Original Paper

Phonetica 1997;54:94–111

Received: July 19, 1996
Accepted: January 6, 1997

Phonetic Structures of Banawá, an Endangered Language

Peter Ladefoged[a], *Jenny Ladefoged*[a], *Daniel Everett*[b]

[a] Phonetics Laboratory, Linguistics Department, UCLA, Los Angeles, Calif.,
[b] Linguistics Department, University of Pittsburgh, Pa., USA

Abstract

This paper describes the phonetic characteristics of Banawá, an endangered language spoken in Brazil. The qualities of the Banawá vowels are described in terms of their formant frequencies. The places of articulation of each consonant, the voice onset time and the manner of articulation are documented. The structure of syllables and words is delimited, and the location of stressed syllables is described and verified experimentally.

Nobody knows exactly how many languages there are in the world today, but there is little doubt that there are more now than there will be in our children's time. The best estimates of the numbers of languages and speakers are those in the *Ethnologue* [Grimes, 1992], which lists nearly 7,000 distinct languages. Over half of the languages listed are spoken by under 10,000 people, about a quarter of them are spoken by under 1,000 people, and about 10% by under 100. It will be very difficult for languages spoken by small numbers of people to survive.

One of these endangered languages is Banawá, an Arawan language, spoken by about 75 speakers deep in the rain forest in northern Brazil (see fig. 1). The group has been known to non-indigenous people for a comparatively short time, as they were first contacted by missionaries only 30 years ago. Since that time they have acquired many Brazilian characteristics; they have already learned to play football, Brazil's national pastime, which they eagerly watch on television when they manage to get out of the forest. They will probably become more completely assimilated into Brazilian culture in the near future, and their language may not be spoken in a generation or two. Irrespective of one's views on cultural change, there is an obvious scientific necessity to document the phonetic structures of the language while speakers are still available. This paper will provide some of this documentation through projects sponsored by the National Science Foundation to which we are very grateful.

When describing the phonetic structures of any language we would like to describe all and only the linguistically relevant phonetic facts. Unfortunately, the resources that we have available for the description of a single language do not permit this. This account of Banawá is limited in two ways. Firstly, we will describe only the major phonetic structures – the quality of the vowels in limited contexts, the principal

KARGER

E-Mail karger@karger.ch
Fax+41 61 306 12 34
http://www.karger.ch

Peter Ladefoged
Phonetics Laboratory
Linguistics Department
UCLA
Los Angeles, CA 90095–1543 (USA)

Fig. 1. The location of the approximately 75 speakers of Banawá.

consonantal properties, and some aspects of the suprasegmental features. This is all that can be done within the limits of the present NSF project, if we are to consider a number of endangered languages for which even this information is not available. Secondly, it is based on the speech of only five male speakers. We were unable to collect data from a larger group including female speakers. However, the limitation to only five speakers may not be as unsatisfactory as it first appears. There are probably less than 30 adult male speakers of Banawá, so a sample of five speakers still represents a larger percentage of the adult male population than is used as a basis for most phonetic descriptions.

The first necessity for an account of the phonetic structures of any language is a good phonological description of that language. There are two previous publications on the language [Buller et al., 1993; Everett, 1995] which, supplemented by the third author's knowledge of the language, enabled us to determine appropriate material to record. The data for this study consist principally of recordings of 6 male speakers made in July 1995. The recordings of one of the speakers proved unusable for quantitative work, as he spoke so quietly and shyly that we eventually had to abandon attempts to make reliable analyses. Two of the speakers were recorded at the SIL Center in Porto Velho, where they worked extensively with us in preparing the word lists and other materials. The other speakers were recorded in the Banawá settlement, which is in the jungle about 120 miles due North of Porto Velho. All the recordings were made on a DAT recorder, using a close-talking, noise-canceling microphone for each speaker individually. The frequency response was substantially flat throughout the audible range, and the signal/noise ratio was greater than 45 dB.

Table 1. Words illustrating the vowels of Banawá in stressed syllables after **b** and **t**, and in unstressed syllables after **b** and **f**

Stressed		Unstressed	
bita	mosquito	**ibi**	each other
befa	other	**ibe**	a strip
bata	to pick	**iba**	to put/place
bufa	put on water	**ibufa**	to dump into water
tifa	drink water	**tafi**	eating
tefe	food (m.)	**tafe**	food (f.)
tafa	to eat	**tafa**	to eat
tufa	to block in	**tafu**	to eat

Vowels

Banawá has four vowels, **i, e, a, u,** illustrated by the words in table 1. We analyzed two tokens of each of the vowels in the words in table 1, as spoken by each speaker (a total of 5 speakers with 2 tokens of 16 words = 160 vowels), using a Kay CSL system. Formant frequencies were determined from observations of the formant histories throughout the word, and superimposed FFT and LPC spectra made at the most steady-state portions of each vowel.

Even with good computer analysis facilities, the determination of the formants – the resonances of the vocal tract, as opposed to the peaks in the observed spectra – cannot be made entirely algorithmically. There are often cases where LPC spectra either do not provide a sharp peak or provide a spectral peak that is clearly not a resonance of the vocal tract. In addition, the peaks in FFT spectra may not be well defined. If the number of points in the FFT is large (the equivalent of a narrow-band filter) then the individual harmonics will be shown. The formant resonances will not each be defined by a single peak, but will have to be estimated as being somewhere between the two highest harmonics in the appropriate region. If the number of points in the FFT is small (the equivalent of a wide-band filter) each formant will usually be defined by a single peak in the spectrum. But when two formants are close together there may be only a single peak for the two of them. Furthermore the time interval will reflect only a particular part of a glottal pulse, which will sometimes be when the glottis is closed, and sometimes when it is open; the vocal tract resonances will be different in these two conditions. All these factors lead to the possibility of error in the determination of the formant frequencies from the acoustic data.

One way of checking on the reliability of the measurement procedures is to compare the first two formants in the two tokens of each word, as was first done by Peterson and Barney [1952]. There is a confounding factor in this procedure in that comparison of the two sets of measurements shows both differences in the tokens produced, and differences due to the accuracy of the measurement techniques; the speaker might have used a different pronunciation, or there might have been measurement error. Variation between tokens could be avoided by comparing two sets of measurement of the same token. However, errors in determining the true resonances of the vocal tract are more likely to arise when considering different tokens which may have slight differences in the glottal source (a common and unnoticed variation in the production of the same vowel in a language without phonologically contrastive phonation

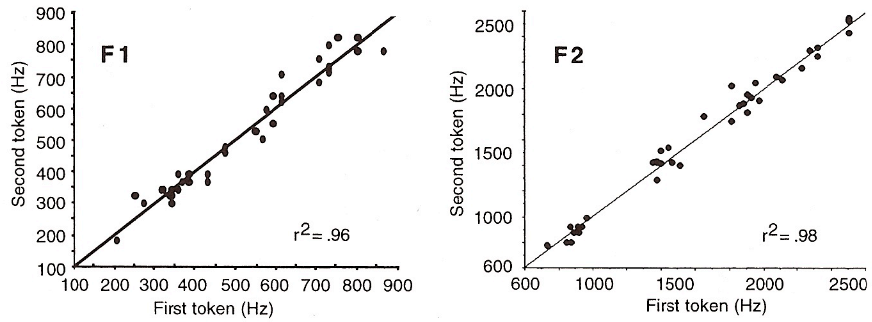

Fig. 2. The correlation between two measurements of each of F1 and F2, one measurement in the first token of each of the words containing stressed vowels, and the other in the second token of the same word.

types). The FFT and LPC procedures are affected by differences in the glottal source function even when the formant resonances remain the same. We want to have a valid representation of each vowel type as produced by each speaker separately, irrespective of any phonation type difference or any other non-contextual variation that might have been made. Accordingly, we can conveniently combine the two sources of variation (within-speaker variation and measurement error) and check the reliability of our representation by comparing two tokens of each vowel as produced by each speaker. (When considering within-speaker variability, it would have been nice if we had had more than two repetitions of each vowel by each speaker. But we did not, and, as shown by Johnson et al. [1995], between-speaker variability, which we consider later, is a much greater than within-speaker variability.)

The results for the stressed vowels are as shown in figure 2. It may be seen that there is a good correlation between the two sets of measurements. The higher values of F1 have some differences, but the measurements of F2 are very similar in the two repetitions of the same word.

The same procedure was used in a comparison of the unstressed vowels, as shown in figure 3. The F1 differences are slightly greater, but the F2 differences are very much the same as in the case of the stressed vowels. For both the stressed and the unstressed vowels, the high correlations indicate that the measurements of the formant frequencies are reliable.

Figure 4 shows the first two formant frequencies of the vowels as produced by the five speakers saying two tokens of each of the words in table 1 illustrating the stressed vowels (20 points for each vowel, some of which are overlapping). The values on the scales are in hertz, but the distances on the scales are arranged so that equal distances represent equal bark intervals of each formant. The ellipse in this figure encloses four vowels of speaker 4 which are significantly different in their F1 values from those of F1 in the same words as spoken by the other 4 speakers ($p<0.01$ in an ANOVA of F1 by speaker by vowel). This speaker has an aberrant vowel, with a higher F1 (a more open vowel) than that of other speakers.

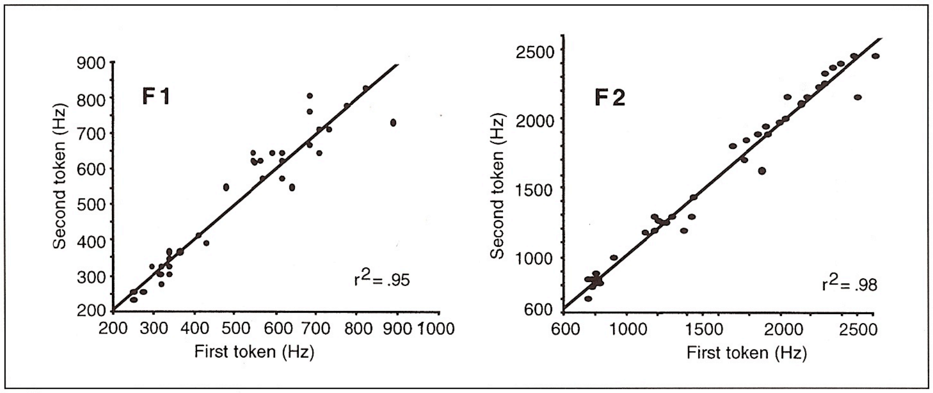

Fig. 3. The correlation between two measurements of each of F1 and F2, the one measurement in the first token of each of the words containing unstressed vowels, and the other in the second token of the same word.

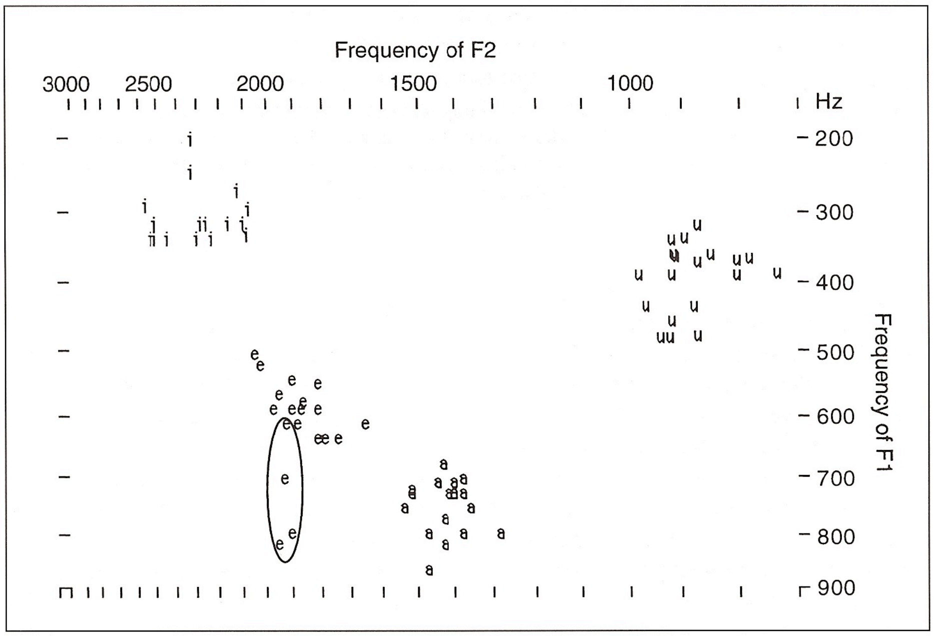

Fig. 4. The first two formant frequencies of the stressed vowels in table 1, as spoken by 5 speakers. The ellipse encloses the four vowels of speaker 4 which are discussed in the text. See text for discussion of the scales and axes.

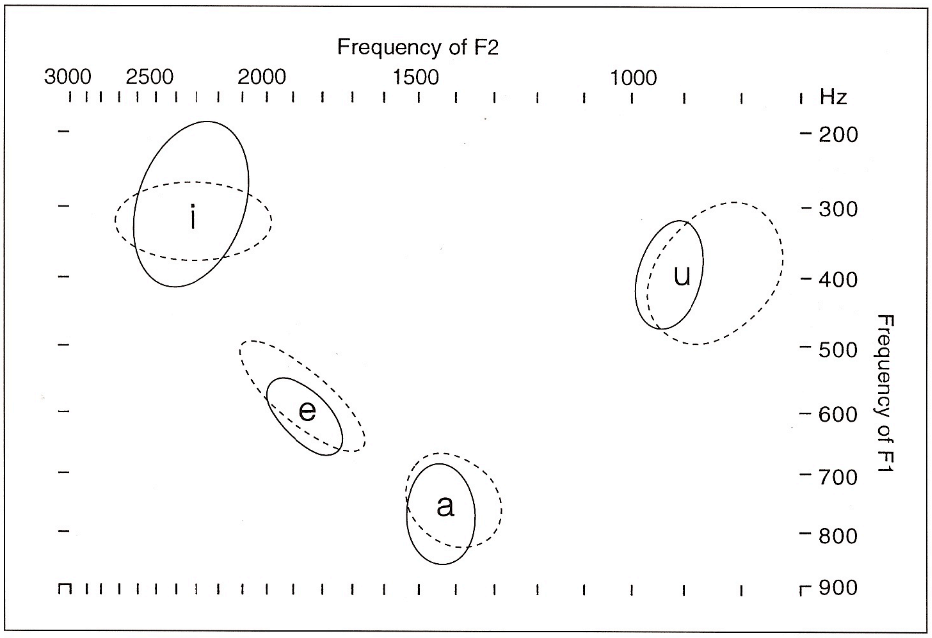

Fig. 5. Formant plots of the Banawá vowels in table 1, as produced by five speakers. The ellipses drawn with solid lines show areas containing all points within two standard deviations of the mean for vowels in stressed syllables after a coronal consonant. The dashed ellipses show the same range for vowels in stressed syllables after a bilabial consonant. The vowel symbols are placed at the grand mean for each vowel, irrespective of context.

Figure 5 shows the variation in the first two formant frequencies of individual tokens in figure 4, except that speaker 4's **e** vowels have been omitted from the calculations. The first step in plotting this graph was to determine the principal components of the variation in the F1–F2 space shown in figure 4, considering vowels after labials and vowels after coronals separately. Ellipses were centered on the means for the vowels in each of these contexts. The axes of the ellipses were angled along the principal components, and the radii were made equivalent to two standard deviations of the mean variation in these dimensions. Assuming our five speakers constitute a representative sample of the approximately 30 adult male speakers of Banawá, we can say that 95% of them, i.e. about 26 or 27 Banawá adult males, have vowels with formant frequencies that lie within the ellipses shown in figure 5, with some possible extra variation for the vowel **e**, where one of our speakers had a different vowel quality.

The differences between vowels after bilabials and vowels after coronals are shown in table 2. As can be seen from both figure 5 and table 2, formant frequencies are generally slightly lower after a bilabial consonant than after a coronal consonant. An analysis of variance shows that the lowering of F2 in **u** is very significant ($p<0.001$), but that none of the other differences are even probably significant ($p<0.05$).

Table 2. Differences (Hz) in formant frequencies after bilabials as compared to vowels after coronals

	i	**e**	**a**	**u**
F1	23.9	– 33.0	–27.5	– 5.2
F2	– 5.0	11.4	–36.6	–46.9**
F3	–56.5	–130.4	95.0	–18.3

** = $p < 0.001$.

Table 3. Statistics for mean formant frequencies for all five speakers (except for **e** in the case of speaker 4) for the first three formants for the four vowels

	Count	Mean	SD	SE
F1				
i	20	311	47.1	10.5
e	16	591	41.0	10.2
a	20	758	46.6	10.4
u	20	396	48.6	10.9
F2				
i	20	2,320	164.0	36.7
e	16	1,869	101.1	25.2
a	20	1,422	58.9	13.2
u	20	883	63.8	14.3
F3				
i	20	2,822	181.8	40.6
e	16	2,741	178.5	44.6
a	20	2,593	279.3	62.4
u	20	2,685	164.1	36.7

The plots in figure 5 show that the three vowels **i**, **e**, **a** differ in vowel height from one another, taking F1 as a measure of the feature Height. There is also an apparent difference in height between **i** and **u**. An analysis of variance indicates that when all 5 speakers are taken together, this difference is significant ($p < 0.01$). This is an interesting point in that in work in progress on other endangered languages with a limited number of vowels is showing that the high back vowel is often not as high as the front vowel. In Banawá, the high back vowel has been written in two different ways. Everett transcribes it as **u** [Everett, 1995, 1996; and in Buller et al., 1993], but Buller and Buller in their field notes [pers. commun.] transcribe it as **o**. The ANOVA might be taken as indicating that Bullers' transcription is preferable; however, the same analysis shows that there is an interaction between speakers and vowels, and for two of the five speakers there is no statistical difference in Height between **i** and **u**. In Banawá, the difference in Height between **i** and **u** seems to be speaker dependent; in this paper we will continue to use **u**, the more usual symbol for a high back vowel. The statistics for all five speakers (except for **e** in the case of speaker 4) for the first three formants for the four vowels are shown in table 3.

The ellipses in figure 5 show that feature values cannot be given context-free specifications. Thus [front] (one of the possible values of the feature Back) means one thing when it occurs in the context of [high] (one of the possible values of the feature

Height), and another when it occurs with [low]. If we regard the implementation of the feature Back as being dependent on F2, then [front] has a value of 2,320 Hz when it is in conjunction with [high] for the vowel **i,** and 1,869 Hz for the vowel **e** when it is in conjunction with [low]. Another way of specifying the implementation of the feature Back is to regard it as a function of the distance between the first two formants (the old Jakobsonian definition of Diffuse). In that case, the feature specification [front] would have a value of 2,320 – 311 = 2,009 Hz for a [high] vowel, but this same feature specification would have a value of 1,869 – 591 = 1,278 Hz for a [low] vowel. Whatever measure one chooses, it appears that the phonetic interpretation of Back depends on whether a high vowel or a low vowel is being described. The context-dependent nature of feature specifications has long been noted by phoneticians [Ladefoged, 1972]. We will consider another example in this paper when we discuss differences in voicing.

The tables and figures show that these vowels are more in line with the theory proposed by Lindblom [1990], requiring vowels to be what he would call adequately dispersed in the auditory vowel space, rather than with the articulatory/acoustic notion of quantal vowels proposed by Stevens [1989]. The low vowel **a** is too far forward (has too high an F2) to be equated with the quantal vowel **ɑ**, which Stevens defined as one in which F1 is maximal and F2 minimal so that they are close together. He notes that this requires 'a *backed* and low tongue position' [Stevens 1989, p. 14, emphasis added]. Banawá, like many other languages with five or fewer vowels, has a low central rather than a low back vowel.

Consonants

The consonants of Banawá are as shown in table 4, and illustrated in tables 5 and 6. The more specific symbols **ɸ** and **ɾ** are used in the chart in table 4; elsewhere, so that the examples are more readable, the more general symbols **f** and **r** are used. Table 5 shows two examples of each consonant in word initial position before **a** and **i**, and table 6 has two examples of each consonant in word medial position between **a_a** and between **i_i**.

The voiced stops **b**, **d**, **ɟ** are voiced throughout (although, as we will see, the voicing in **ɟ** is often not fully evident). The voiceless stops **t** and **k** are virtually unaspirated. Measurements were made of two tokens of each of the words in table 7. The mean voice onset times (VOTs) for **t** and **k** in the different phonetic contexts provided by these words are shown in figure 6. An analysis of variance showed that there was no difference ($p > 0.05$) between the measurements for initial and medial occurrences, and these measurements have been pooled. Accordingly each column represents the mean of 6 measurements, except in two cases when faults in recording the first two speakers resulted in only four valid measurements of the first word, **tisi**, 'fall'.

The first point to note about figure 6 is that the VOT for **t** (the white and black columns) is much less than that for **k** (the shaded columns). An analysis of variance shows that for each speaker this is a significant difference ($p < 0.01$). The mean VOT for velars is 44 ms, whereas that for dentals is 22 ms. There is also a smaller but still significant difference ($p < 0.01$) in the VOT of **t** before **i** (the white columns) and before **a** (the black columns). There is no significant difference in the VOT of **k** before different vowels, three speakers having a longer VOT before **a**, and the other two before **i**.

Table 4. Banawá consonants

	Bilabial	Dental	Palatal	Velar	Labial-velar	Glottal
Stop	**b**	**t d**	**ɟ**	**k**		
Nasal	**m**	**n**				
Tap		**ɾ**				
Fricative	**ɸ**	**s**				
Approximant					**w**	**h**

Table 5. Examples of each consonant in word initial position before **a** and **i**

baka	(name)	**bisi**	to pinch
tafa	eat	**tisi**	fall
daka	(name)	**disi**	wasp
ɟaka	walk	**ɟiri**	to shock
kaka	toucan	**kisi**	descend
maka	snake	**misi**	up
naka	sticky	**nisa**	down
faki	twist	**fisi**	monkey
saka	jab, pierce	**sisi**	a few
rawi	write	**risa**	down on (f.)
waka	break	**wisi**	to cut
haku	spider	**hisi**	sniff

Table 6. Examples of each consonant in word medial position between **a_a** (or **u_a**) and **i_i**

baba	(name)	**kibi**	full
bata	rotten	**kiti**	strong
bada	(name)	**bidi**	small
baɟa	palm fronds	**diɟi**	to wobble
baka	(name)	**kiki**	to look to the side
bama	catfish	**kimi**	corn
baa ana	she hits	**kini**	green
bafa	Bafa (river)	**kifi**	to cross a bridge
basa	to put a stick up high	**kisi**	to descend
baa ara	another woman	**kiri**	I am itching
baa uwa	I hit	**tiwi**	did you see?
haha	to laugh	**kihi**	potato

Table 7. The words used for VOT measurements

	t		k	
Initial	**tisi**	fall	**kisi**	descend
	tima	upstream	**kimi**	corn
	tafa	eat	**kaka**	toucan
	tama	vine	**kama**	come
Medial	**kiti**	strong	**kiki**	to look to the side
	bata	rotten	**baka**	(name)

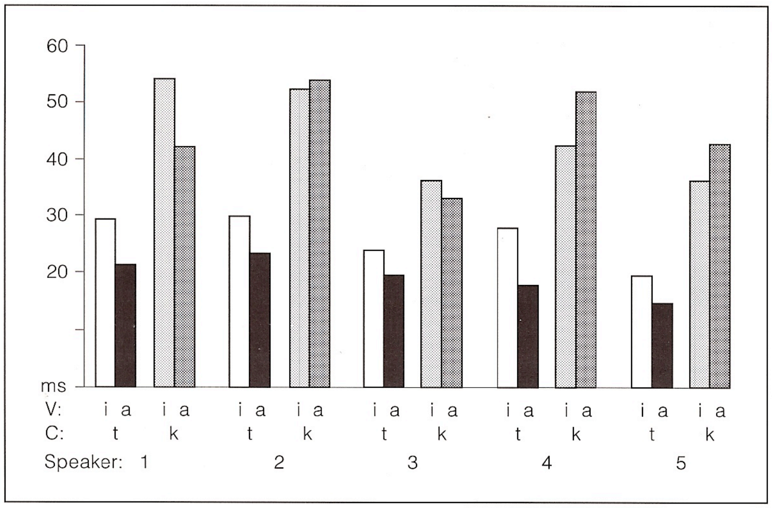

Fig. 6. The VOTs foar the two vioceless stops **t** and **k** as produced by the five speakers before the vowels **i** and **a**.

The dependence of VOT on place of articulation (and, to a lesser extent, on the quality of the following vowel) has been noted for many languages, going back at least to Fischer-Jørgensen [1954]. This may be another example of the interdependence of features. One has to know whether one is talking about dental or velar stops before one can say how aspirated a stop might be. In other words, the value of the feature Voicing cannot be stated without taking into account the co-occurring value of the feature Place. This may be partly an artifact of our form of description. The best phonetic exponent of the feature Voicing may not be VOT but a measure of the difference in timing between two gestures, one for the stop, and one for the glottal adjustments. If this measure could be stated correctly, at least some of the differences in VOT between velar and dental stops might be seen to be due to concomitant aerodynamic and articulatory differences that we do not yet fully understand. However, some of the differences between places of articulation are language specific. In their extensive study, Lisker and Abramson [1964] report that Dutch unaspirated **t** has a mean VOT of 15 ms and **k** of 25 ms, a difference of 10 ms, whereas Cantonese **t** has a mean VOT of 14 ms and k of 34 ms, a difference of 20 ms. In Banawá the mean VOT dentals is 22 ms and that for velars is 44 ms, a difference of 22 ms. We have to know both what language is being described and which place of articulation is involved before we can assign a phonetic value to the feature Voicing.

All our speakers used a bilabial fricative **ɸ** rather than **f**. As noted, we have used the more common symbol, in accordance with the principle of using a simple phonemic transcription whenever possible in order to make the examples easier to read. We use the more exotic symbol only when specifically referring to the bilabial character of this fricative. An aspect of all the bilabial sounds, particularly noticeable in **m** and **ɸ**, but also evident in **b**, is that the release of the articulation has a **w**-like component, so that these sounds might have been transcribed $\mathbf{m^w}$, $\mathbf{ɸ^w}$, $\mathbf{b^w}$. This component, which

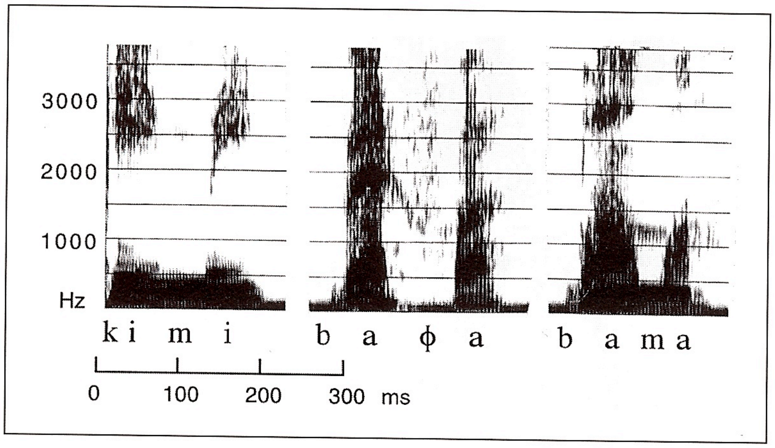

Fig. 7. Spectrograms of **kimi**, 'corn', **bafa**, 'Bafa' (name of a river), and **bama**, 'catfish'.

can be clearly seen in the spectrograms in figure 7, may not be a labial velar of the kind regularly symbolized by **w**. The lowering of the formants (notably F2) may be simply due to the protrusion of the lips that occurs during the course of these sounds.

All the consonants **t**, **d**, **n**, **s** have laminal articulations. Three of our speakers used interdental articulations for **t**, **d**, **n**. Some of our speakers had rather poor dentition, so it was possible to observe that the fricative **s** was usually made a little further back (it was never interdental); but it was always further forward than the typical English alveolar **s**. The tap which we have symbolized **ɾ** has a dental articulation in which the tongue tip is raised from the floor of the mouth, moving slightly forward as it strikes the anterior part of the alveolar ridge. Spectrograms of this sound in initial and medial position are shown in figure 8.

The voiceless approximant **h** may be accompanied by nasalization of the adjacent vowels, but this is by no means certain. The nasalization of vowels adjacent to **h** (and ʔ in other languages) appears to be an areal feature of languages spoken in this region. In the case of Banawá, all our speakers had vowels in words such as **haha** 'to laugh' and **kihi** 'potato' that were clearly breathy (and therefore had auditory characteristics in common with nasalized vowels), but spectrographic analysis did not reveal any evidence of a nasal pole/zero pair in any of these vowels. Unfortunately, we were unable to obtain any aerodynamic or other physiological records that would throw more light on this problem.

Previous accounts of Banawá [Everett, 1995, 1996, Buller et al., 1993] suggested that the consonants **ɹ** and **w** are not part of the phonological inventory of Banawá, but are simply the result of glide formation when underlying **i** and **u** are syllable initial before a vowel. This is somewhat reminiscent of a proposal by Jakobson et al. [1951] in which they suggested that English words such as 'yield' and 'woo' could be transcribed as /iiild/ and /uuu/. This is possible in English because there is no contrast between word-initial or final **j**, **w** and **i**, **u**, respectively. There are good phonological reasons for these identifications in Banawá. Morphological structures and interlexical

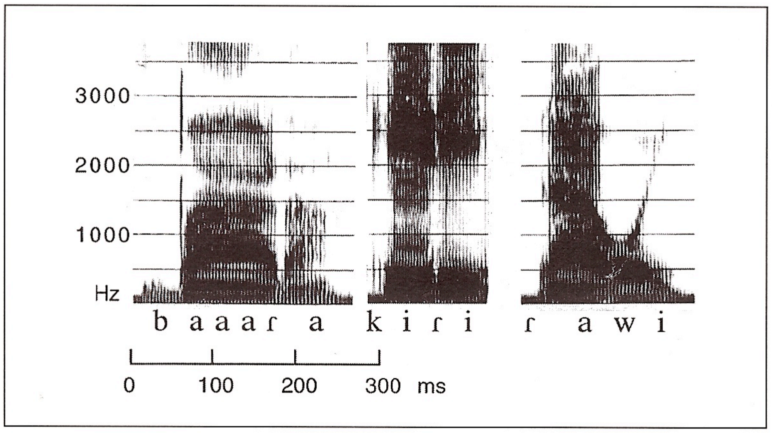

Fig. 8. Spectrograms of **baa aɾa**, 'another woman', **kiɾi**, 'I am itching', **ɾawi**, 'write'.

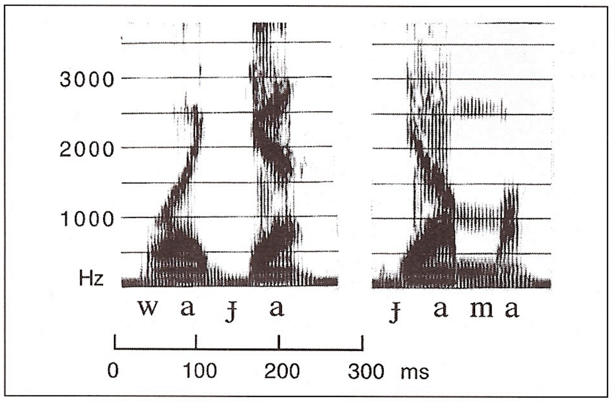

Fig. 9. Spectrograms of **baɟa**, 'palm fronds' and **ɟaka**, 'walk'.

forms can cause /i/ to vary between **i** and **ɟ**, and /u/ to vary between **u** and **w**. But none of our Banawá speakers had a glide **j**; for all of them the consonant series included a palatal stop **ɟ**, as exemplified in the spectrograms in figure 9.

Syllable and Word Structure

The most familiar building blocks of words and syllables are the consonant and vowel units denoted by C and V. The description of many languages, including Banawá, also requires a unit of timing, the mora, asssociated with each vowel, V. Syllables in Banawá can be described in terms of the following statements:

(1) Syllables have one or two moras.

Table 8. Banawá words containing sequences of vowels (all syllable boundaries are shown by a period [.])

.bii.	fan
.faa.	water
.kai.ɟa.ra.	to take pride in oneself
.ka.wa.rie.	to cook
.ka.wa.ri.sei.	rafter
.me.na.nau.	they play
.ra.bi.kai.	sick
.u.wa.ria.	one

Every syllable has at least one vowel, but there are no syllables with longer strings of vowels such as CVVV, which would have three moras.

(2) There are no final consonants.

(3) There are no consonant clusters.

If there is a consonantal onset to a syllable, it consists of only a single consonant.

As far as words are concerned, we cannot say how long they could be, but we can say:

(4) All Banawá words have at least two moras.

Words can have one or more syllables, but if there is only one syllable it must have two vowels.

Some typical Banawá words are given in table 8. Syllable boundaries within these words have been marked, showing how many syllables there are, and making it apparent that syllables can be CV or CVV, with VV denoting a vowel sequence of two moras within a single syllable. On most occasions, this sequence is very similar to the sequence of vowels that occurs in a diphthong in traditional descriptions. All VV sequences in Banawá form diphthongs, except for those in monosyllabic words, where like vowels can follow one another. We can therefore say that there are no vowel length contrasts in Banawá. The sequences of identical vowels that do occur in monosyllabic words do not contrast with short vowels in these words, as short vowels constitute only a single mora and a monosyllabic word, like all Banawá words, must have at least two moras, in accordance with (4).

Banawá syllables can contain nearly all the possible sequences of unlike vowels. Of the 12 possible sequences, **ie**, **ia**, **iu**, **ei**, **ea**, **eu**, **ai**, **ae**, **au**, **ui**, **ue**, **ua**, only two **ea** and **ae**, do not occur. One way of describing this situation is by (5):

(5) Except for monosyllabic words, adjacent vowels in the same word cannot have the same sonority.

This would still forbid long vowels, as like vowels obviously have the same sonority. If we also define sonority to make it clear that there is no difference in sonority between **e** and **a**, then (5) would constrain Banawá so that it would not have the sequences **ea** and **ae**. Our definition of sonority would also have to make it plain that **i** has more sonority than **u**, so that sequences such as **iu** and **ui** are allowed, and that both **i** and **u** have less sonority than **e** and **a**, allowing the other possible sequences of vowels.

There are, however, two objections to this proposal. Firstly, if we define sonority in terms of acoustic intensity, then the use of a ternary scale in which **(e=a)>i>u** requires special motivation. Why should there be a division between **i** and **u**, but not between **e** and **a**? Secondly, it provides little in the way of an explanation of the observed facts. Why should a difference in sonority (whatever that is) be required for a sequence of vowels? There is another possible explanation for the lack of sequences such as **ea** and **ae**, which appear to be generally disfavored among the world's languages. It may be that the two elements in each of these diphthongs do not differ sufficiently in quality (a much more salient property than sonority). A related possibility is that **ea** and **ae** are not sufficiently contrastive with **ia** and **ai**, respectively, and that vowel systems tend to have only the pairs that contrast most in vowel quality. We could rephrase (5) so that it would provide an alternative explanatory account of the facts:

(5′) Except for monosyllabic words, adjacent non-high vowels cannot occur in the same word.

Thus the sequences of [high] vowels **iu** and **ui** are permitted, but the sequences **ea** and **ae**, which contain the vowels that are most alike in quality are not permitted.

We must, however, note that this statement is not a fully satisfactory account of the facts, as it does not mention the systemic pressure to avoid having the similar pair **ea** and **ia**, and the other the similar pair **ae** and **ai**. We want to describe the phonetic structure of a language in a way that explains why it is natural for things to be as they are. When we are trying to describe the phonetic structure in terms of a simple set of statements, we have a problem. A single constraint may account for the facts, but there may be two or more reasons for the observed phenomena. There are nearly always systemic (paradigmatic) forces constraining the possibilities that can occur at a given point in the structure, and syntagmatic forces constraining possible sequences. Thus a language may avoid having **ae** and **ai** because they are too similar, and also fail to have sequences such as **ae** because the two vowels are too alike as indicated by (5′). In an ideal explanatory description we would have both types of statements, those giving the syntagmatic constraints ensuring that diphthongs consist of elements with sufficiently different qualities, and those giving the paradigmatic constraints ensuring that all the diphthongs that can occur at a given point in the system are sufficiently different from one another. However, general principles of simplicity do not encourage us to describe the situation in terms of two formal statement when one will do. Recent work by Flemming [1995] has discussed the necessity of considering paradigmatic pressures. But at the moment we do not have a way of combining statements about syntagmatic and paradigmatic pressures into a single theory applicable to the description of the phonetic structures of a language.

Suprasegmental Structure

Banawá has a very simple suprasegmental structure. Neither tone nor length play any role in distinguishing words, and stress is entirely predictable. Stress in Banawá occurs on the first mora in a word, and on every other mora after the first. The only exceptions to this rule are words with three or more moras beginning with a vowel. In such words the onsetless first vowel is not counted as a mora in the stress assignment process. A formal statement of the stress rule would embody two additional constraints:

(6) An initial vowel in a word cannot form a mora in words with three or more moras.

(7) Stress occurs on alternate moras within a word, beginning with the first mora.

Table 9. Words exemplifying Banawá stress placement (syllable boundaries are shown by a period [.])

(a)	**.'u.wi.**	cry
(b)	**.'bi.ta.**	mosquito
(c)	**.'wa.ra.'bu.**	ear
(d)	**.'wa.na.'ku.ri.**	spider
(e)	**.'re.re'u.ka.'na.**	crank
(f)	**.u.'wia.**	go out (as of a fire)
(g)	**.u.'wa.ri'a.**	one

Table 10. Further words exemplifying Banawá stress placement (syllable boundaries are shown by a period [.])

(a)	**.'kai.'ɹa.ra.**	to take pride in oneself
(b)	**.'reu.'ka.na.**	to stir
(c)	**.'su.ki'a.**	dark

These constraints are exemplified by the words in table 9. Stress falls on the first vowel of the two syllable words in (a) and (b). It falls on the first and third vowels in the three and four syllable words in (c) and (d), and on the first, third and fifth vowels in the five syllable word in (e). Note that stress falls on the first vowel of the two mora word in (a), but in the longer words in (f) and (g), the initial vowel is skipped, so that stress falls on the second vowel in (f) and on the second and fourth vowels in (g).

With the exception of (e), (f) and (g), each of the words in table 9 has only one mora in each syllable. We can see that stress is assigned by reference to moras rather than syllables by considering the additional words that have syllables with two moras given in table 10.

In (a) **'kai** must be a single syllable. Part of the evidence for this is that there cannot be two syllables, **'ka.i** because **i** cannot form a syllable by itself; all word-internal syllables have to have an initial consonant. In so far as it provides additional confirmation, we can also note that the **'kai** syllable in (a) sounds much like the English syllable [**kai**] as in 'kite'. The second syllable, **'ɹa**, is stressed, so stress cannot be said to be on alternate syllables; with alternate syllable stress the syllable **ra** rather than the syllable **ɹa** would have been stressed. All this shows that stress asssignment must be mora based rather than syllable based. The same arguments apply in (b). We must interpret **'reu** as a single syllable, as **u** cannot be a separate syllable without an onset. As the following syllable **ka** is stressed, stress must be on alternate moras not alternate syllables. Similarly, the two vowels at the end of (c) must belong to the same syllable. In this case the stress falls on the second mora in this last syllable of the word.

These observations on Banawá stress were verified by tests with three of our 6 Banawá speakers. Before we can discuss this work we must be fully explicit concerning the nature of stress. All aspects of speech are the results of muscular actions, and stress is no exception. In our view stressed syllables are produced by increased activity of the respiratory muscles, typically the internal intercostal muscles [Ladefoged, 1967]. One of the few arguments against this notion is that of Adams [1979], who makes it quite clear that she disagrees with Ladefoged [1967]. Adams [1979, p. 117] recorded EMG activity from the internal intercostals and says 'It should be emphasized … that in no case … were localized burst of internal intercostal activity found to correlate with the incidence of stressed syllables.' It is certainly true that one cannot see bursts of internal intercostal activity occurring on stressed syllables in Adams' published data. But this does not mean that such activity did not occur in the utterances she recorded. Her data reflect the activity of a large number of motor units firing almost simultaneously, so that the action potentials are superimposed on one another in the record. In the studies summarized by Ladefoged [1967] fewer units were recorded, so that the separate bursts of activity are more apparent. One cannot see bursts of internal intercostal activity in Adams's records because of the density of the activity and the time scale of the recordings. But that does not mean that they are not there. One cannot see the formant structure in the waveforms that accompanied her EMG data; but no one would doubt that the formants are there and are varying in accordance with the sounds being produced. It just takes some signal processing of the waveform to make the variations evident. In addition there are many other studies supporting the notion that stressed syllables involve extra respiratory effort (though not necessarily of the internal intercostal muscles). Stressed syllables have a higher subglottal pressure [Ladefoged, 1967]. There is no way to explain peaks of subglottal pressure in different places in noun-verb pairs (recorded within frames) such as 'an overflow' and 'to overflow' without postulating controlled variations in respiratory activity that can be correlated with the positions of the stressed syllables.

If we consider stress to be a motor gesture, then one way of determining where stressed syllables occur is to get speakers to tap on something while they speak. This activity is a very basic gesture that can be easily coordinated with other movements. Speakers can often beat time on the stresses, or nod their heads, or make some other gesture such as tapping on a table. It is difficult to get speakers to tap only on the unstressed syllables, or on every occurrence of some other phonetic feature, such as whenever the tongue touches the alveolar ridge, or every time the lips close. But speakers usually have no difficulty in tapping on stressed syllables. At least in a language that has lexical stress, as long as a word is said with its normal rhythm, beats will occur on stressed, and only on stressed, syllables.

We instructed our main Banawá speakers to tap on stressed syllables in Portuguese, a language with which they were familiar, demonstrating on Portuguese examples what was required, and getting them to practice in that language first. We then asked them to say a number of Banawá phrases, each time saying the phrase twice in a natural way and repeating it a third time while tapping on appropriate syllables. Occasionally, when asked to tap in this way, the speakers responded by trying to tap on every syllable; but when they did this they altered the rhythm of the sentence they were producing, saying it syllable by syllable. We would then ask them to say it at a regular speed, and to tap on just the principal parts. They responded by producing the required sentence, and repeating it with taps on the stressed syllables with no changes in rhythm. We used our two main speakers to explain the task to a third speaker who was not so familiar with Portuguese. Unfortunately circumstances did not permit us trying this task with other speakers.

Figure 10 exemplifies this process. The upper part of the figure is the waveform of the phrase **'tu.wi'a**, 'to allow not to accompany'. Below is another token of this phrase, said while tapping on a glass which produced a sharp ring (and overloaded the recording). The lower part of the figure is the phrase **'tu.wi.'ɹa**, 'poorly done'. This part of the figure also shows one token without taps and below it one token with taps. In each case, the token accompanied by taps was said slightly more forcibly, and

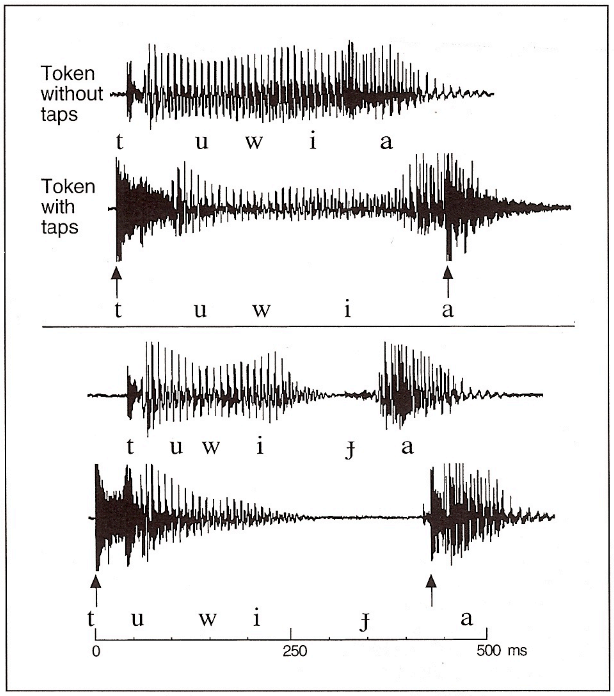

Fig. 10. Two tokens of each of two Banawá phrases **'tu.wi'a**, 'to allow not to accompany' and **'tu.wi.'ɟa**, 'poorly done'. In each case the second token is accompanied by taps on a glass near the microphone. The arrows mark the onsets of these taps.

therefore slightly more slowly. In the first phrase the second tap ocurs on the last vowel in this two syllable word, which is the third mora in the word. In the second phrase the tap occurs as the palatal stop is released, making this phrase have three syllables, with stresses on the first and last. This same pattern was evident in all three of the speakers who were recorded while producing taps.

In all the Banawá phrases we tested in this way, speakers tapped on the mora predicted by the account of the location of stresses given above. The phrases used in this part of the study were not controlled for segmental content in ways that would have enabled us to use acoustic measures such as duration and intensity as correlates of stress. We hope that a later study of additional material will be able to investigate this topic. On the basis of our present data the constraint based description of stress placement seems to provide a good indication of speakers' behavior.

Conclusion

This is a far from complete account of the phonetic characteristics of Banawá. We have described the qualities of the Banawá vowels, noting the slight allophonic differences between vowels in a coronal and labial context; but many other small allophonic differences could have been noted. We have described the places of articulation of each of the consonants, and given an account of the variations in VOT. The manner of articulation has also been documented for most of the consonants. But little has been said about allophonic variations. The structure of syllables and words has been delimited, and the location of stressed syllables verified; but there has been no analysis of intonational features. Much more work remains to be done. But we have been able to present the major phonetic characteristics of this endangered language.

Acknowledgements

Many thanks to our ever helpful Banawá consultants. We are also grateful to friends at the SIL Center at Porto Velho, Brazil, for their hospitality and help. Helpful comments on earlier versions of this paper were made by Robert Kirschner, Donca Steriade, Richard Wright and two anonymous reviewers. This research was supported by NSF Grant 9511118 to P.L. and I.M. and NSF Grant 9631322 to D.E.

References

Adams, C.: English speech rhythm and the foreign learner (Mouton, The Hague 1979).
Buller, E.; Buller B.; Everett, D.: Stress placement, syllable structure, and minimality in Banawá. Int. J. Amer. Ling. *59:* 280–293 (1993).
Everett, D.: Prosodic levels and constraints in Banawá and Surwaha (Rutgers Optimality Archive, Rutgers 1996).
Everett, D.: Sistemas prosodicas da familia Arawa; in Estudos Fonologicos de Linguas Indigenas Brasileiras (Editora da Universidade Federal do Rio de Janeiro, Rio de Janeiro 1995).
Fischer-Jørgensen, E.: Acoustic analysis of stop consonants. Misc. Phon. *2:* 42–59 (1954).
Flemming, E.: Auditory features in phonology; UCLA PhD Diss. (1995).
Grimes, B.: Ethnologue (Summer Institute of Linguistics, Dallas 1992).
Johnson, K.; Ladefoged, P.; Lindau, M.: Individual differences in vowel production. J. Acoust. Soc. Amer. *94:* 701–714 (1993).
Ladefoged, P.: Phonetic prerequisites for a distinctive feature theory; in Papers in Linguistics and Phonetics in Memory of Pierre Delattre (Mouton, The Hague 1972).
Ladefoged, P.: Three areas of experimental phonetics (Oxford University Press, Oxford 1967).
Lindblom, B.: Models of phonetic variation and selection, vol. 11, pp. 65–100; PERILUS (Phonetic Experimental Research, Institute of Linguistics, University of Stockholm (1990).
Lisker, L.; Abramson A.: Cross-language study of voicing in initial stops: Acoustical measurements. Word *20:* 384–422 (1964).
Peterson, G.; Barney, H.: Control methods used in a study of the vowels. J. Acoust. Soc. Amer. *24:* 175–184 (1952).
Stevens, K.N.: On the quantal nature of speech. J. Phonet. *17:* 3–46 (1989).