Language Endangerment Threatens Phonetic Diversity

Ettien Koffi

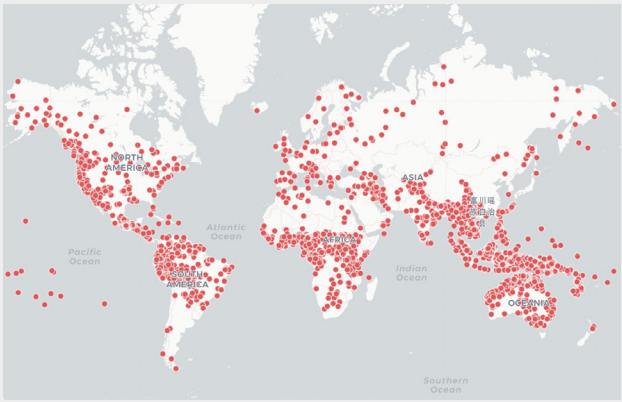
Language Endangerment at a Critical Point

A pandemic of linguicide has been decimating minority languages all over the world at an alarming rate. Its virulence has reached such a velocity that UNESCO (2003) projected that 90% of the world's languages will be dead by the year 2100. Statistically speaking, according to McWorther (2003), "a language dies roughly every two weeks." Currently, 2,923 of the world's 7,111 languages are in serious stages of endangerment (Eberhard et al., 2019). The situation is so dire that the United Nations (UN) called attention to it by declaring 2019 the International Year of Endangered Language (IYIL 19). Now, the UN is taking an unprecedented step by declaring the decade from 2022 to 2032 the International Decade of Indigenous Languages (IDIL 22-32; see bit.ly/39uudCv). This article highlights what language endangerment entails for phonetic diversity and what experts have done and are doing to stem the tide. It also highlights speech synthesis as a new model of language documentation that can help preserve phonetic diversity even if a critically endangered language breathes its last. Speech synthesis is a technique used so that smart devices can speak and understand language. Siri and Alexa are among the products of speech synthesis. More details are given in **A Simpler Speech Synthesis Model**.

Root Causes of Endangerment

There are several comorbidity factors that conspire to cause a language to die. Among them are genocides, epidemics, natural disasters, migration, urbanization, political/economic imperialism, population size, and the interruption of intergenerational language acquisition. For the sake of brevity, I focus only on the last three causes of endangerment. All things being equal, the smaller the number of people who speak a given language, the slimmer its chance of survival in the linguistic jungle where bigger languages eat smaller ones. With regard to size, it is estimated that languages that are spoken by 5,000 people or less are in grave danger of extinction (Krauss, 1992). Nearly 40% of the world's languages fall into this category. In fact, Tucker and Wright (2020) estimate that 25% of them have fewer than 1,000 speakers. Indigenous languages also become endangered when their speakers find themselves under the political and economic rule of a dominant language (Mufwene and Vigoroux, 2008). Under such circumstances, minority languages lose their marketability value, which means that any language that does not afford its speakers better chances for socioeconomic mobility is sooner or later abandoned for one that does. When this happens, parents no longer see any value in passing that language on to the next generation (Baugh, 2009; Koffi, 2012). Batibo (2009) and Sands (2017) among others contend that these aforementioned comorbidity factors are the ones that are pushing 2,923 indigenous language to the brink of extinction. **Figure 1** gives a visual display of the breadth and depth of language endangerment worldwide.

Figure 1. *Map of endangered languages. Each* ***red dot*** *represents an endangered language. An interactive map with the names of each language is available at bit.ly/3ta9efO. Map used with permission, copyright © 2020* Ethnologue.



https://doi.org/10.1121/AT.2021.17.2.23

Table 1. *Safe languages*

EGIDS	Status	Descriptions of Language
EGIDS 0	International	Used internationally, i.e., English, French, Spanish.
EGIDS 1	National	Used in education, work, mass media, and government at a national level, i.e., Hindi, Swahili, Thai.
EGIDS 2	Provincial	Used in education, work, mass media, and government within major administrative subdivisions of a nation, i.e., Yoruba, Igbo, Hausa, (Nigeria, West Africa).
EGIDS 3	Wider communication	Used at work and in mass media without official status to transcend differences across a region, i.e., Akan (Ghana), Lingala (Democratic Republic of the Congo), Luganda (Uganda), Ewe (Togo).
EGIDS 4	Educational	In vigorous use, with standardization and literature being sustained through a widespread system of institutionally supported education, i.e. Bambara (Mali), Wolof (Senegal), Olusamia (Kenya).
EGIDS 5	Developing	In vigorous use, with literature in a standardized form being used by some though this is not widespread or sustainable, i.e., Aja, Gun, Lokpa (Benin).
EGIDS 6a	Vigorous	Used for face-to-face communication by all generations and the situation is sustainable, i.e., Kabye, Moba (Togo), Dewoin (Liberia).

EGIDS, Expanded Graded Intergenerational Disruption Scale.

Table 2. *Endangered languages*

EGIDS	Status	Descriptions of Language
EGIDS 6b	Threatened	Used for face-to-face communication within all generations, but it is losing users, i.e., Aizi, Krumen (Côte d'Ivoire), !Xóõ (Botswana).
EGIDS 7	Shifting	Child-bearing generation can use the language among themselves, but it is not being transmitted to children, i.e., Frisian, Walser (Germany), Svan (Georgia).
EGIDS 8a	Moribund	Only remaining active users are members of the grandparent generation and older, i.e., Jeri Kuo, Beti, Mbre (Côte d'Ivoire).
EGIDS 8b	Nearly extinct	Only remaining users of the language are members of the grandparent generation or older who have little opportunity to use the language, i.e., Breton (France), Odut (Nigeria), Saami (Sweden).
EGIDS 9	Dormant	Serves as a reminder of heritage identity for an ethnic community, but no one has more than symbolic proficiency, i.e., Saami (Norway), Cornish, Polari (United Kingdom).
EGIDS 10	Extinct	No longer used and no one retains a sense of ethnic identify, i.e., Maidu, Mahican, Iowa-Oto (United States).

Assessment of Gravity

Experts rely on specific metrics to assess the levels of linguistic vitality or endangerment. UNESCO (2010) uses seven indices to classify languages: (1) safe; (2) stable yet threatened; (3) vulnerable; (4) definitely endangered; (5) severely endangered; (6) critically endangered; and (7) extinct. Eberhard et al. (2019) use a 13-point "Expanded Graded Intergenerational Disruption Scale" (EGIDS). This is the method of classification used in this article because it is more descriptive (**Table 1**).

Languages with EGIDS 0-3 have been referred to by some language advocates as "killer" languages because their spread has proven to be harmful to the languages in **Table 2**.

The IDIL 22-32 declaration concerns all the languages listed in **Figure 2**. However, languages with EGIDS 8a, 8b, and 9 should receive greater attention. Languages with EGIDS 9, 312 in all, should be prioritized because they are on the brink of extinction.

Hospice Versus Palliative Care Mindsets

What documentation model is best to preserve phonetic diversity before critically endangered languages die? To answer this question, we can analogize to health care professionals who manage terminally ill patients. Two types of mindsets guide the care they provide: the hospice care mindset (HCM) and the palliative care mindset (PCM). There is a subtle but important distinction between the two, even though, in both cases, the patient is critically ill and the specter of death is omnipresent. In HCM, the operative assumption is that patients have six months or less to live. Therefore, every effort is made to keep them medically comfortable before death. In PCM, the view is that although death may be around the corner and patients may be in a critical state for some time to come, there is still a glimmer of hope that they may pull through. Consequently, care providers do not give up easily on these patients. I contend that if language planners want to preserve phonetic diversity during IDIL 22-32, they should operate under the assumptions of PCM. Crystal's (2000) six "postulates" in **Table 3** are likened to care plans that can be used to save critically endangered languages from extinction.

Four of the six postulates lie outside the expertise or control of linguists. The two that a linguist can operate within are Postulates 5 and 6. In fact, it can be argued that linguists have used various approaches of Postulate 5 to document languages. Unfortunately, the strategies that are currently being used are good only at preserving vestiges of dying languages and they are incapable of preventing moribund languages from dying. If the same methods continue to be used during IDIL 22-32, languages with EGIDS 8a/8b and EGIDS 9 will have no fighting chance of survival. However, if current models are used in conjunction with speech synthesis, moribund languages will not only survive but some may even thrive again. However, before rolling out this new documentation approach, we need to take a detour to familiarize ourselves with past and current methods of language documentation because they will contribute to the technologization of endangered languages as stated in Postulate 6.

Figure 2. *Worldwide distribution of endangered languages in different regions having Expanded Graded Intergenerational Disruption Scale (EGIDS) 6b to 9.*

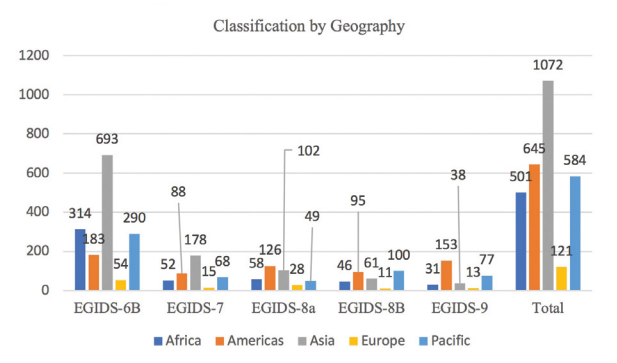


Table 3. *Postulates and their description*

Postulates	Description of Endangered Language
Postulate 1	Each will progress if speakers increase their prestige within the dominant community.
Postulate 2	Each will progress if speakers increase their wealth relative to the dominant community.
Postulate 3	Each will progress if speakers increase their legitimate power in the eyes of the dominant community.
Postulate 4	Each will progress if speakers have a strong presence in the education system.
Postulate 5	Each will progress if speakers can write their language down.
Postulate 6	Each will progress if speakers can make use of electronic technology.

Graphicization

The first line of defense against language death is graphicization. It consists of providing an unwritten language with a spelling system that represents its sound system adequately (Coulmas, 1989). Graphicization means a difference between language death and language extinction. If a language dies while it has not been reduced to writing, this dead language goes completely extinct, leaving absolutely no trace in the annals of languages. However, if a language dies after having been graphicized, it will never go extinct because it can be reconstructed on the basis on written records. Unfortunately, 1,409 languages of Africa's 2,902 languages are yet to be reduced to writing (Koffi, 2012).

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Symbols to the right in a cell are voiced, to the left are voiceless. Shaded areas denote articulations judged impossible.

Figure 3. *International Phonetic Alphabet (IPA) System for Consonants.* ***Top row****, where sounds are formed in the mouth.* ***Left column****, how the airstream is modified in the production of individual speech sounds. Available at bit.ly/3py8Oh1, license by Creative Commons.*

Phonetic Transcription

The International Phonetic Alphabet (International Phonetic Association, 1996) is used to transcribe speech sounds phonetically. It is a technique that was devised by linguists in 1886 to represent speech sounds as accurately as humanly possible. The IPA symbols used in transcription are displayed in **Figures 3** and **4**. The unique symbols in **Figures 3** and **4A** are arranged in such a way that they represent the areas in the mouth where a sound is produced, how air from the lungs or ambient air is modified, and what the tongue and other articulators do when speech sounds are made. Each symbol represents one and only one sound, and each sound is represented by one and only one symbol.

The IPA system is patterned after the table of elements used in chemistry and consists of some 86 unique symbols for consonants, 28 vowels, 33 diacritics (signs placed above, below, or next to symbols to show a slight modification in pronunciation), and 5 pitch levels (Pullum and Ladusaw, 1986). This impressive arsenal has been used and continues to be used to document endangered languages all over the world. Gambarage (2017) refers to it as "the lingua franca for linguists," adding that "the IPA transcription is not intended for native speakers; it is for the wider linguistic community who will use the data for analysis and comparative studies." Ladefoged (2003) notes that it was the primary documentation method during the first half of the twentieth century. Despite its popularity among linguists, critics have long charged that IPA transcription is far from ideal because it relies exclusively on the transcriber's own hearing acuity.

Figure 4. ***A:*** *Vowel chart, how vowels are produced.* ***Left column****, mouth aperture.* ***Top row****, movement of the tongue.* ***Front****, the tongue moves forward toward the front teeth;* ***central****, a neutral tongue position;* ***back****, the tongue retracts toward the back of the mouth.* ***B:*** *Diacritics chart, signs placed under, above, or next to symbols to indicate subtle modifications in pronunciation.*

A VOWELS

	Front	Central	Back
Close	i • y	ɨ • ʉ	ɯ • u
	ɪ ʏ		ʊ
Close-mid	e • ø	ɘ • ɵ	ɤ • o
		ə	
Open-mid	ɛ • œ	ɜ • ɞ	ʌ • ɔ
	æ	ɐ	
Open	a • ɶ		ɑ • ɒ

Where symbols appear in pairs, the one to the right represents a rounded vowel.

B DIACRITICS

̥	Voiceless	n̥ d̥	̤	Breathy voiced	b̤ a̤	̪	Dental	t̪ d̪
̬	Voiced	s̬ t̬	̰	Creaky voiced	b̰ a̰	̺	Apical	t̺ d̺
ʰ	Aspirated	tʰ dʰ	̼	Linguolabial	t̼ d̼	̻	Laminal	t̻ d̻
̹	More rounded	ɔ̹	ʷ	Labialized	tʷ dʷ	̃	Nasalized	ẽ
̜	Less rounded	ɔ̜	ʲ	Palatalized	tʲ dʲ	ⁿ	Nasal release	dⁿ
̟	Advanced	u̟	ˠ	Velarized	tˠ dˠ	ˡ	Lateral release	dˡ
̠	Retracted	e̠	ˤ	Pharyngealized	tˤ dˤ	̚	No audible release	d̚
̈	Centralized	ë	̴	Velarized or pharyngealized	ɫ			
̽	Mid-centralized	e̽	̝	Raised	e̝ (ɹ̝ = voiced alveolar fricative)			
̩	Syllabic	n̩	̞	Lowered	e̞ (β̞ = voiced bilabial approximant)			
̯	Non-syllabic	e̯	̘	Advanced Tongue Root	e̘			
˞	Rhoticity	ə˞ a˞	̙	Retracted Tongue Root	e̙			

Some diacritics may be placed above a symbol with a descender, e.g. ŋ̊

We all know that "the ear is too easily fooled" (Baken and Orlikoff, 2000).

Audio Recording

Rousselot expressed similar misgivings about IPA transcription some 114 years ago. In his obituary, Scripture (1925) wrote the following, "Coming to Paris in 1880, Rousselot was initiated into questions of Romance linguistics by Gaston Paris... Finding it impossible to detect the finer details of spoken sounds by ear, he complained to Gaston Paris, who said, 'Only mechanical registration will give you accurate knowledge. Attempts have been made in Marey's laboratory. Go and see!'" This exchange led Rousselot to invent the kymograph machine to record and account for speech sounds as accurately as possible (**Figure 5**).

Figure 5. *Abbé Rousselot and his kymograph machine. Available at bit.ly/3acqZ5K, license by Creative Commons.*

Rousselot's kymograph machine set the stage for the invention of other audio recording devices. The kymograph drew graphs in the frequency and intensity domains on the basis of the articulatory effort exerted in producing specific sounds. The first major breakthrough came in 1935 with the invention of the magnetic reel-to-reel tape recorder. The next seminal moment came in 1963 when Philips introduced the portable cassette player. The adjective "portable" must be taken under advisement because these early audio recorders were bulky and heavy. Ian Maddieson tells of his incredible phonetic field recording adventures in Africa. Their audio recorders and their accessories were so heavy that he and Peter Ladefoged hired porters to carry their equipment as they trudged from place to place. Ladefoged and Maddieson made good use of their equipment for language documentation. The University of California, Los Angeles (UCLA) Phonetic Lab Archive (1996; available at bit.ly/30fe3aJ) and their book *The Sounds of the World's Language* (Ladefoged and Maddieson, 1996) are the fruit of their dedication to documenting phonetic diversity. Ladefoged (1968) was first among linguists to have used this method for documenting 61 of the 1,436 Niger-Congo languages spoken in West Africa. Increasingly, documentary linguists are making use of portable camera recorders to capture aspects of everyday life and festivals. Some have produced excellent documentaries about endangered languages using digital audiovisual devices (available at bit.ly/2NYdu2t).

Visualization Technologies

Linguists have used inventiveness and creativity in adapting medical devices for language documentation. Ladefoged (2003) displays many "mom-and-pop" palatographs (devices to capture imprints of where articulators come into contact to produce sounds), airflow equipment, and a laryngoscope machine. Nowadays, researchers are using very sophisticated tools such as portable ultrasound and electroglottography (Whalen et al., 2011). Butcher (2013), Esling (2013), and Tabain (2013) provide excellent descriptions of various technologies used currently in language documentation and acoustic phonetic fieldwork.

Sound Spectrography

The sound spectrograph machine was invented at Bell Laboratories in the 1950s. Unfortunately, its prohibitive cost prevented its widespread usage. Fromkin (1985) indicates that in 1952, for the whole of England, there was only one sound spectrograph, which was owned by David Abercombie. When Fry (1955) needed to conduct his groundbreaking experiment on the acoustic correlates of lexical stress in English, he had to travel all the way from England to the Bell Laboratories in New Jersey to conduct his experiments. Fortunately, in 1995, Paul Boersma and David Weenink (2020) developed Praat, a free software platform for online speech analysis. It has become so popular that Watt (2013) refers to it as "the industry standard for acoustic analysis of speech." Ladefoged (2003) likens

spectrographs to a magnifying glass that bring aspects of speech sounds into sharp focus. They make it possible for speech sounds to be examined in three dimensions: frequency, intensity, and duration.

Africa and the Preservation of the World's Phonetic Diversity

Now that the UN has declared IDIL 22-32, greater efforts and monetary investments should be made for the documentation of all critically endangered languages. African languages should receive considerable attention for two reasons. First, critically endangered languages in Africa are underdocumented. Kandybowicz and Torrence (2017) mention 308 such languages. Sadly, they report that 201 African languages have gone extinct without ever being documented. This is a huge loss for the scientific understanding of phonetic diversity because, as Ladefoged (2007) notes, "endangered languages may contain even more unusual phenomena." He opines that they may have as many as 600 consonants and 200 different vowels that are not found in safe languages. The second reason for focusing on African languages is eloquently stated by Clements (2000, p. 123).

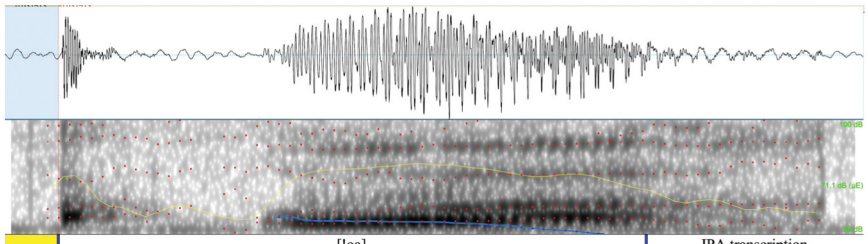
> "The African continent offers a generous sample of the great diversity of phonological systems to be found in the world's languages, as well as some original features of its own. African phonological systems range from the relatively simple to the staggeringly complex. Those on the more complex end of the spectrum contain phonemic contrasts little known elsewhere in the world, rich patterns of morphophonemic alternations, and intricate tonal and accentual systems, all offering stimulating grounds for phonetic and phonological study."

Brenzinger (2007, p.198) concurs and adds, "With about the third of the world's languages, the African continent is among the linguistically richest areas on the planet. Thus, a great deal of the future of linguistic diversity in general depends on what is going to happen to African languages." In other words, if one wants to really document phonetic diversity in the world's languages, one need not look further than the African languages. For example, the African languages are unrivaled by the quantity and diversity of their stop consonants. Stops are sounds that are made with two articulators coming together to block, albeit momentarily, the free flow of air molecules. There are 22 different types of stops in African languages. Plain stop sounds such as [p, p^h, b, t, t^h, d, k, k^h, g] are found in many languages outside of the African continent, including American English. However, the implosives [ɓ, ɗ, ɠ] (see bit.ly/2L74AhG) and the ejectives [p', k', t'] (see bit.ly/2YoxtJ4) occur more frequently in African languages than in languages elsewhere. As for the labiovelars [k͡p] and [g͡b] (see bit.ly/36pm8gj) and the clicks [ʘ, ǀ, !, ǁ, ǂ] (see bit.ly/3t6useB), they are almost exclusively found in African languages, except for an occasional sighting in the indigenous languages in Papua New Guinea (Hale, 1992). The labiovelars [k͡p] and [g͡b] are ubiquitous in most West African languages, whereas clicks are found mostly in a cluster of languages in southern Africa.

A Closer Look at Postulate 6

As noted earlier, most of the documentary efforts of the past three decades or so can be likened to hospice care. The focus is/was clearly on collecting artifacts about dying languages so that when they eventually die, future generations will still have some information about them. As commendable as these efforts are, documented linguistic artifacts cannot be used to generate novel utterances in the dead or dying language. A paradigm shift in documenting endangered languages is therefore needed during IDIL 22-32. For this to happen, documentation must undergo a change of mindset from the hospice care mentality to a palliative care mentality. This means that Postulate 6 must undergird future documentation efforts and goals, an idea echoed by Grenoble and Whaley (2006). This approach is currently being experimented on Jeju, (see bit.ly/3j1xPyS), a moribund language spoken in South Korea.

The new mindset under Postulate 6 calls for documenting endangered languages for use in speech-enabled technologies. The very first step in this process is speech digitalization. Rabiner and Schafer (1978) note that when speech is fully digitalized, it can be used for automatic speech recognition; Text-to-Speech; Speech-to-Text; voiced-enabled assistive technologies; digital transmission and storage of speech; speech synthesis; speaker verification and identification; and enhancement of signal quality. The speech digitalization process that makes these technological applications possible is summarized pictorially by Rabiner and Juang (1993). The methodology consists of extracting, measuring, and collecting all relevant features that are to be used to build voice-enabled intelligent systems. Koffi (2020) has amply explained and described how this can be applied to African languages. An oversimplified demonstration of how one might go about digitalizing a critically endangered language for speech synthesis is now discussed.



[!oa]				IPA transcription
TK O A				Arpabet transcription
!		o	a	Individual segments
	90 ms			VOT measurement of click [!]
74 Hz		141 Hz	131 Hz	F0/Pitch measurement
744 Hz		493 Hz	633 Hz	F1 measurement
1177 Hz		1003 Hz	1179 Hz	F2 measurement
2449 Hz		2507 Hz	2591 Hz	F3 measurement
72 dB		78 dB	76 dB	Intensity measurement
32 ms		89 ms	121 ms	Duration measurement

Figure 6. *Feature extraction of [!oa]. An annotated spectrograph such as this allows researchers to extract a number of essential parameters of a sound as the first step in the digitalization process.* ***Red dots****, formant trackers;* ***blue line****, pitch tracker;* ***yellow****, intensity tracker, VOT, voice onset time. See text for further explanation.*

Exemplification with !Xóõ

!Xóõ, known by the internationally recognized code ISO 639-3: nmn, is a critically endangered language spoken by 2,000 speakers in Botswana and 500 speakers in Namibia. According to Eberhard et al. (2019), its EGIDS rating is 6b. It has achieved celebrity status of a sort among endangered languages in Africa because of the richness of its clicks [ʘ, ǀ, !, ǁ, ǂ]. Ladefoged and Maddieson (1996, p. 246) report that "over 70% of the words in a !Xóõ dictionary begin with a click." The word [!oa] used in this demonstration is found in a sound file located at UCLA Phonetic Archive (available at bit.ly/3ouIC5J). The canonical syllable structure of !Xóõ is consonant vowel (CV) or consonant vowel[1] vowel[2] (CV^1V^2). The spectrograph in **Figure 6** displays the main acoustic correlates that are extracted from that word.

Ten tiers are created in **Figure 6** to extract and measure the relevant features of the word [!oa]. The first tier represents the IPA transcription. The second tier is an Arpabet transcription, which is a unique transcription system very much like the IPA. It was designed in the 1960s to allow a standard keyboard to be used to transcribe speech accurately without the need to resort to IPA symbols.

Because the Arpabet is fully compatible with the American Standard for Information Interchange Codes (ASCII), it is the premier transcription system used in coding for Text-to-Speech and Speech-to-Text for use in many voice-enabled applications (Jurafsky and Martin, 2000). The third tier represents the individual segments in [!oa] and the fourth measures the Voice onset time (VOT; see Blumstein, 2020). The VOT is the time interval between when articulators come together to produce [!]. F0 measures vocal fold vibration. Speech has three key formants, F1, F2, and F3, which appear as areas of concentration of acoustic energy in spectrographs. They are extremely important in speech synthesis. F1 correlates with the degree of opening of the mouth when a segment is produced. F2 indicates the horizontal movement of the tongue. F3 correlates with the state of the lips, whether they are rounded or unrounded. Intensity deals with the loudness of a sound and duration indicates the amount of time (in milliseconds) it takes to produce a sound. Taken together, these extracted tokens help to turn analog speech sounds into digits that become the necessary ingredients for speech synthesis.

Technologizing Endangered Languages

Voice-enabled speech technologies are ubiquitous in contemporary life for the speakers of English and other elite world languages. However, no such technology exists in any of the more than 2,000 indigenous languages of Africa. For the languages on the brink of extinction, those with EGIDS 8a/8b and EGIDS 9, developing voice-enabled software packages for use in mobile devices can be a game changer. Mobile phone usage has exploded in Africa. A 2015 Pew Research Center (see pewrsr.ch/3afYYdp) study found that, between 2002 and 2014, mobile phone usage increased tenfold, from 8% to 83%, in Ghana and other sub-Saharan

African countries. Imagine developing a simple voice-recognition app that allows the speaker of an endangered language to call another person by using voice commands such as "Call X" in the native tongue. Imagine also writing a basic software program for a talking dictionary based on the first 1,000 words in that language. Language revitalization lessons can be designed following the model in the Duolingo app. The app gives the speaker the chance to learn new words and use them in sentences of graded complexity. Technologizing endangered languages can create enthusiasm among their speakers and bring about revitalization along the lines envisioned in Postulate 6.

Methodological Challenges

Technologizing endangered languages is easier said than done. It involves multilayered expertise, including familiarity with speech synthesis. Kent and Read (2000) describe several speech synthetic models. The ones that are currently in vogue rely mostly on the diphone concatenation method. Diphone extraction consists of splicing every sound in two and measuring each half. This is done for every sound that occurs in a language. Unfortunately, this method is more amenable for use in languages that have been well studied phonetically and phonologically. However, for the hundreds of indigenous languages that have not yet been graphacized, using the diaphone method is extremely time consuming. For example, if we accept Clements' (2000, pp. 125, 134) typology that a prototypical African language has 21 consonants and 9 vowels, one would need to extract 900 diphones (30 × 30) for the databank. If 7 correlates are extracted for each diphone, one would need to extract and digitalize approximately 6,300 diphones. If this approach were to be used, !Xóõ would go extinct before a successful speech synthesis is achieved.

A Simpler Speech Synthesis Model

The diphone concatenation method is too onerous. For this reason, the Occam razor principle of scientific inquiry compels us to look for a simpler model that can achieve the same result or better with relatively less effort. One alternative method consists of extracting formant data from syllables. A syllable-based speech synthesis is appealing for at least six compelling reasons. First, the syllable has a long history in human linguistic experiences. Second, both literate and preliterate societies take syllables into account in their songs and in various language games. Third, the syllable is a key building block in learning to read. Fourth, astounding insights have accumulated over the past 40 years that make syllable-based speech synthesis theoretically sound. Fifth, the vast majority of world languages have relatively simple syllable structures. Last but not least, syllable-based synthesis is relatively less time- onsuming because the number of possible syllables is far fewer than the number of diphones. According to Clements (2000), the preferred syllable structure of African languages is CV. This means that there are 189 possible CV syllables (21 consonants × 9 vowels). This number grows to 533 when canonical syllables such as VC, NV, CV^1. and V^2 are taken into account. A syllable-based synthesis requires 3,731 tokens instead of the 6,300 needed for speech synthesis based on diphone concatenation.

Conclusion

It follows from the analytical sketches outlined above that documenting endangered languages in accordance with Postulate 6 is beyond the expertise of a single linguist. Consequently, cross-disciplinary collaboration with people in other fields should be the hallmark of language documentation during IDIL 22-32. Naturally, because linguists have expertise in graphicization, phonetic transcription, Arpabet transcription, fieldwork, and acoustic phonetic feature extraction, they should lead documentation efforts. However, their efforts should be augmented by expertise in engineering (signal processing), coding, computer science, and intelligent systems design. The addition of speech synthesis to the tools that documentary linguists are already using will prove to be extremely beneficial for endangered languages during IDIL 22-32 and beyond. If speakers can use their native tongues to access various technological applications, we believe as Crystal (2000) does, that Postulate 6 will increase the chances of survival of critically endangered languages.

Acknowledgments

I am grateful to Gary Simons, one of the editors of *Ethnologue*, for making the endangered language map available to me and for giving me the Expanded Graded Intergenerational Disruption Scale (EGIDS) data from the Americas, the Pacific, and Asia. Many thanks to Arthur Popper for insightful critiques that have improved the readability of this paper.

References

Baken, R. J., and Orlikoff, R. F. (2000). *Clinical Measurement of Speech and Voice*, 2nd ed. Singular Publishing Group, San Diego, CA.

Batibo, H. (2009). Poverty as a crucial factor in language maintenance and language death: Case studies from Africa. In W. Harbert, S. McConnell-Ginet, A. Miller, and J. Whitman. (Eds.). *Language and Poverty*, Multilingual Matters, Buffalo, NY, pp. 23-36.

Baugh, J. (2009). Econolinguistics. In W. Harbert, S. McConnell-Ginet, A. Miller, and J. Whitman. (Eds.), *Language and Poverty*. Multilingual Matters, Buffalo, NY, pp. 67-76.

Blumstein, S. E. (2020). When speech goes wrong: Evidence from aphasia. *Acoustics Today* 16(4), 12-19. https://doi.org/10.1121/AT.2020.16.4.12.

Boersma, P., and Weenik, D. (2020). *Praat: Doing Phonetics by Computer*. Available at https://bit.ly/3dU4zcR.

Brenzinger, M. (2007). Language endangerment in southern and eastern Africa. In M. Brenzinger (Ed.), *Language Diversity Endangered*. Mouton de Gruyter, New York, NY, pp. 79-204.

Butcher, A. (2013). Research methods in phonetic fieldwork. In M. J. Jones and R. Knight (Eds.), *The Bloomsbury Companion to Phonetics*. Bloomsbury, New York, NY, pp. 57-78.

Clements, N. G. (2000). Phonology. In H. Bernd and D. Nurse (Eds.), *African Languages: An Introduction*. Cambridge University Press, New York, NY, pp.123-160.

Coulmas, F. (1989). *The Writing Systems of the World*. Basil Blackwell, Cambridge, MA.

Crystal, D. (2000). *Language Death*. Cambridge University Press, New York, NY.

Eberhard, D. M., Simons, G. F., and Fennig, C. D. (Eds.). (2019). *Ethnologue: Languages of Africa and Europe*, 22nd ed. SIL International, Dallas, TX.

Esling, J. (2013). Voice and phonation. In M. J. Jones and R. Knight (Eds.), *The Bloomsbury Companion to Phonetics*. Bloomsbury, New York, NY, pp. 110-125.

Fromkin, V.A. (1985). Introduction. In V. A. Fromkin (Ed.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*. Academic Press, Inc., San Diego, CA, pp. 1-14.

Fry, D. B. (1955). *The Physics of Speech*. Cambridge University Press, New York, NY.

Gambarage, J. J. (2017). Unmasking the Bantu orthographic vowels: The challenge for language documentation and description. In J. Kandybowicz and H. Torrence (Eds.), *Africa's Endangered Languages: Documentary and Theoretical Approaches*. Oxford University Press, New York, NY, pp. 449-484.

Grenoble, L. A., and Whaley L. J. (2006). *Saving Languages: An Introduction to Language Revitalization*. Cambridge University Press, New York, NY.

Hale, K. (1992). Language endangerment and the human value of linguistic diversity. *Language* 68(1), 35-42.

International Phonetic Association (1999). *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*. Cambridge University Press, New York, NY.

Jurasfsky, D., and Martin, J. H. (2000). *Speech and Language Processing: An Introduction to Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall, Upper Saddle River, NJ.

Kandybowicz, J., and Torrence, H. (2017). Africa's endangered languages: An overview. In J. Kandybowicz and H. Torrence (Eds.), *Africa's Endangered Languages: Documentary and Theoretical Approaches*. Oxford University Press, New York, NY, pp. 427-447.

Koffi, E. (2012). *Paradigm Shift in Language Planning and Policy: Game-Theoretic Solutions*. De Gruyter Mouton, New York.

Koffi, E. (2020). A tutorial on acoustic phonetic feature extraction for automatic speech recognition (ASR) and text-to-speech (TTS) applications in African languages. *Linguistic Portfolios* 10, 130-153.

Krauss, M. (1992). The world's language in crisis. *Language* 68(1), 4-10.

Ladefoged, P. (1968). *A Phonetic Study of West African Languages: An Auditory-Instrumental Survey*, 2nd ed. Cambridge University Press, New York.

Ladefoged, P. (2003). *Phonetic Data Analysis: An Introduction to Fieldwork and Instrumental Techniques*. Blackwell Publishing, Malden, MA.

Ladefoged, P. (2007). *Preserving the Sounds of Endangered Languages*. Available at https://bit.ly/2ZZZJCP.

Ladefoged, P., and Maddieson, I. (1996). *The Sounds of the World Languages*. Blackwell Publishers, Malden, MA.

Maddieson, I., and Ladefoged, P. (1989). Multiply articulated segments and the feature hierarchy. *UCLA Working Papers on Phonetics* 72, 116-138.

McWhorter, J. (2003). *The Power of Babel: A Natural History of Language*. Perennial, New York, NY.

Mufwene, S., and Vigoroux, C. B. (2008). Colonization, globalization and language vitality in Africa: An introduction. In C. Vigoroux and S. Mufwene (Eds.), *Globalization and Language Vitality: Perspectives from Africa*. Continuum, New York, NY, pp. 1-50.

Pulum, G. K., and Ladusaw, W. A. (1986). *Phonetic Symbol Guide*. The University of Chicago Press, Chicago, IL.

Rabiner, L. R., and Juang, B.-H. (1993). *Fundamentals of Speech Recognition*. Prentice Hall, Englewood Cliffs, NJ.

Rabiner, L. R., and Schafer, R. W. (1978). *Digital Processing of Speech Signals*. Prentice Hall, Englewood Cliffs, NJ.

Sands, B. (2017). The challenge of documenting Africa's least-known languages. In J. Kandybowicz and H. Torrence (Eds.), *Africa's Endangered Languages: Documentary and Theoretical Approaches*. Oxford University Press, New York, NY, pp., 11-38.

Scripture, E. W. (1925). Obituary: Abbe Rousselot. *Nature* 115(2883), 164-165.

Tabain, M. (2013). Research methods in speech production. In M. J. Jones and R.-A. Knight (Eds.), *The Bloomsbury Companion to Phonetics*. Bloomsbury, New York, NY, pp. 39-56.

Tucker, B. V., and Wright, R. (2020). Speech acoustics of the world's languages. *Acoustics Today* 16(2), 56-64. https://doi.org/10.1121/AT.2020.16.2.56.

UNESCO (2010). *Atlas of the World Languages in Danger*, 3rd ed. UNESCO, Paris, France.

University of California, Los Angeles (UCLA) (1996). *UCLA Phonetic Lab Archive*. Available at https://bit.ly/30fe3aJ.

Watt, D. (2013). Research methods in speech acoustics. In M. J. Jones and R.-A. Knight (Eds.), *The Bloomsbury Companion to Phonetics*. Bloomsbury, New York, NY, pp. 79-97.

Whalen, D. H., Dicanio, C. T., and Shaw, P. A. (2011). Phonetics of endangered languages. *Acoustics Today* 7(4), 35-42.

About the Author

Ettien Koffi
enkoffi@stcloudstate.edu

Department of English
St. Cloud State University
Webster Hall
720 Fourth Avenue South
St. Cloud, Minnesota 56301-4498, USA

Ettien Koffi earned a BA in linguistics from the University of Abidjan (Abidjan, Côte d'Ivoire) and a PhD in linguistics from Indiana University (Bloomington, IN). He is a professor of linguistics at Saint Cloud State University (St. Cloud, MN); the author of five books; and the author/coauthor of several dozen articles. His acoustic phonetic research is synergetic, encompassing L2 acoustic phonetics of English (speech intelligibility from the perspective of the Critical Band Theory), sociophonetics of Central Minnesota English, general acoustic phonetics of Anyi (a West African language), and acoustic phonetic feature extraction for application in Automatic Speech Recognition, Text-to-Speech, Speech-to-Text, Voice Biometrics, and Intelligent Systems.