_ttscorpus5_001.wav) while the second segment is the selected phonetically rich sentence.

(Ibibio_ttscorpus5_001 "bON akam kuukpa mba.")

(ibibio_ttscorpus5_002 "akefeefeRe ajak ikOt abasi.")

(ibibio_ttscorpus5_003 "eJe amaanam aNwaNa ke mme owo enie ntreubOk ke usVN OmmO keetekeet.")

(ibibio_ttscorpus5_004 "abasi amaasiak usVN ubOkkO OnO nditO isred.")

(ibibio_ttscorpus5_005 "ejIn OmO adiben eJe akaaisaN.")

(Ibibio_ttscorpus5_006 "eteidVN kiNsidi.")

Figure 2. Sample phonetically balanced sentences extracted from the utterance file (utt.scm)

3.1.3 Speaker selection and voice recording

During the selection of a speaker and voice recording the minimum requirements specified in [7] should be considered and the convention defined in Festival when processing the speech files, carefully followed. One who possesses the required skills (for instance, voice stability and stamina) should be selected to do the recordings. The recordings should be carried out in an acoustically damped laboratory with high quality (noise canceling) recording equipment. For our synthesizer development, we used the audio recording laboratory of the Universitaet Bielefeld, Germany.

3.1.4 Speech annotation

The recorded files are then converted into wave forms and annotated. For Ibibio, we annotated according to the following tiers:

(i) Sentence level
(ii) Word level
(iii) Syllable level
(iv) Sound level
(v) Tone level (both underlying and surface tones)

A Praat picture file showing an extract of the pitch pattern and the various annotation tiers for one of the sentences is shown in Figure 3.

3.1.5 Voice integration and synthesis

In Festival, the process of voice integration includes:

(i) *Voice adaptation:* To adapt Festival to a new voice, some scheme files require

modification. One of the major modifications is in the letter-to-sound (LTS) rule (phone set and syllabification modules) file. This file specifies the grammatical structure of a language (i.e. how the vowel and consonant sets and their phonemic combinations are pronounced and how syllables are segmented). We modified the *Nina* voice template, a female Spanish voice by replacing the phone set and syllabification rules therein with Ibibio rules. Figures 4 (a) and 4 (b) shows a snippet of the Scheme code for Ibibio LTS rules:

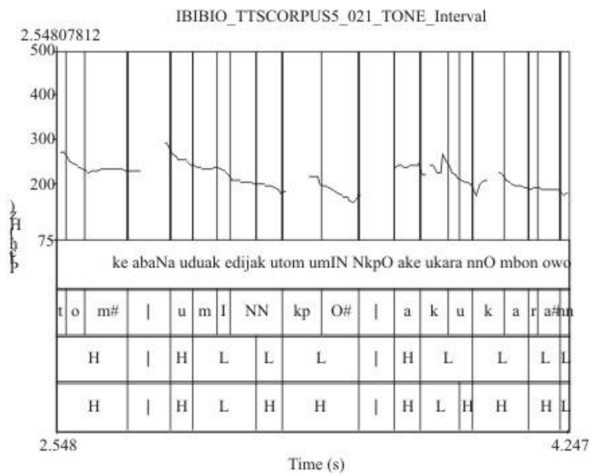


Figure 3. A sample annotated file including underlying tones

```
(;; Sets used in the rules
;; Vowels set
(V a aa e ee I ii o oo u uu O OO @ V U )
;; Consonant set
(C b p pp t tt d f s m mm n nn N NN Nw R
   h j k kk kp r w J )
;; Exception set
(E e ee I) )
( ;; LTS rules  (G-2-P)
;; Some vowel sounds
( [ a a ] = aa );; bring spaced words
   together
( [ a ] = a )
( [ e e ] = ee )
( [ e ] = e )
( [ I I ] = ii )
( [ I ] = I )
( [ I ] = I )
( [ o o ] = oo )
( [ o ] = o )
( [ u u ] = uu )
( [ O ] = O )
( [ V ] = V )
( [ @ ] = @ )
( [ U ] = U )
…
; Some consonant sounds
( [ b ] = b )
( [ h ] = h )
( [ d ] = d )
( [ f ] = f )
( [ R ] = R )
( [ j ] = j )
( [ k p ] = kp )
( [ k k ] = kk )
( [ k ] = k )
( [ m m ] = mm )
( [ m ] = m )
( [ N w ] = Nw )
( [ N N ] = NN )
( [ N ] = N )
( [ J ] = J )
( [ n n ] = nn )
```

Figure 4 (a). A snippet of LTS rule (from the phone set module – lts.scm)

```
(lts.ruleset
  uyo_ibibio_syl
 (  (V a aa e ee i ii I o oo u uu O OO @ V U A)
   (C b p pp t tt d f s m mm n nn N NN Nw R h j
k kk kp r w J ) )
  ;; Rules will add - at syllable boundary
 (  ;; for syllabification convenient (nw -> x, Nw -
> y, I -> c)
    ;; represent long voewls - CV
    ( C [ a a ]  = aa  )
    ( C [ i i ]  = ii )
    ( C [ u u ]  = uu  )
    ( C [ O O ] = OO  )
    ( C [ e e ]  = ee  )
    ( C [ o o ]  = oo  )
;; Represent Diphthongs
    ( C [ a i ]  = ai  )
    ( C [ e i ]  = ei  )
    ( C [ u i ]  = oi  )
    ( C [ O i ] = Oi  )
    ( C [ u i ]  = ui  )
…
;; C-CV
    ( C  [ b ] V =  - b )
    ( C  [ h ] V =  - h )
     ( C  [ d ] V =  - d )
     ( C  [ k p ] V =  - kp )
     ( C  [ k ] V =  - k )
;; V-CV
     ( V  [ b ] V = - b )
    ( V  [ h ] V = - h )
    ( V  [ d ] V = - d )
    ( V  [ k ] V = - k )
    ( V [ p ] V  = - p )
   ( V [ t ] V = - t )
  ( V [ k p ] V = - kp )
```

Figure 4 (b). A snippet of the LTS rule (syllabify module – lts.scm)

(ii) *Voice synthesis:* This process implements the adaptation procedure. The procedure has to do with setting directory structures/paths, utterance building and development of the appropriate target-cost functions. The detail directory structure and utterance building procedure and a design of some finite state transducers (FSTs) that models the tonal typology for African tone

languages are presented in [7].

System compilation: The compilation is done in four phases: generate pitchmarks, generate utterances, generate normalized coefficients and generate Linear Predictive Coding (LPC) coefficients [7].

4. Project Perspective and Sustainability Strategies

4.1 Integrating tones

4.1.1 Finite State (FS) processing of tones

Tone modeling and integration in speech synthesis is 'AI' complete. So far, progress has been made concerning tone synthesizers for Mandarin [10, 11]. In [10] for instance, a Mandarin tone FST is presented. The FST as shown in Figure 5 models tone Sandhi in the Tianjin FST variety. The tone sequences of canonical lexical forms are mapped to other tone sequences; both input and output tones are from the lexical tone inventory. Strategies for tone modeling in Ibibio exists in [12].

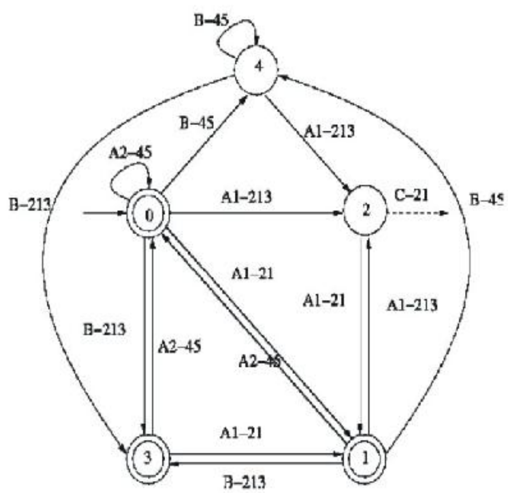


Figure 5 Mandarin FST. Source [10]

It is well known that the basic structure of pitch systems can be modeled by regular (linear) languages and regular (linear) grammars (finite state automata, FSA). Finite state (FS) modeling holds even where hierarchies are involved in pitch systems organization. Some hierarchies so far treated are merely right-branching or left-branching [13-14] and therefore FS-equivalent.

In the phonological modeling of intonation, the research of [15] leads the field of intonational phonology. The work builds on metrical [16] and autosegmental phonology [17-18] as well as Bruce's analysis of Swedish word accents [19].

Pierrehumbert [15] represents an intonational phrase (the largest prosodic constituent) as a sequence of high (H) or low (L) tones. H and L are members of a primary phonological opposition. The tones do not interact with each other but merely follow each other serially in an utterance. Three types of tones were identified in the research:

(i) pitch accents, either as single tones (H*,L*) or bitonal (H*+L, H+L*, L*+H, L+H*). Pitch accents are assigned to prosodic words; the "*" indicates the association and alignment between the tone and the accented syllable of the prosodic word.

(ii) Phrase accents, marked by the "-" symbol (H- , L-). Phrase accents indicate the offset pitch of intermediate phrases and thus control the pitch movement between a pitch accent and a boundary tone.

(iii) Boundary tones, denoted by the "%" symbol (H%,L%). Boundary tones are aligned with the edges of an intonational phrase. The initial and final boundary tones control the onset and offset pitch respectively, of the intonational phrase.

The model thus introduces a three level hierarchy of intonational domains which obey the strict layer hypothesis: An intonational phrase consists of one or more intermediate phrases; each intermediate phrase is composed of one or more prosodic words. The intonation contour of an utterance is described as a series of relative (H and L) tones. Well formed sequences are predicted by the finite state grammar (FSG) [13] in Figure 6.

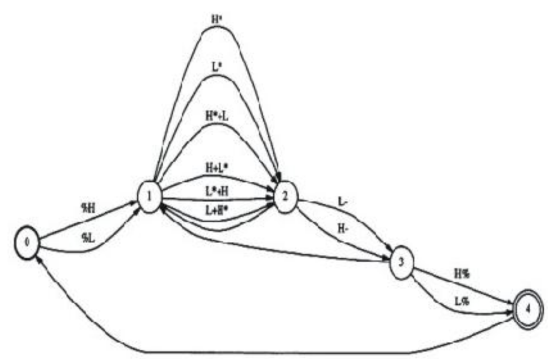


Figure 6. Phonological model with pitch accent, phrase accent and boundary tone labeling. Source [15]

4.1.2 Ibibio prosodic typology: tones

In Ibibio, two contrastive tones occur: the High (H) and Low (L), plus a downstep high (!H) feature. A downstepped H tone may occur after a preceding H tone. Other tonal realizations are High-Low (HL) falling and Low-High (LH) rising contour tones. Each syllable bears either a level tone or one of the contour tones. Thus a disyllabic word may produce the following tone patterns: H-H, H-L, H-!H, L-L and L-H bearing word restrictions [20].

Ibibio possesses the tone-sandhi assimilation, i.e.

produces a characteristic perceptual and experimental effect called tone terracing, i.e., the morphological (lexical) tone patterns are realized in sequence, in terms of phonetic pitch patterns. As observed in Figure 3, the tone sequences are realized at a fairly high level at the beginning of a sequence, and at certain well defined points, the entire pitch register appears to be downstepped to a new level. This process may iteratively occur for several times.

A FST that models morphophonemic and morphosyntactic tone in Ibibio is presented in Figure 8.

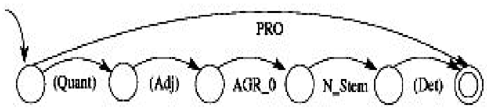


Figure 8. Ibibio noun phrase FST. Source [12]

As observed in [12], there are four tonological categories determined by the part of speech in Ibibio. These categories include

(a) Nouns: lexical tone with phonemic functionality, comparable with tone in East Asian languages; òbû ‘crayfish’ - óbǔ ‘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-yàá-kǎ ‘I will go (sometime)’
(d) Composition template morphemes: in word-formation, superimposed patterns which function as ‘interfixes’: ènò ‘gift’, àbàsì ‘God’ form a compound: ènò + high Tone + àbàsì → ènóábàsì (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.

A non-recursive Ibibio NP and VP FSTs is shown in Figure 9. In this Figure, the networks characterize the two most interesting phrasal units of Ibibio and form the basis for their tonal interactions:

Noun phrases: The noun phrase may be pronominal (PRO), or consist of a sequence of optional quantifier (QUANT), obligatory person and number agreement prefix (AGR 0), optional adjective (ADJ), obligatory noun stem (N STEM) and optional determiner or numeral (DET).

Verbal units: The verb has agglutinating prefixes, and consists of a sequence optional modality (MOD), agreement (AGR 1), tense (TENSE), aspect (ASPECT), a second agreement prefix (AGR 2) conditioned by (second) person and number.

Tonal properties and interactions: A number of elements are relevant for tone assignment.

In Figure 10, a generic model architecture design for a TTS system is presented. A simulation of the terracing FST is carried out in [21]. This design is currently being implemented using the HTS system.

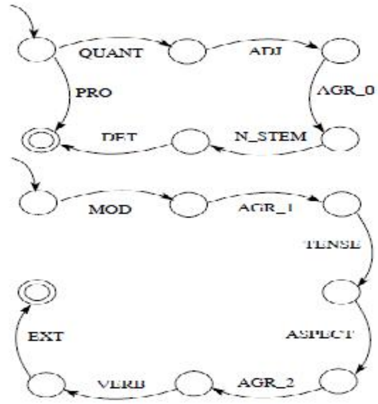


Figure 9. A non-recursive Ibibio NP and VP FSTs. Source [21]

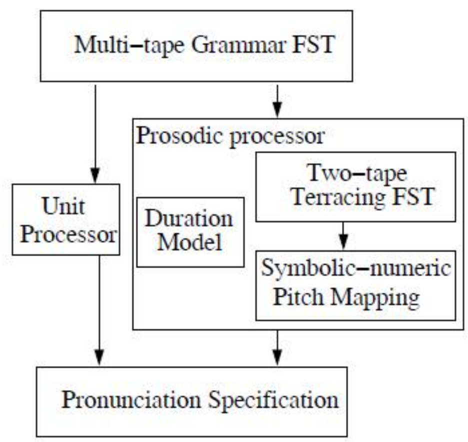


Figure 10. Architecture of a FST based tone language processor.

4.1.3 Applications of intonation

Intonation forms a central part of human speech communication, it does not only convey diverse linguistic information, but also information about the speaker, the speaker’s mood and attitude. The applications of intonation are enormous and cuts across diverse fields. One attraction of intonational features in many technical applications is that they are quite resistant to transmission distortions and noise. This property makes a synthesizer more intelligible and will be pursued in the course of this research.

Various applications of intonation are discussed in [16]. We summarize these in three categories:

Technological applications: Speech synthesis, speech recognition, speaker verification and language identification.

Medical applications: Rehabilitation of voice disorders and in psychiatry.

Educational applications: Prosodic training of ones voice, foreign language training and providing feedback for the medically challenged (the deaf).

4.2 The HMM-Based Speech Synthesis System (HTS)

HTS is a statistical parametric process that generates the average of some sets of similarly sounding speech segments. This technique contrasts with the unit-selection synthesis that has become common place within the last decade, which retains natural unmodified speech units.

Works on unit-selection synthesis has investigated the optimal size units to be selected for synthesis. The longer the unit, the larger the database must generally be to cover the required domain. Experiments have been demonstrated with different unit sizes [17]. The experiments prove that smaller units can be better as they offer more potential joining points. However, continuity can also be affected when more joining points are available, and demands for a more robust algorithm due to the many parameters to choose from, by varying the size of units, varying the size of the databases and limiting the synthesis domain. Black [18] has highlighted the different directions in constructing the best unit-selection synthesizer for the targeted application. The coverage of "more data" may seem an easy direction to adopt, but with speech databases growing to tens of hours of data, time-dependent voice-quality variations have posed great difficulty [22-24]. In addition, very large databases require substantial computing resources that limit unit-selection techniques on embedded devices or where multiple voices and multiple languages are required.

These drawbacks have informed the use of HTS in furthering our synthesis project. HTS requires minimal hours of speech recordings (training data). To further the Ibibio synthesis development, we intend to use 2-3 hrs of recordings. We are in the process of corpus enrichment for an unrestricted domain synthesis system. We have resorted to using HTS because of the following reasons:

(i) integration of tones will be less complicated since tone patterns can be parametrically handled. This is a far more better approach than other synthesis systems

(ii) excellent coverage of acoustic space would be achieved

(iii) the synthesizer will be more robust and simple because of its minimal database requirement

(iv) replicating the synthesizer for other tone languages and integrating same into mobile phones and other industrial applications or production of affordable devices similar to the simple computer (Simputer) used in India, for massive information dissemination will pose less difficulty since HTS supports multilingualism.

In a typical statistical parametric speech synthesis system, parametric representations of speech including spectral and excitation parameters are extracted from the speech database and modeled by applying a set of generative models (HMMs). Usually, the maximum likelihood criterion is used to estimate the model parameters, thus:

$$\hat{\lambda} = \arg\max_{\lambda} \{p\,(O \mid W, \lambda)\} \qquad (1)$$

where λ is a set of model parameters, O is a set of training data (selected corpus) and W is a set of word sequences corresponding to O. The speech parameters, o are then generated for a given word sequence to be synthesized, w, from the set of estimated models, $\hat{\lambda}$, in order to produce the optimal output probabilities from $\hat{\lambda}$, i.e.:

$$\hat{o} = \arg\max_{o} \{p(o \mid w, \hat{\lambda})\} \qquad (2)$$

Finally, a speech waveform reconstruction is obtained from the parametric representations of speech. Although any generative model may be suitable, but HMMs are most widely used. Statistical parametric speech synthesis with HMMs hereinafter referred to as HMMs-based Speech Synthesis have been extensively dealt with in [25-28].

Figure 11, is a HMM-based Speech Synthesis workflow diagram. It has two sections: the training and synthesis sections. The training section performs the maximum likelihood estimation of equation (1) using the EM algorithm [29]. This process is closely related to that of Speech recognition. The major difference being that both spectrum (for instance, mel-cepstral coefficients [30] and their dynamic properties) and excitation (e.g., log F_0 and its dynamic properties) parameters are extracted from a database of natural speech and modeled by a set of multi-stream context dependent HMMs. One other difference is that the linguistic, phonetic and prosodic contexts are taken into account. For instance, the contexts used in the HTS English recipes provided in [27], has the following context levels: phoneme, syllable, word, phrase and utterance.

The synthesis section maximizes equation (2). It can be viewed as the inverse operation of speech recognition. First, a given word sequence is converted into a context-dependent label sequence, and then the utterance HMM is constructed by concatenating the context-dependent HMMs according to the label sequence. Second, the speech parameter generation algorithm generates production sequences of spectral and excitation parameter for the utterance HMM. Variants of the speech parameter algorithms exist in [31-32]. A speech waveform is finally synthesized from the generated spectral and excitation parameters via excitation generation and speech synthesis filtering [33]. The speech parameter generation algorithm is described as follows:

Assume for simplicity of notation that each state output distribution of a single stream, single multiple variant Gaussian distribution is given as

$$b_j\left(o_t\right) = \psi\left(o_t\,;\, \mu_j\,,\, \textstyle\sum_j\right) \qquad (3)$$

where o_t is the state-output vector at frame $t, b_j(\bullet), \mu_j$ *and* $\sum_j$ correspond to the j-th state-output distribution and its mean vector and a co-variance matrix. Using the HMM-based speech synthesis framework, equation (2) can be approximated as [23]

$$\hat{o} = \arg\max_o \{p \,|\, (|\, w, \hat{\lambda})\} \qquad (4)$$

$$= \arg\max_o \left\{ \sum_q p\,(o, q \,|\, w, \hat{\lambda}) \right\} \qquad (5)$$

$$\approx \arg\max_o \max_q \left\{ p\,(o, q \,|\, w, \hat{\lambda}) \right\} \qquad (6)$$

$$= \arg\max_o \max_q \left\{ P\,(q \,|\, w, \hat{\lambda}) \bullet p\,(o \,|\, q, \hat{\lambda}) \right\} \qquad (7)$$

$$\approx \arg\max_o \left\{ p\,(o \,|\, \hat{q}, \hat{\lambda}) \right\} \qquad (8)$$

$$= \arg\max_o \left\{ \psi\,(o\,;\, \mu_{\hat{q}}\,,\, \textstyle\sum_{\hat{q}}) \right\} \qquad (9)$$

where $o = \left[o_1^T, \cdots, o_T^T\right]^T$ is the expected state-output vector sequence,

$q = \{q_1, \cdots, q_T\}$ is a sequence of states

$\mu_q = \left[\mu_{q_1}^T, \cdots, \mu_{q_T}^T\right]^T$ is the mean vector for q

$\sum_q = diag\left[\sum_{q_1}, \cdots, \sum_{q_T}\right]$ denotes the covariance matrix

T denotes the total number of frames in o.

Here, the state sequence $\hat{q}$ is determined in order to maximize its state-duration probability as:

$$\hat{q} = \arg\max_q \left\{ P\left(q \,|\, w, \hat{\lambda}\right) \right\} \qquad (10)$$

If the state-output vector, o_t, consists of M-dimensional static feature, c_t, and its first-order dynamic (delta) feature, Δc_t, then

$$o_t = \left[C_t^T\;, \Delta c_t^T\right]^T \qquad (11)$$

and the dynamic feature is calculated as [25]

$$\Delta c_t = c_t - c_{t-1} \qquad (12)$$

The relationship between o_t *and* c_t can be arranged in a matrix form thus

$$\underset{o}{\begin{bmatrix} \vdots \\ c_{t-1} \\ \Delta c_{t-1} \\ c_t \\ \Delta c_t \\ c_{t+1} \\ \Delta c_{t+1} \\ \vdots \end{bmatrix}} = \underset{W}{\begin{bmatrix} \cdots & \vdots & \vdots & \vdots & \vdots & \cdots \\ \cdots & 0 & I & 0 & 0 & \cdots \\ \cdots & I & I & 0 & 0 & \cdots \\ \cdots & 0 & 0 & I & 0 & \cdots \\ \cdots & 0 & -I & I & 0 & \cdots \\ \cdots & 0 & 0 & 0 & I & \cdots \\ \cdots & 0 & 0 & -I & I & \cdots \\ \cdots & \vdots & \vdots & \vdots & \vdots & \cdots \end{bmatrix}} \underset{c}{\begin{bmatrix} \vdots \\ c_{t-2} \\ c_{t-1} \\ c_t \\ c_{t+1} \\ \vdots \end{bmatrix}} \qquad (13)$$

Where $c = \left[c_1^T, \cdots, c_t^T\right]^T$ is a static feature vector sequence and W is a matrix which appends features dynamically to c. Here I and 0 corresponds to the identity and zero matrices. We observe that the state-output vectors are linear transformations of the static features, as such, maximizing $\psi\left(o; \mu_{\hat{q}}, \sum_{\hat{q}}\right)$ with respect to, o, is equivalent to that with respect to, c, i.e.

$$\hat{c} = \arg\max_c \left\{ \psi\,(Wc;\, \mu_{\hat{q}}\,,\, \textstyle\sum_{\hat{q}}) \right\} \qquad (14)$$

Equating $\partial \log \psi\,(Wc;\, \mu_{\hat{q}}, \sum_{\hat{q}}) / \partial c$ *to* 0, we obtain a set of linear equations for the determination of $\hat{c}$ as

$$W^T \textstyle\sum_{\hat{q}}^{-1} W\hat{c} = W^T \textstyle\sum_{\hat{q}}^{-1} \mu_{\hat{q}} \qquad (15)$$

$W^T \sum_{\hat{q}}^{-1} W$ is a positive-definite and band-symmetric structure and can be solved very efficiently. This converts the trajectory of $\hat{c}$ into a non-piece-wise stationary form due to its association with dynamic features which contribute to the likelihood and is therefore consistent with HMM parameters. As shown in figure 12, the trajectory of $\hat{c}$ becomes smooth rather than piece wise, due to the effect of the dynamic constraints.

4.2.1 Advantages of statistical parametric synthesis

Numerous advantages of statistical parametric synthesis over the unit-selection synthesis abound. Some of these advantages include:

- (i) flexibility in modifying and adapting voice characteristics, speaking styles, and emotions.
- (ii) coverage of acoustic space
- (iii) database size is not critical
- (iv) multi-lingual support
- (v) low memory requirements
- (vi) footprints reduction [34-36]
- (vii) stepwise quality improvements and more “robust” in handling speech quality degradation
- (viii) exploration of new HMM technologies
- (ix) provision of new frame-work for joint optimization of the front-end (text analysis) and back-end (waveform generation) modules of TTS systems
- (x) based on mathematically, well-defined statistical principles

4.2.2 Some drawbacks of HTS

HTS is not without drawbacks. Some of the drawbacks which are continually refined include:

- (i) buzzy sounding (synthesized) speech [26, 37]

(ii) the speech parameters are generated from acoustic models, which assumptions does not hold for real speech.
(iii) over-smoothing [38-40].

Figure 11. A workflow diagram for HMM synthesis (HTS)

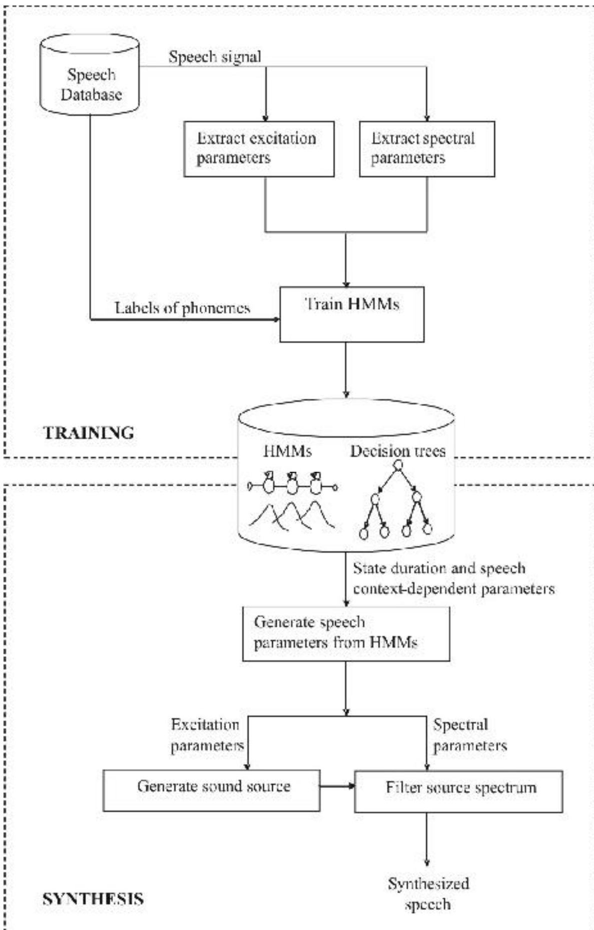


Figure 12. Illustrating the effect of dynamic constraints in HTS

4.3 Funding

The Ibibio synthesizer development emerged as a parallel project during a tripartite DAAD "Hochschulpartnershaften" funded project between Universität Bielefeld, Germany; the University of Uyo, Nigeria and Universite de Cocody, Abidjan, Côte d'Ivoire and was funded by the Local Language Speech Technology Initiative (LLSTI). LLSTI provided a one-year industry scholarship and computing equipment for the speech synthesis project and was anchored by Outside Echo, a non-governmental organization in the UK, interested in aiding the development of synthesizers for minority languages.

To ensure sustainability of this initiative, a Science and Technology Education Post Basic (STEP-B) Nigeria/World Bank project funding was sourced. The project proposal "Towards a generic Text-to-Speech applications for African tone languages", has been approved for funding. The STEP-B project is a World Bank/Federal Government of Nigeria funded project. The project aims at improving the standard and infrastructure of science and technology education/research in Nigeria.

We are optimistic that the integration of tone into the current synthesizer will be completed in the course of the project. Our vision of the STEP-B project is to evolve into a center of excellence – Centre for Speech Technology Research (CSTR), which will serve as a network hub for speech research in Nigeria, where specialized degree courses would be offered. This idea, we hope, will strengthen our international collaboration network and explore more collaborative ties with other institutions across the globe. An international workshop is planned in the first year of the project (2010), amongst series of conferences/workshops.

A summary of project objectives include

(i) to build on existing basic proof of concept prototypes
(ii) to create a substantive standardized text and speech corpus archive for education, research and development, following international standards (for format, storage, metadata and access) and best practices.
(iii) to provide a standardized framework for language development
(iv) to develop sets of re-useable text and speech resources (tools, machine-readable dictionaries and grammars)
(v) to develop a software toolkit for further applications in the field in the domains of public information services and tutorial systems
(vi) to provide an open-source, state-of-the-art strategy and adaptation procedures that is generic for other African languages, in consultation with colleagues
(vii) to exhibit an affordable information system with a broad domain coverage
(viii) to develop and train experts in the field of Human Language Technology (HLT) by embarking on rigorous and massive training of staff and students (by experienced professionals) through workshops, conferences and graduate programmes
(ix) to establish and maintain a collaborative research network with local and international partners.

The outlined objectives will be achieved within the project's two years timeline (2010-2012). Plans on fostering international collaboration and investing in massive training of manpower are expected. We propose practice on three synthesis systems namely: Festival, Bonn Open Speech Synthesis (BOSS) system and Hidden Markov Model TTS (HTS) system. The reason for this is to explore as much as possible, various state of the art technologies that will perfect manpower training and expertise.

4.5 By-products

This section presents the initial by-products (data and speech resources), obtained from the Ibibio speech synthesis development and dwells on a new speech resource. Outlined below are the tangential outcomes/deliverables:

(i) Ibibio corpora/lexicons
(ii) annotated Ibibio speech files
(iii) Ibibio e-dictionary [41]
(iv) Ibibio recordings and video files
(v) Ibibio e-concordance
(vi) Nkari, Annang, Ilue, Ekit and Oro language resources
(vii) Medefaidrin language documents scan

In the next section, we discuss ways of evolving the last by-product (the Medefaidrin language documents scan), into a technology (e-preservation) system. We intend to incorporate the Ibibio e-dictionary into this archive to form a multilingual Web dictionary.

5. The talking Medefaidrin Web dictionary

This section proceeds by presenting in Figure 13, the architecture of a talking Medefaidrin Web dictionary. Medefaidrin is a language spoken by a Christian religious group known as Oberi Okaime Church in Ibiono Ibom and Itu Local Government Areas of Akwa Ibom State of Nigeria. The language, according to the adherents, was revealed to the founder of the church in 1927 by the Holy Spirit, which is known as 'Seminant' in this language. One interesting fact about the language is that the users of the Medefaidrin language have a different first language, which is the Ibibio language. The Medefaidrin language also has its unique script for writing. Figures 14 and 15 shows a scan of the Medefaidrin corpus, captured from dilapidated exercise books and a digitally revived copy, filtered for colouration, brightness and contrast to preserve the script electronically. Medefaidrin is unfortunately endangered and thus, the need for appropriate documentation of the language. We implement a dictionary structure similar to [42] with a minimal target language.

The architecture implements a modular XML design and is dynamic in nature. We design the database (back-end) with Mysql and script the front end with PHP, adding other flavours and formats with Dreamweaver and Flash. The Web design and data upload is in progress. Figure 16 is the front-end implementation of the design.

The search engine works with a dynamic algorithm defined in the administration section and is context sensitive. This is to ensure flexibility of the search engine and improve the search performance. A sound link is incorporated to actually simulate the sound. Meanings in English and Ibibio are also provided to enhance easy understanding of the language.

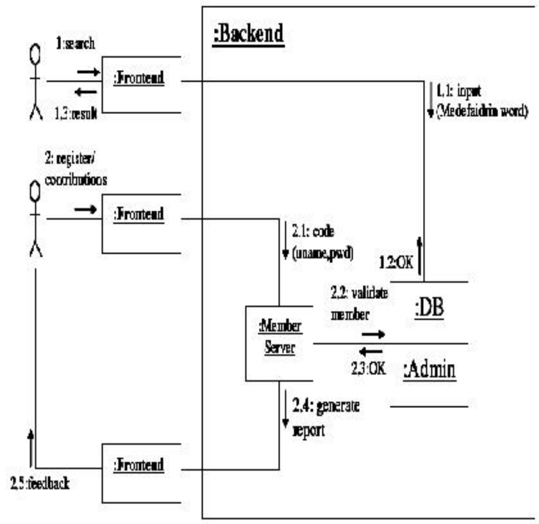


Figure 13.The Medefaidrin Web dictionary architecture

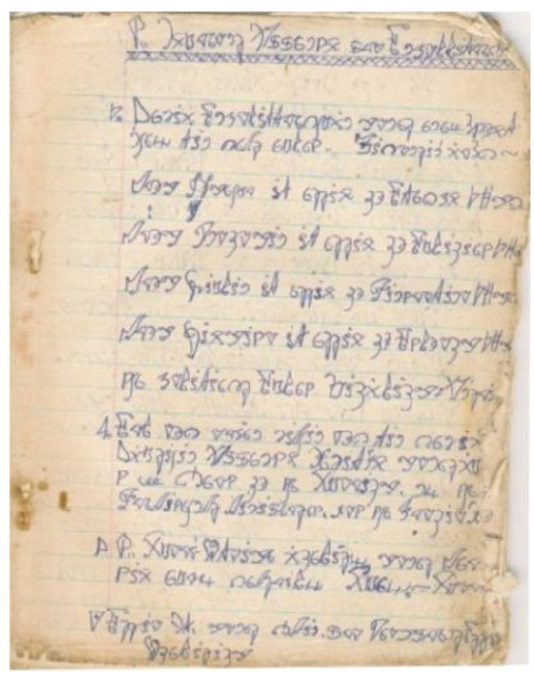
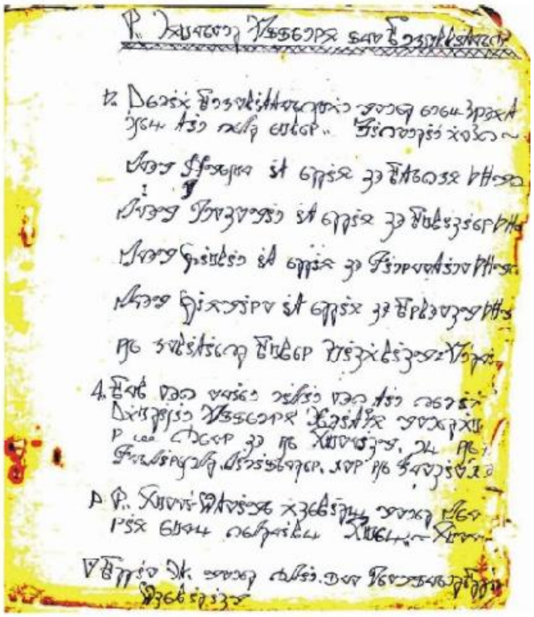


Figure 14. A sample scan copy of the Medefaidrin corpus

The Web dictionary will serve as a useful tool for learning and research on spoken languages. The project will also benefit the local communities from where these data were captured. The Website shall accept contributions from interested persons and will be hosted after data upload. The site, we hope, will help revitalize the language, which had been neglected. More information on the Medefaidrin language can be found in [43-47] and the multi-lingual construction methodology is available in [48]



15. Digitally revived electronic version to preserve the corpus

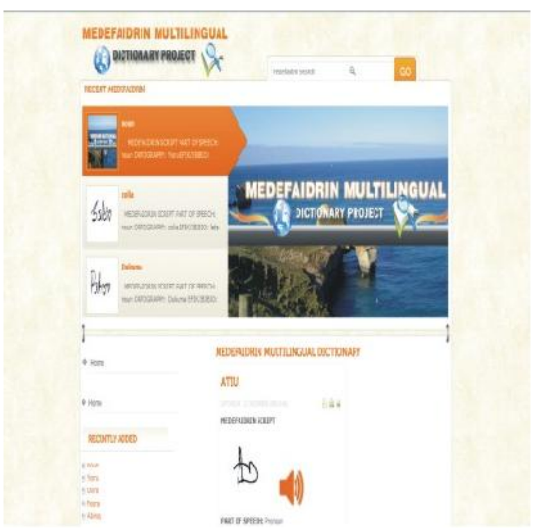


Figure 16. A design implementation of the Talking Medefaidrin Web dictionary

6. A Unified Model

Unifying linguistic resources has in recent times received considerable attention. In [49], the need for a unified meta-modeling of linguistics resources has been stressed. We propose in Figure 15, a client-server model that allows for the interoperability of linguistics resources. Here, the server side has a database of heterogeneous resources and interoperability mapping tools. The clients on the other hand interface the server through applications mapping. The proposed model is generic and will encourage collaboration, content exchange and management of resources. It also enforces adherence to standards and agreeable formats. The clients are plug-and-plays to the server and details behind the generic interface are made transparent to the clients.

The following are requirements for a successful implementation of the proposed scheme:

(i) An open forum for features review of software formalisms
(ii) A Generic XML-based exchange format
(iii) Converters that rewrite formalisms into system specific expressions

It is expected that these linguistics resources require standardization, and when standardizing the linguistic resources, the following key issues should be dealt with

(i) Managing the trade-off between interoperability and viability of linguistic representation.
(ii) Documenting and maintaining document formats
(iii) Unifying queries, presentation and management of the resources.

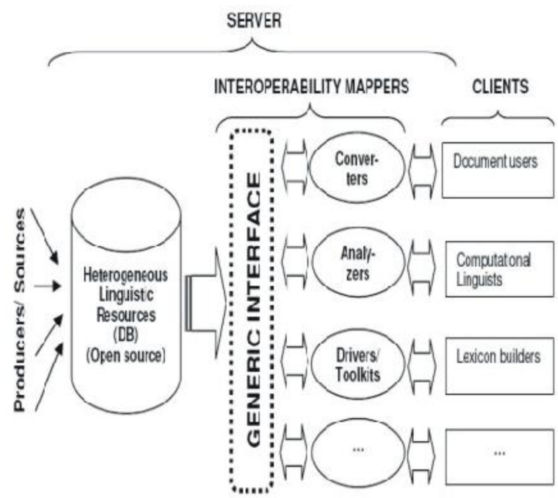


Figure 17. A unified model for linguistics resources

7. Conclusion and Outlook

In this paper, particular attention has been to creating open source toolkit components that will impact on the society. The major groups that would benefit from this effort include the language communities and the telecommunication industries. We anticipate that on successful implementation of tone into the current synthesizer, developing speech synthesis applications for other African tone languages would have succeeded. We would have also addressed some of the many open “AI complete” issues in speech technology.

Acknowledgements

We acknowledge the STEP-B Nigeria/World Bank for providing the grant (grant number: FME/STEP B/79/3/14) to further the speech synthesis research project.

We are also grateful to Reverend Richmel Ekit and the members of the Oberi Okaime Christian Church, Edem Urua, Ibiono Ibom Local Government Area of Akwa Ibom State, Nigeria for providing the Medefaidrin data.

References

[1] D. Gibbon. Workable Efficient Language Documentation: a Report and a Vision. ELSENews 11.3. Autumn, pp. 3-5, 2002.

[2] K. Boves, R. Calson, E. Hinrichs, D. House, S. Krauwer, L. Lemnitzer, M. Vainio and P. Wittenburg. Resources for Speech Research: Present and Future Infrastructure Needs. In *Proceedings of 2009 Interspeech Conference*, Brighton, pp.1803 – 1806, 2009.

[3] N. Campbel, W. Hess and B. Moebius and J-V. Santex. The ISCA Special Interest group on Speech Synthesis. http://www.ims.uni-stuttgart.de/~moebius/papers/synsig-paper.pdf

[4] D. Mostefa, M. N. Garcia, O. Hamon and N. Moreau. Deliverable 16: *Evaluation Reports. ELDA Technical Report*, 2006.

[5] C. T. Bennett and A. W. Black. The Blizzard Challenge. In *Proceedings of Blizzard Challenge 2006 Workshop*, Pittsburg, PA, 2006.

[6] A. Bounafonte, J. Adeu, J. Esquerra, S. Gallego, A. Moreno and J. Perez. Corpus and Voices for Catalan Speech Synthesis. In *Proceedings of Language Resources and Evaluation Conference (LREC),* pp. 3325-3329, 2008.

[7] M. E. Ekpenyong, E-A. Urua and D. Gibbon. Towards an Unrestricted Domain TTS system for African Tone Languages. *International Journal of Speech Technology*. 11: pp. 87-96, 2008.

[8] M. Ekpenyong, M. Udoinyang and E. Urua. A Robust Language Processor for African Tone Language Systems. *Georgian Electronic Scientific Journal : Computer Sciences and Telecommunications*. 6(23): pp. 3-12, 2009.

[9] D. Gibbon, E. Urua and M. Ekpenyong. Data Creation for Ibibio Speech Synthesis. LLSTI Project Report. Third Parners Workshop, Lisbon, 2004.

[10] Y. Xu. Contextual tonal variations in Mandarin, *Journal of Phonetics*, 25, pp. 61-83, 1997.

[11] Y. Xu. Effects of tone and focus on the formation and alignment of F_o Contours. *Journal of Phonetics* 27, 55 – 105, 1999.

[12] D. Gibbon, E-A. Urua; M. Ekpenyong. Problems and Solutions in African Tone Language Text–To–Speech. *MULTILING-2006*, ISCA Publications Archive, Paper 14. http://www.isca-speech.org/archive/ml06/ml06/ml06_014.html, 2006.

[13] D. Gibbon & E. A. Urua. Morphotonology for TTS in Niger-Congo Languages. In *Proceedings of Speech Prosody Conference*, Dresden, Germany, http:///www.isca-speech.org/archive, 2006.

[14] Steedman, Mark. Syntax, Intonation and focus. In Ewan Khein and Frank Veltman (eds.) *Natural Language and Speech* Mahwah, N. J. Lawrence Erlbaum Associates, pp.331-342, 1991.

[15] J. B. Pierrehumbert. The Phonology and Phonetics of English Intonation. PhD Dissertation, MIT, MA, 1980.

[16] A. Botinis, B. Granstrom and B. Mobius. Developments and paradigms in intonation research. *Speech Communication*. 33 (2001), pp. 263 – 296, 2001.

[17] S. Kishore and A. Black. Unit size in Unit Selection Speech Synthesis. In *Proceedings of Interspeech*, pp. 1317-1320, 2003.

[18] A. Black. Perfect synthesis for all of the people all the time. In *Proceedings of IEEE Speech synthesis* workshop, 2002.

[19] Bruce, G., Granström, B., Gustafson, K., House, D. Prosodic phrasing in Swedish. In: *Proceedings of theESCA Workshop on Prosody*, Lund, Sweden, pp. 180-183, 1993.

[20] E. Urua and S. Bird. The phonetic description of Ibibio tones. In proceedings of the Edinburgh Linguistics department conference, pp. 210-216, 1994.

[21] D. Gibbon. Tone and Timing: Two problems and two methods for prosodic typology. In International symposium on tonal aspects of languages with emphasis on tone languages, Beijing, China, 2004. http://www.isca-speech.org/archive/tall2004/tal_065.pdf

[22] Y. Tylianou. Assessment and correction of voice quality variabilities in large speech databases for concatenative speech synthesis. In *Proceedings of ICASSP*, pp. 377-380, 1999.

[23] H. Kawai and M. Tsuzaki. A study on time-dependent voice quality variations in a large-scale single speaker speech corpus used for speech synthesis. In *Proceedings of IEEE speech synthesis* workshop, 2002.

[24] Y. Shi, E. Chang, H. Peng and M. Chu. Power Spectral density based channel equalization of large speech database for concatenative TTS system. In *Proceedings of ICSLP,* pp. 2369-2372, 2002.

[25] T. Masuko, K. Yokuda, T. Kobayashi, and S. Imai. Speech synthesis from HMMs using dynamic features. In *Proceedings of ICASSP*, pp. 389 – 392, 1996

[26] T. Toshimura, K. Tokuda, Masuko, T. Kobayaashi, Y. Kitamura. Simultaneous modeling of spectrum, pitch and duration in HMM-based speech synthesis. In *Proceedings of Eurospeech*. pp. 2347 – 2350, 1999.

[27] K. Tokuda, H. Zen, J. Yamagishi, A. Black, T. Masuko, S. Sako, T. Toda, T. Nose, K. Oura,.

The HMM-based speech synthesis system (HTS). http://hts.sp.nitech.ac.jp/ . 2008.

[28] W. Tachiwa, S. Furui. A study of speech synthesis using HMMs. *In proceedings of spring meeting* of ASJ, pp. 239-240, 1999.

[29] A. Dempster, N. Laird, D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistics Society* 39, pp. 1-38, 1977.

[30] T. Fukada, K. Tokuda, T. Kobayashi, S. Imai. An adaptive algorithm for mel-cepstral analysis of speech. *In Proceedings of ICASSP*. pp. 137-140, 1992.

[31] K. Tokuda, T. Yoshimura, T. Masuko, T. Kobayashi, T. Kitamura. Speech parameter generation algorithm for HMM-based speech synthesis. *In proceedings of ICASSP*, pp. 1315-1318, 2000.

[32] S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X-Y. Liu, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, P. Woodland. The Hidden Markov Model Toolkit (HTK) version 3.4. http://htk.eng.cam.ac.uk/, 2006.

[33] S. Imai, K. Sumita, C. Furuichi. Mel log spectrum approximation (MLSA) filter for speech synthesis. Electronics and communications in Japan. 66(2), pp. 10-18, 1983.

[34] H. Zen, T. Toda, M. Nakamura, T. Tokuda. Details of the Nitech HMM-based speech synthesis system for the Blizzard Challenge 2005. *IEICE Transactions on Information System*. E90-D (1), pp. 325 – 333, 2007.

[35] Y. Morioka, S. Kataoka, H. Zen, Y. Nankaku, K. Tokuda, T. Kitamura. Miniaturization of HMM-based speech synthesis. In *Proceedings of Autumn Meeting* of ASJ. pp. 325 – 326, 2004.

[36] K. Oura, H. Zen, Y. Nankaku, A. Lee, K. Tokuda. Tying variance for HMM-based speech synthesis. In *Proceedings of Autumn Meeting* of ASJ. pp. 421-422, (in Japanese), 2008.

[37] T. Toshimura, K. Tokuda, Masuko, T. Kobayaashi, T. Kitamura. Mixed excitation for HMM-based speech synthesis. In *Proceedings of Eurospeech*. pp. 2263 – 2266, 2001.

[38] Z.-H. Ling, R.-H. Wang. Minimum unit selection error training for HMM-based unit selection speech synthesis system. In *Proceedings of ICASSP*. pp. 3949 – 3952, 2008.

[39] K. Oura, H. Zen, Y. Nankaku, A. Lee, K. Tokuda. Postfiltering for HMM_based speech synthesis using mel-LSPs. In *Proceedings of Autumn Meeting* of ASJ. pp. 367-368, 2007.

[40] J. Yu, M. Zhang, J. Tao, X. Wang. A novel HMM-based TTS system using both continuous HMMs and discrete HMMs. In *Proceedings of ICASSP*. pp. 709 – 712, 2007.

[41] E-A. Urua, M. Ekpenyong and D. Gibbon. ABUILD Language Documentation Curriculum Materials (Lexicography). Uyo Ibibio Dictionary. Preprinted Draft, Version 01. http://coral.lili.uni-bielefeld.de/Courses/Summer04/Lexicography/IbibioDictionary/ibibio_dictionary01.pdf

[42] M. Lafourcade. Multilingual Dictionary Construction and Services. Case Study with Fe* Projects. In *Proceedings of PACLING '97*, Meisei University, Ohme, Tokyo, 1997.

[43] R. F. G. Adams. Oberi Okaime: A New African Language and Script. *Africa: Journal of the International African Institute*. 17(1): 24-34. Edinburgh University Press, 1947.

[44] E-A. Urua. Medefaidrin, the Spirit Language of the 1920s in Ibibio Land. In O-M. Ndimele, I. I. L.Udoh and O. Anyanwu (eds.) *Critical Issues in the Study of Linguistics, Languages and Literatures in Nigeria : A Festshrift for Conrad Max Benedict Brann*. Festschrift Series 7, Port Harcourt, Nigeria : LAN in Collaboration with M&S Grand Orbit Communications Ltd and Emhai Press. 2, 2008.

[45] M. B. Abasiattai. The Oberi Okaime Christian Mission : Towards a History of an ibibio Independent Church in *Africa : Journal of the International African Institute*. 59(4), pp. 496-516, 1989.

[46] M. B. Abasiattai. The Ibibio Oberi Okaime : The Tenth Wonder of the World. A Lecture Presented at the First Faculty of Arts Distinguished Lecture Series, University of Uyo, Uyo. 2008.

[47] H. Kathleen. Oberi Okaime Script, Texts and Counting Systems in *Bulletin de Institut Francais d'Afrique Noire.Serie B*, Sciences Humanities, Dakar. 23(1-2), pp. 291-308.

[48] E-A. Urua and M. Ekpenyong. A Robust XML-Platform for Medefaidrin Multilingual Lexical Database Development. In. *Language Development Beyond Borders: A festschrift in honour of Professor Dr. Dafydd Gibbon*, Urua, E-A., M. Ekpenyong & F. Ahoua (eds.), Uyo, Nigeria: Emsel Press, 2009.

[49] M. Ekpenyong. A Unified Meta Model for Interoperability of Linguistic Resources. In *Language Development Beyond Borders: A festschrift in honour of Professor Dr. Dafydd Gibbon*, Urua, E-A., M. Ekpenyong & F. Ahoua (eds.), Uyo, Nigeria: Emsel Press, 2009.

Moses E. Ekpenyong obtained his B.Sc. in Computer Science/Statistics in 1995, M.Sc. in Computer Science in 2004 and is currently on his Ph.D. research programme in Speech Technology. He is a lecturer in the Department of Computer Science in the University of Uyo and belongs to the following professional bodies: Nigerian Association of Mathematical Physics (NAMP), Nigeria Computer Society (NCS), International Speech and Communications Association (ISCA), Nigeria Mathematical Society (NMS), Nigerian Statistical Association (NSA), West-African Linguistics Society (WALS) and African Human Language Technology (HLT) Society. He is an International Scholar and has been involved in collaborative/interdisciplinary research projects on Language Documentation, Speech Technology and Wireless communications technology both nationally and internationally. He has a plethora of publications in his areas of specialization - speech and wireless communication technologies.

Eno-Abasi E. Urua, is a Professor of Linguistics in the Department of Linguistics and Nigerian languages, University of Uyo. She obtained her B.A. Ed (English) from the University of Calabar in 1980, M.A. and Ph.D. degrees in Linguistics, from the University of Ibadan in 1987 and 1990 respectively. Her research interests are in Phonetics and Phonology, with a focus on the phonology of Lower Cross languages. Her other research interests are in the area of language documentation and description (Documentary Linguistics) and speech annotation. Over the years, Professor Urua has done extensive and insightful research on the tonal phonology of Lower Cross languages. She has over 40 publications to her credit in both local and international journals. Among these, Ibibio Phonetics and Phonology (2000) remains a reference point to her colleagues and students. She is a member of many professional associations, including the Linguistic Association of Nigeria (LAN), the West African Linguistic Society (WALS), World Congress of African Linguistics (WOCAL), the African Academy of Languages (ACALAN), the Association of Nigerian Language Teachers (ANLAT), and the National Association of Women Academics (NAWACS). She is the Editor of several professional journals, including Journal of the Linguistic Association of Nigeria (JOLAN), USEM, journal of Language, Linguistics and Culture, Journal of Women Academics (JOWACS), and is a reviewer to local and international journals. She is currently the Director, International Programmes in the University of Uyo.

Victor J. Ekong graduated with a B. Sc. in Mathematics, from the University of Ibadan in 1968. He holds an M.S. (1970) and Ph.D. (1972) in Computer and Information Science from the Ohio State University. He taught at Florida International University, Miami, USA, 1972 - 1974, the Obafemi Awolowo University, Ile-Ife, 1974 -1993, and the University of Uyo, currently. From 1997 to 2002, Dr Ekong was the Acting Director of the Uganda Educational Computer Center, where he assisted in establishing both e-mail and Internet for the Institute of Teacher Education, Kyambogo (ITEK), now the University of Kyambogo. In the Obafemi Awolowo University, he supervised a number of M.Sc. students in Numerical Analysis and Compiler Design. He is currently the Director of Computer Services in the University of Uyo.

Okure U. Obot obtained his B.Sc. (Hons) in Mathematics/Computer Science (1994) from the University of Port Harcourt, Nigeria, M. Tech. and Ph.D. in Computer Science from the Federal University of Technology, Akure in 1999 and 2007, respectively. He lectures in the Department of Computer Science, University of Uyo, Nigeria. His research interest is in Soft computing, Knowledge Engineering and Computational Linguistics. Dr. Obot is a member of the Nigeria Computer Society and the African HLT Society. He has published in both National and International Journals in his area of specialization.

Imelda I. Udoh, received her Bachelor's, Master's and Ph.D degrees from the University of Calabar in 1983, 1986, and 1998 (Inter-language Phonology). She is currently an Associate Professor and a member of staff of the Department of Linguistics and Nigerian Languages, University of Uyo since 1986, and is currently the Acting Head of Department. Dr. Udoh has been a Research Visitor and Scholar to the School of Oriental and African Studies (SOAS), University College, London, University of California, Berkeley, and University of Calabar. She has done extensive research in Descriptive and Documentary Linguistics on the languages of the Upper and Lower Cross River spoken in Akwa Ibom and Cross River States of Nigeria, which has yielded over forty publications including articles in both national and international journals. She has a passion for minority languages in Nigeria. She is a member of several professional associations like West African Linguistics Society (WALS), the Reading Association of Nigeria (RAN), the Nigerian Language Teachers' Association (ANLAT), the Linguistic Association of Nigeria (LAN) , of which she was Secretary from 2004-2008.