APPENDIX 1

Northeast Coastal Bantu

Defined by following sound changes:

i. */β/ > v (e.g. before į)

ii. */β/ > w (e.g. before a)

iii. */lį/ > Zi (before vowels)

I. Sabaki-Ruvu

Languages sharing loss of Cl. 5 prefix before C-stems

A. Sabaki-Seuta

Languages sharing the change */NC̥/ > (N̥)C̥^{h}

1. Sabaki (Kenyan Coastal)

Languages sharing change */lį/ > dzi/ji/ɟi

a. Pokomo-Mijikenda

Languages sharing */t/ > h shift

i. Pokomo

ii. Mijikenda

Languages share */kw, gw/ > k͡p, g͡b

b. Swahili

2. Seuta

Languages sharing */lį/ > zi

a. Zigula-Ngulu

b. Shambala-Bondei

B. Ruvu

Languages sharing change */NC̥/ > N^{h}

1. Doe
Nhwele (Kwere)
Zalamo
Kutu-Kami
Lugulu

2. Sagala

3. Gogo-Kagulu

II. Saghala

Studies in African Linguistics
Supplement 6, November 1976

PERCEPTION OF TONES OF BISYLLABIC NOUNS IN YORUBA*

Jean-Marie Hombert
University of California, Los Angeles

Although an increasing number of acoustic studies provide accurate data concerning the fundamental frequency (Fo) shapes (i.e. the Fo variations as a function of time for a given tone) of various tone systems[1], perceptual studies of tone languages are almost non-existant.[2]

The purpose of this study is to investigate the perceptual cues used by Yoruba speakers to distinguish among the six tone patterns found in bisyllabic nouns.

0. The Yoruba tone system

Yoruba has been analyzed as having three level lexically contrastive tones: High (H), Mid (M) and Low (L) (e.g. Ward [1952]). Phonetic investigations show however that the low tone is in fact a low falling tone at least in word final position [Hombert 1976a; La Velle 1974]. Two main sandhi rules are found: a) a high tone is changed into a rising tone after a low tone, and b) a low tone is changed into a falling tone after a high tone.[3] Most Yoruba nouns are bisyllabic and have a V_1CV_2 shape. V_1 is either Low or Mid; V_2 can have any of the three Yoruba tones. Thus, the following six patterns are found: LL [_ ↘], LM [_ –], LH, which is realized as a Low-Rising sequence according to the first Sandhi rule just mentioned, [_ ↗], ML [– ↘], MM [– –] and MH [– ¯].

*I wish to thank the members of the Tone Seminar at the University of California, Los Angeles, for their comments and criticisms on an earlier version of this paper. This research was supported by a NSF grant made to the, Phonetics Laboratory, University of California, Los Angeles.

[1]See recapitulation in Hombert [1976b].

[2]See recapitulation in Gandour [forthcoming].

[3]Rule b) is not going to apply to the data presented in this paper since VCV nouns cannot have a high tone on the first vowel.

1. Perceptual factors[4] and the Yoruba tone space

1.1. Experimental paradigm. The following experiment was carried out in order to determine the perceptual factors used by Yoruba speakers to distinguish bisyllabic nouns differing essentially[5] in tone patterns. Eight Yoruba subjects were asked to evaluate the perceptual distance (on a scale from 0 to 10) between tone patterns of VCV nouns presented by pairs. They were instructed to circle "0" if the two tone patterns sounded identical and to move away from the "0" according to the magnitude of the perceived difference between two given tone patterns, the value of "10" representing extreme differences between two stimuli. The complete set of stimuli was played to the subjects in order to enable them to establish a scale for their judgments. As shown in Table 1, 15 different comparisons can be made with the six tone patterns.

Table 1. Possible tonal contrasts found in VCV nouns

	LM	LH	ML	MM	MH
LL	1	2	3	4	5
LM		6	7	8	9
LH			10	11	12
ML				13	14
MM					15

The words used as stimuli in this experiment had been previously recorded by a Yoruba speaker in a sound-treated room. They were read in the carrier frame[6] Kíni X ní Yorùbá? 'What is X in Yoruba?'. They were sub-

[4]The terms "cues", "factors", and "parameters" will be used interchangeably in this paper.

[5]Since we used natural stimuli, Fo differences were not the only differences present in our paired stimuli. Secondary cues such as duration and amplitude differences were not removed from our stimuli. The second part of the paper will deal with this problem.

[6]The carrier phrase was used in order to get a more natural intonation during the recording. Only the test words (i.e. VCV nouns without the carrier phrase) were used during the perceptual experiment.

sequently extracted from the frame and analyzed (Fo, duration, intensity) by a PDP 12 computer. In order to minimize semantic influences, two sets of minimal pairs were presented for each tone contrast. Each pair was presented five times in a randomized fashion making a total of 300 judgments per subject (15 tonal contrasts X 2 sets of minimal pairs X 2 (same pairs presented in reversed order) X 5 repetitions). There was a 1 sec silence between the first and the second noun forming one pair and 3.5 sec between the second noun and the first member of the next pair. The stimuli were presented to the subjects through earphones at a comfortable level (about 70 dB). The 30 minimal pairs used in this experiment are presented in Table 2.

Table 2. Minimal pairs illustrating the 15 tonal contrasts[8]

1.	LL/LM	ìlù	'drum'	ọ̀gọ̀	'stupid person'
		ìlu	'awl'	ọ̀gọ	'club, stick'
2.	LL/LH	ọ̀wọ̀	'respect'	ìlù	'drum'
		ọ̀wọ́	'group'	ìlú	'city'
3.	LL/ML	ọ̀wọ̀	'respect'	ọ̀kọ̀	'spear'
		ọwọ̀	'broom'	ọkọ̀	'vehicle'
4.	LL/MM	ọ̀kọ̀	'spear'	òkò	'pebble'
		ọkọ	'husband'	oko	'form'
5.	LL/MH	ọ̀wọ̀	'respect'	ọ̀kọ̀	'spear'
		ọwọ́	'hand'	ọkọ́	'hoe'
6.	LM/LH	ìlu	'awl'	òro	'wild mango'
		ìlú	'city'	òró	place name
7.	LM/ML	òro	'wild mango'	òdo	'zero'
		orò	a god	odò	'river'
8.	LM/MM	òro	'wild mango'	àwo	'plate'
		oro	a kind of tree	awo	'secret'

[7]The Fo extraction was done by CEPSTRUM method. The software used for this analysis was written by Lloyd Rice and Peter Ladefoged.

[8]The following conventions are used to indicate tones: V́ for high tone; V̀ for low tone; mid tone is unmarked.

Table 2 (cont.)

9.	LM/MH	òro	'wild mango'	èwe	'youth'
		oró	'poison'	ewé	'leaf'
10.	LH/ML	ọ̀wọ́	'group'	òró	place name
		ọwọ̀	'broom'	orò	a god
11.	LH/MM	òró	place name	ìwọ́	'umbilical cord'
		oro	a kind of tree	iwọ	'poison'
12.	LH/MH	ọ̀wọ́	'group'	òró	place name
		ọwọ́	'hand'	oró	'poison'
13.	ML/MM	orò	a god	ọkọ̀	'vehicle'
		oro	a kind of tree	ọkọ	'husband'
14.	ML/MH	ọwọ̀	'broom'	ọkọ̀	'vehicle'
		ọwọ́	'hand'	ọkọ́	'hoe'
15.	MM/MH	oro	a kind of tree	ọkọ	'husband'
		oró	'poison'	ọkọ́	'hoe'

Figure 1 shows the average shape of the Fo of the six tone patterns. All the words with identical tone patterns were averaged together[9] (i.e. a total of 10 words). In order to be able to average the Fo of vowels of different length, the duration of each vowel was normalized by dividing the overall vowel duration in three intervals. An averaged Fo value was computed for each interval and then similar tone patterns were averaged together.

1.2. Results and discussion. One subject gave very inconsistent responses and was discarded. The data from seven subjects were analyzed by factor analysis techniques. The purpose of these statistical techniques is to seek to discover the structure of factors which underlies a set of data or, in other words, to extract the relevant factors which are associated with subjects' judgments of perceptual distance. For example, in the case of this study, one is interested in extracting the factors used by Yoruba subjects to evaluate perceptual distance. Most factor analyses do not have a unique solution (i.e. a unique set of factors); they provide the experimenter

[9]This means that sometimes identical words were averaged together when they were used more than once to contrast with other tone patterns (see Table 2).

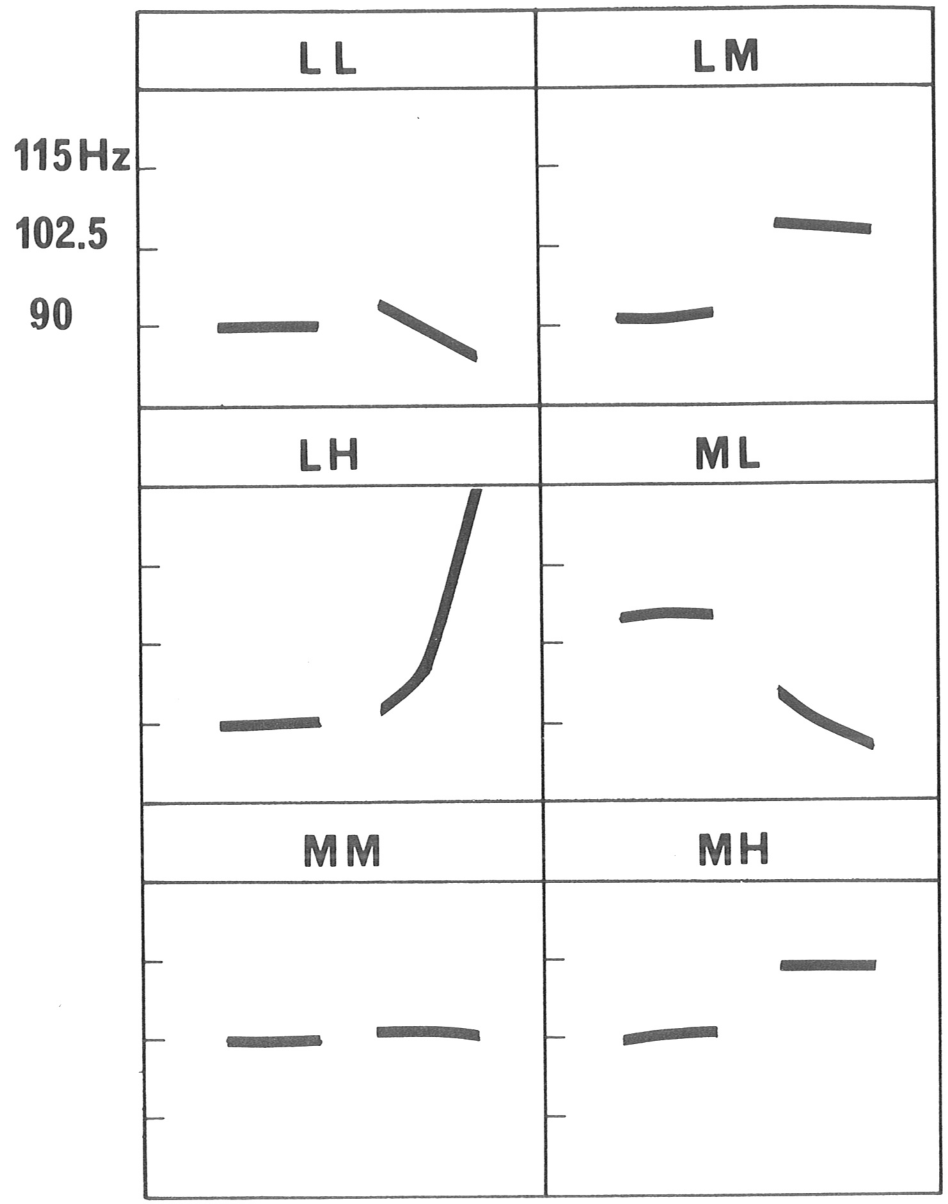


Figure 1. Averaged shape of Fo for the six tone patterns found in VCV nouns (one speaker, ten tokens)

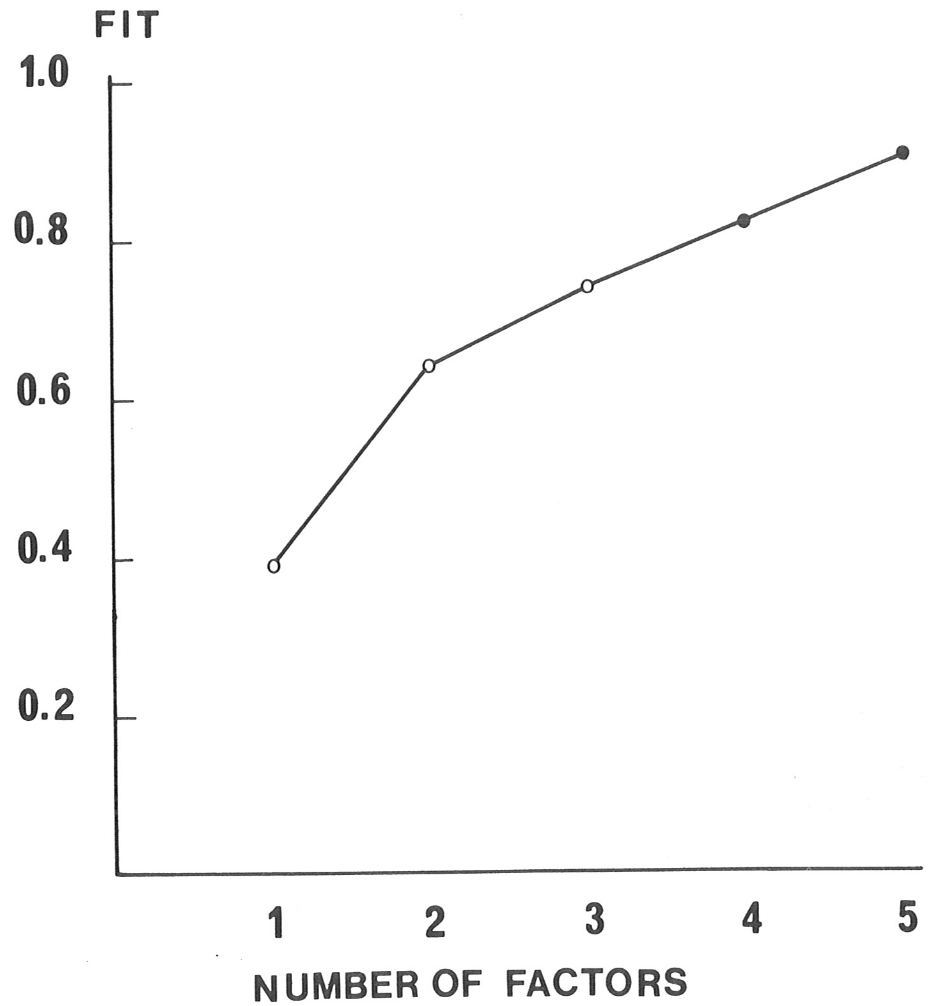
with a set of solutions which are mathematically equivalent and it is the experimenter's role to choose one of these solutions either by using his knowledge of what the solution should look like or by designing subsequent experiments which will empirically justify his choice.

The PARAFAC technique [Harshman 1970] used in this study is an extension of the standard model of factor analysis. One of the advantages of PARAFAC is that it gives a unique solution provided the data are adequate.[10] Because of this uniqueness the solution is claimed to be more explanatory than the solutions provided by more traditional techniques of factor analysis. The greater the number of factors extracted by PARAFAC, the more the variations in the data can be accounted for, but one has to decide whether the increase in the predictability of the data is significant enough to motivate the corresponding increase in the complexity of the solution. The number of factors to be extracted is systematically increased by the experimenter. A point will be reached at which the solution provided by PARAFAC will cease to be unique. One way of choosing among the analyses which have a unique solution is to plot the percentage of the variance accounted for by a given solution as a function of the number of factors. If a sharp change in the slope of the curve is found, it suggests that the amount of improvement of prediction of the data obtained by adding another factor is no longer significant. As can be seen on Figure 2, this criterion favors the two-factor solution. Although the three-factor solution was also a unique solution it was rejected because the amount of improvement of prediction is not much different between 2 and 3 factors from what it is between 3 and 4 or 4 and 5 (4 and 5 factor solutions are non-unique).

The output of a PARAFAC analysis is a set of factors and for each factor, the set of stimuli judged by the subjects is hierarchized. The two-factor solution is presented under this format in Figure 3.

Obviously no label is attached a priori to the factors extracted by PARAFAC. It is the experimenter's role first to interpret and give a label to these factors and second to justify these labels by subsequent statistical analyses.

[10]For the determination of the adequacy of the data set see Harsman [1970].



(FIT = correlation squared; ○ = unique solution; • = non-unique solution)

Figure 2. Goodness of fit as a function of the number of factors extended

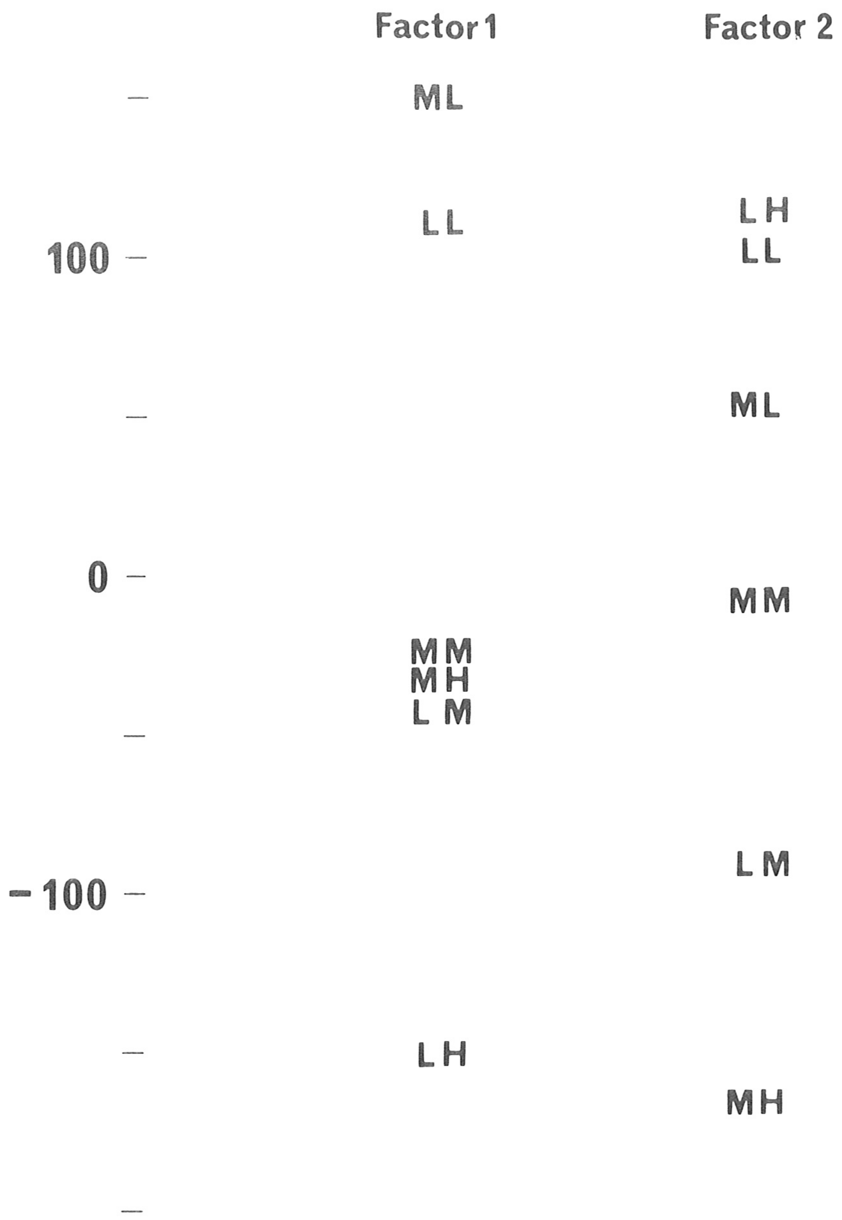


Figure 3. Factor loadings of the two-factor solution

By looking at Figure 3 we can see that Factor 1 separates the six stimuli into three groups (Group 1: ML and LL; Group 2: MM, MH, and LM; and Group 3: LH). This factor seems to be related to the direction of change of the tone of V_2 (i.e. falling, level, and rising). Another possible interpretation of Factor 1 would be to consider the end point of V_2 but this would lead to the wrong prediction that the MH sequence should be grouped with the LH sequence rather than with MM and LM sequences. The interpretation of Factor 2 is more complex. Since the three stimuli which have a changing Fo in the second vowel are ordered before the three stimuli which have a steady state Fo associated with the second vowel, it is likely that the contour/level distinction is playing an important role. But the fact that MM, LM, and MH are separated by large intervals indicates that cues other than the contour/level distinction are included in this factor. Two secondary cues can be considered: 1) The difference between the offset of V_1 and the onset of V_2. This cue would explain why ML is differentiated from LL, and MM from LM and MH. 2) The end point or the averaged Fo of V_2. This cue would explain the distinction made between LM and MH. The fact that this cue would predict that the LH sequence should be at the bottom of the scale can be accounted for by considering that the contour/level cue has a greater perceptual weight than the end point parameter.

In order to clarify and justify the labels attached to the three factors, regression analysis techniques were used. The purpose of this type of analysis is to see to what extent the factor loadings per stimulus (i.e. the values attributed to each stimulus for each factor by the PARAFAC analysis) can be predicted from the actual acoustic values of the stimuli. For instance, if it is claimed that Factor 1 represents the direction of change of the Fo of the second vowel it is important to know how much of the hierarchy provided by Factor 1 can be predicted by using only the actual Fo values of the second vowel. Moreover, regression analysis will help to make precise the content of a given factor; for instance in the case mentioned above, regression analysis will tell us whether the direction of change of the Fo of V_2 or the offset value of the Fo of V_2 (or any linear combination of these two values) is a better approximation of Factor 1. Regression analysis shows that the direction and the amount of change of Fo during V_2 account for 81% of the hierarchy predicted by Factor 1. This figure goes up to 92% when the end point of the Fo of V_2 is also taken into consideration.

In the case of Factor 2, 63% of the hierarchy is predicted when the amount of Fo change at V_2, i.e. contour/level distinction irrespective of the direction of change, is considered alone; 90% when both the amount of Fo change and the averaged Fo of V_2 are taken into account, and finally 99% when a third one, that is the Fo difference between V_2 onset and V_1 offset, is added. These experimental results help quantify the respective role of various acoustic cues, i.e. direction of Fo change, amount of Fo change, Fo value at the end point of a vowel, Fo differences between two syllables, familiar to linguists working with tone languages. Although it is very speculative at this point, one could be tempted to relate these factors to neural activity in the auditory system, e.g. certain neural units responding only to changing frequency stimuli but not to steady state stimuli.

In summary, this experiment suggests that Yoruba speakers used two main factors when judging the distance between the six tone patterns found in bisyllabic nouns. The first factor is mainly associated with direction of change of Fo of V_2 and also to some extent to the averaged Fo value of V_2. The second factor can be analyzed as resulting from the combination of three cues: amount of Fo change of V_2, averaged Fo value of V_2, and Fo difference between onset of V_2 and offset of V_1.

2. The role of secondary cues

The second part of this study investigates the relative importance of primary and secondary cues used to distinguish a mid tone from a low tone in word final position in Yoruba. By comparing LL and ML sequences to LM and MM sequences respectively it was found not only that the final low tone in the first two sequences starts at a lower Fo than a mid tone in the corresponding LM and MM sequences but also that the low tone has a falling Fo (vs. level Fo in the case of a mid tone). The low tone is also shorter in duration (by about 30 msec) and lower in amplitude (by about 6 dB). The purpose of this experiment was to determine whether words like ọkọ̀ 'vehicle' and ọ̀gọ̀ 'stupid person' would be perceived as ọkọ 'husband' and

ọ̀gọ 'stick' by Yoruba speakers after being modified[11] in terms of amplitude or duration or magnitude of Fo contour.

2.1. Experimental paradigm

Seven Yoruba subjects participated in this experiment. They were asked to judge the meaning of these stimuli, which were presented to them through a loudspeaker at a comfortable level (about 70 dB). In the first part of the experiment they could circle on their answer sheet either ọkọ̀ 'vehicle' or ọkọ 'husband' and in the second part they had a choice between ọ̀gọ̀ 'stupid person' and ọ̀gọ 'stick'. Ten stimuli were used:

1) The original LL and ML words, i.e. ọ̀gọ̀ and ọkọ̀ .

2) Low tones with increased amplitude: The amplitude of the second vowel of the LL and ML sequences was raised by 7 dB.

3) Low tones with lengthened duration: The duration of the second vowel of the LL and ML sequences was lengthened by 4 glottal pulses (about 50 msec). In order to avoid abrupt changes in the slope of this low tone, every other glottal pulse was repeated once (instead of repeating the same glottal pulse four times).

4) Low tones with steady state fundamental frequency: The fundamental frequency of the second vowel of the LL and ML sequences was changed into a level Fo by repeating the third glottal pulse (and keeping the same duration as the original low tone).

5) Low tones with steady state fundamental frequency, lengthened duration, and increased amplitude: The three cues were combined; the third glottal pulse of the second vowel was repeated, the duration was increased by 4 glottal pulses and the amplitude was raised by 7 dB.

Each of these ten stimuli was presented ten times in a randomized order. There was a 3.5 sec silence between each stimulus in which subjects could give their response.

---

[11]These modifications were made on the digitized wave form (sampling rate of 10 kHz) using programs written by Louis Goldstein, Steven Greenberg, Jean-Marie Hombert and Lloyd Rice.

2.2. Results and discussion

The ten stimuli used in this experiment were either original LL and ML sequences or modified versions of these original sequences. However, Yoruba speakers perceived some of these stimuli as LM and MM sequences, respectively. This shift in judgments did not occur with stimuli of Type 2 (increased amplitude) and 3 (increased duration) but occurred with Types 4 (steady state Fo) and 5 (three types of modification combined). These results are presented in Table 3.

Table 3. Number of times (out of Fo judgments) a low tone was identified as a mid tone (7 subjects)

	Stimulus 4 (steady state Fo)	Stimulus 5 (steady state Fo, increased duration, increased amplitude)
LL identified as LM	55	57
ML identified as MM	32	37

These results show that to a great extent a low tone in word final position will be identified as such only if it has a falling Fo. If a tone has a steady state Fo it will be identified as a mid tone even if its Fo value is within the range of values associated with low tone[12]. It is interesting to notice that this misidentification will not be as frequent when the preceding tone is a mid tone as opposed to a low tone (see Table 3). A preliminary explanation is that the interval between low and mid (in a LM sequence) is smaller than the interval between mid and low (in a ML sequence). There is a tendency for the number of "mistakes" to increase when the three modifications are combined (stimulus 5). But this improvement was found to be non significant. Shifts in judgment did not occur when only the duration or only the amplitude was increased. It could be the case that since the functional load of the tone system is so important, secondary

[12]It should be pointed out that by replacing the original falling Fo by a steady state Fo, I removed at the same time possible additional cues such as slightly different phonation types (e.g. creakiness) between the beginning and the end of the vowel.

cues are overridden by Fo. It would be interesting to test the role of amplitude in the perception of tones in languages in which the functional load of the tone system is much lower (e.g. Bantu languages).

In summary, the results of this second experiment show that falling Fo is the main perceptual cue to identify low tones in word final position. Although production data indicate a strong correlation between Fo and amplitude, amplitude information does not appear to be a major cue used by Yoruba speakers to distinguish between low and mid tones in word final position.

REFERENCES

Gandour, Jack. Forthcoming. "Perception of tones." In Victoria A. Fromkin (ed.), Tone: A Linguistic Survey. New York: Academic Press.

Harshman, Richard A. 1970. "Foundations of the PARAFAC procedure: models and conditions for an 'explanatory' multi-modal factor analysis." Working Papers in Phonetics, no. 16. University of California, Los Angeles.

Hombert, Jean-Marie. 1975. "Towards a theory of tonogenesis: an empirical, physiologically, and perceptually-based account of the development of tonal contrasts in languages." Doctoral dissertation, University of California, Berkeley. Forthcoming revision, Report of the Phonology Laboratory, no. 2, University of California, Berkeley.

Hombert, Jean-Marie. 1976a. "Consonant types, vowel height, and tone in Yoruba." In Working Papers in Phonetics, no. 33, pp. 40-54. University of California, Los Angeles.

Hombert, Jean-Marie. 1976b. "Tone space and universals of tone systems." Paper presented, annual meeting of the Linguistic Society of America, Philadelphia.

La Velle, Carl R. 1974. "An experimental study of Yoruba tone." In William R. Leben (ed.), Papers from the Fifth Annual Conference on African Linguistics, pp. 185-194. Studies in African Linguistics, supp. 5.

Ward, Ida C. 1952. Introduction to the Yoruba Language. Cambridge: Heffer and Sons.