Ann. Bull. RILP, No. 9, 137-150 (1975)

PROGRESS REPORT ON AN ACOUSTICAL STUDY OF PITCH IN TIBETAN

Olle Kjellin

1. Introduction.

According to Kitamura (1974) there are four stressed tones and one neutral tone in Lhasa Tibetan. The four stressed tones of Kitamura's phonemic classification (based on auditory impressions) are:

(1)
- (I) high level: ¯ma 'wound', ¯maa 'to the wound';
- (II) low to mid rising: ´ma 'mother', ´maa 'butter';
- (III) high to mid falling: \`maa 'battle';
- (IV) low-mid-low (convex): ^maa 'down'.

Kjellin (1974), the first attempt at a generative-phonological treatment of Tibetan tones, was based on this analysis with some modifications in order to make the treatment more economical:

(2)
- (I) same
- (II) only low in short syllables: _ma 'mother' (under influence of Woo 1969), but low to high rising in long syllables: ´maa 'butter';
- (III) high to low falling: \`maa 'battle';
- (IV) only low-low in the underlying representation, which was found to need this amount of tailoring to fit the sandhi rules that work in compounds. In the output, however, the isolation form was the same as in Kitamura's analysis.

For the first three tones, the surface isolation forms could also be used as underlying form. The so-called neutral tone was ignored in Kjellin (1974), and it has not yet been instrumentally investigated, either.

2. Aim and Procedures

2.1. The aim of this study was only to provide a more reliable basis (than the auditory "transcriptional" phonetic analysis) for the phonological analysis and for the experiments to follow. This paper reports on the results of the analysis of a small part of the recorded material. The data obtained so far show consistency and seem to provide the relevant basis aimed at.

2.2. The subject was Sonam Gyatso ("Sonam Rinboche"), aged 41. He spent his first seven years in Lhasa, but then he was picked out as the incarnation of the chief monk of a major monastery (Ngor Gonba) of the Sakya sect. So from the age of seven he lived and studied at the monastery, near Shigatse in the Tsang province to the West of Lhasa, until he was about 25 years old, when he was forced to flee to India. Within a short time he went on to Japan, where he has been living for more than twelve years working as a researcher on the Tibetan language, culture, and religion, and as an informant for e. g. H. Kitamura's investigations of the Tibetan language. This latter fact is the main reason why I based my work on Kitamura's tonal analysis (with modifications) rather than that of e. g. Richter(1964) or that of Sedlâcek (1959). (Richter's and Sedlâcek's informant was only available to me for a short time.)

Sonam himself considers his idiolect representative of the Lhasa dialect. An important question is whether Sonam's Tibetan has become unnatural in any respect during his long exile: Fortunately it does not appear to have done so, according to other Tibetans and to Kitamura, and this is probably guaranteed by the fact that he has never lost contact with his language, due to the nature of his work.

2.3. High-quality recordings were made in a floating anechoic chamber through a Sony ECM-54 electret condenser microphone on a TEAC R-740 tape recorder.

Pitch extraction was performed on a FACOM 230-25 computer, but for the averages and curves the smaller DEC PDP-9 was used.

2.4. The six monosyllables mentioned in §1 (¯ma, ´ma, ¯maa, ´maa, `maa, ^maa) as well as the twelve actually possible (out of the 36 theoretically possible) disyllabic combinations of these tones, plus again the six monosyllabic types but with other initials (s, w, y, th, 't/d, and h/ɦ), were each recorded twenty times, ten in isolation alternating with ten in the carrier sentence, "´thalen ___ redan?" 'I suppose the next word is ___'. Thus, for example, "`maa; ´thalen `maa redan; `maa; ´thalen `maa redan; `maa ... etc.", with a breath pause between each.

Out of these more than one thousand utterances only sixty have been completely analysed at present, viz. the 6 ma(a)'s x 10 times without carrier sentence.

The analysis was done by computer with the pitch extraction program of Fujisaki and Tanabe (1972). This pitch extractor is roughly speaking a peak detector by which the fundamental period (T_0) of the glottal wave is measured and then the fundamental frequency (F_0), which is heard as pitch, is calculated.

The technique is based on approximate deconvolution of the source and transfer characteristics of speech so that the derived time-domain function is free from subsidiary peaks due to formants, but still, in contrast to cepstrum techniques (Noll 1964, 1966) retains the advantages of the autocorrelation function. However, as this technique, containing for example no less than three Fourier transforms, is rather time-consuming, it is combined with a method by which the peaks of a reasonably steady waveform can be detected directly from the waveform. Only when there is an abrupt change in pitch, as e. g. at the onset of the voiced portion, need the pitch extractor resort to the more complex method. Thus the process is relatively fast (Fujisaki and Mitsui 1973).

The speech signal is sampled at 10 kHz for the digital processing, and after computing, the F_0 value etc. for every 0.0128 sec (12.8 msec) is interpolated and printed out. These print-outs were compared with oscillograms, narrow-band spectrograms, and wide-band spectrograms in order to determine the exact moment of oral release. This moment was used for alignment so that all values of the same syllable type could be plotted together and also averaged, on the assumption that the moment of oral release has some phonological-organisational function in natural speech.

3. Results and Discussion.

The results are presented in extenso in figures 1 to 7. The "a" figures show all the tokens of each type plotted together. The "b" figures show the arithmetic averages of each type. Figure 7 shows all of these averages plotted together.

As the "a" figures show, each pitch contour type is produced with considerable consistency, not only of the contour itself but also of the absolute pitch. The small undulations are due to irregularities in vocal fold vibration. These irregularities naturally are larger in the vicinity of voice onset and cessation, and they can (probably) safely be ignored; they have no noticeable effect on linguistic entities.

The few tokens that included errors made by the computer have been rejected and are not plotted in any figure, but "errors" made by the subject's vocal folds are plotted. Such errors and non-errors were detected in the manual analysis and comparison with the oscillograms. In those cases where a few tokens are longer than the others, the averages of only three tokens or less have not been plotted.

It must further be noted that contours with a falling pitch frequently end with a glottal-stop-like laryngeal gesture, seen as an aperiodic waveform which the pitch extractor judges as "voiceless"; this is especially noticeable in Fig. 6.

I have no explanation for the phenomenon observable in the closure period of the initial consonant (sonorant): a rapid decrease of the F_0 value beginning about 40 or 50 Hz higher than the target, which is reached about 35 ms before the oral release. This fall may be an irrelevant artefact arising utterance-initially.

On the other hand it is worth noting that there does seem to exist a pitch target related to each phonological segment, and this target may be relevant for the tone as such, provided that, in the case of the initial segment, it is appropriate to interpret the end point of the F_0-fall as a phonological target. In ´maa (fig. 4) the low target of the m is 'stretched out' across most of the immediately following vowel, and the last vowel contains a high target; ¯ma of Fig. 1 contains only high targets, etc. There do not seem to be any intermediate targets (mid tones). Consequently, our modifications of Kitamura's tones as given in (2) were partly correct but still need further amendment, as in (3) (hereafter in this paper we shall use upper-case letters for segments with high pitch targets, and lower-case letters for segments with low pitch targets).

(3)

(I)	high level:	MA	'wound',	MAA	'to the wound';
(II)	low-high:	mA	'mother',	maA	'butter';
(III)	high-low:	MAa	'battle';		
(IV)	low-high-low:	mAa	'down'.		

As is evident from the graphs, e. g. Fig. 7, high pitch target is generally between 130 and 140 Hz, slightly downdrifting in the level tones. Low pitch target is between 110 and 120 Hz. In one case, the A of MAa, the pitch jumps to 150 Hz. This may be viewed as a means of increasing the 'distance' to MAA, which is de facto slightly falling although linguistically level.

That the A of mAa is high, not mid or low, is slightly surprising, as is also the fact that mA is not ma, that is, it is not just low, it does rise to high. Particularly the latter example is interesting in the light of Woo's proposal (1969) that contour tones (i. e. non-level tones) do not occur in short syllables but result where there is a sequence of sonorant segments with different pitch targes excluding the intial consonant (Woo 1969, e. g. p. 122).

In the next few paragraphs I shall take issue with Woo on this point.

Woo (1969) does admit that contour tones "may be perceived" on short vowels, but only "as the result of stress and intonation mechanisms interacting with the pitch production mechanism, or as a result of the transition from a particular type of consonant to the vowel, or from the vowel to a particular type of consonant" (p. 126). However, the Tibetan m and M do not appear to be two particular types of consonant.* About sonorants, Woo says, "Convention (II) states that initial non-syllabic sonorants, i. e. nasals and glides, cannot be contrastive with respect to pitch" (p. 84), and she concludes this "convention" by specifying initial sonorants as mid. It is not clear whether her conventions are intended only for Mandarin, but if she intended them as universal (as true conventions are), the Tibetan data seem to provide counterevidence. The other sonorants in Tibetan also contrast in this way: WA-wA, YA-yA, RA-rA, LA-lA. Further, the non-sonorants SA-sA, SYA-syA (palato-alveolar) show the same contrast, but here it is very difficult to perceive the difference, since these are voiceless, and the whole 'onus' rests with the short A, which in sA and syA contains the slightly rising transition from the preceding low target. In fact, a simple listening test showed that the subject himself could not always correctly identify his own SA, sA, etc., when they were picked out of a tape where he had recorded the alphabet and were presented to him in random order together with many other items picked in the same way!

But Woo concludes, "although there are reports of contour tones in syllables with short vowels, the contours never appear to be contrastive in the short syllables as they are in long syllables" (p. 132).

There is, however, a solution: The initial consonant (and also the final, if there is one) must be considered part of the tone in a segmental analysis

---

* Actually they can be regarded as different types of consonant: They can be shown - on an independent basis - to derive from different underlying contexts, that is, they are not "tonally determined allophones". It may be that the pitches are determined by the consonant types and the consonant cluster types. If so, then Woo would be correct at least on this very point with the further implication that the Tibetan idiolect we have observed is not a 'tone language'. In fact, from purely phonological considerations we can show how to predict all possible pitch contours in this language. Work is now in progress to see whether there also are physiological grounds for such assumptions.

of tone. When this is done, it turns out that, in fact, all "tones" or pitch contours of this idiolect, whether monosyllabic or polysyllabic, are predictable, phonetically; they are not lexical. (How the pitches are to be derived is described in a separate paper in this volume.) So we regard Tibetan as a non-tonal language, and think that Woo's observations may very well be valid for tone languages.

It is hoped that when this acoustical study is completed it will be of value for the description and reconstruction of Tibetan languages.

Acknowledgments

My thanks are especially due to H. Fujisaki for discussions and advice, and to S. Sekimoto and Y. Mitsui for their assistance at the computers.

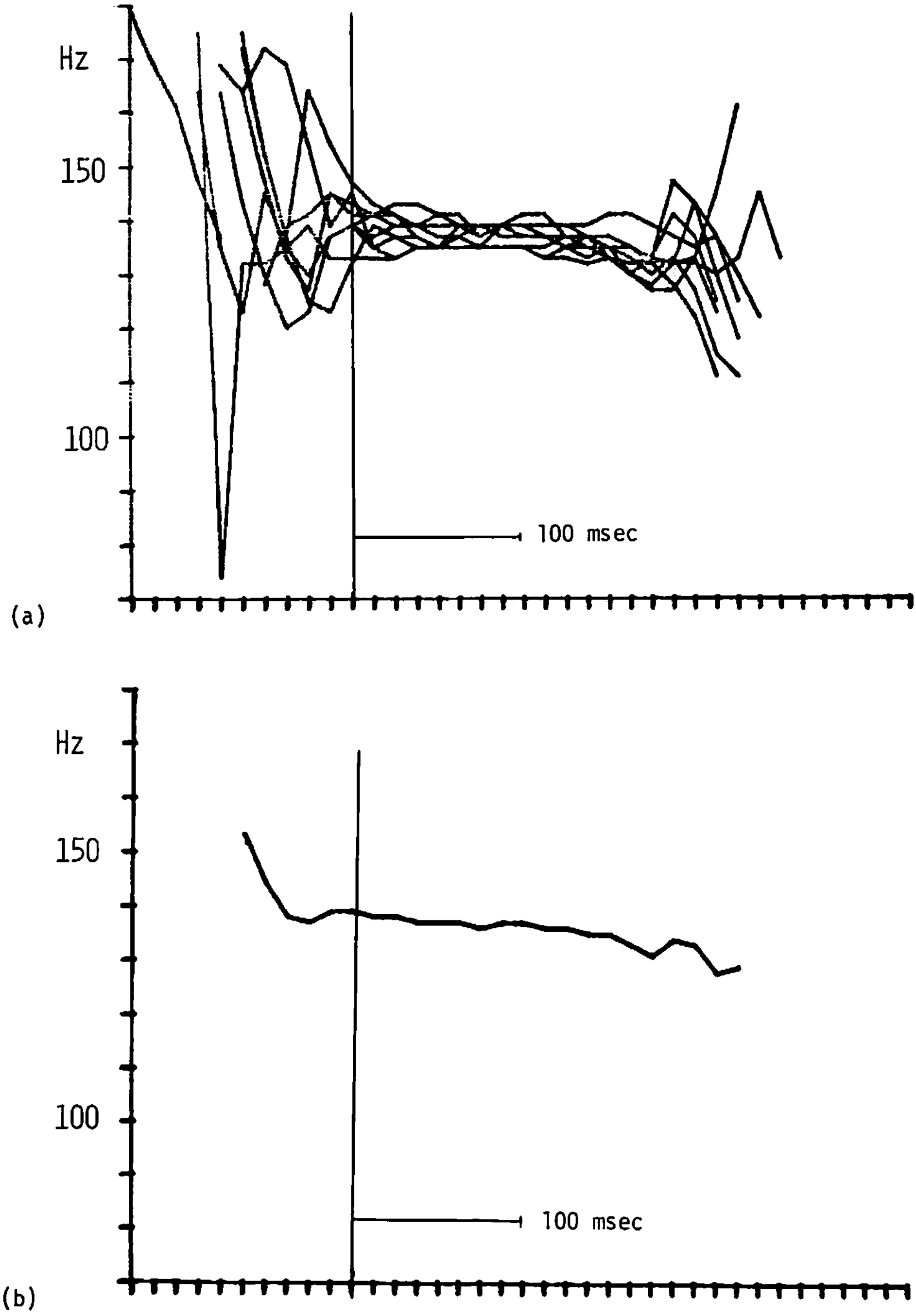


Fig. 1. MA

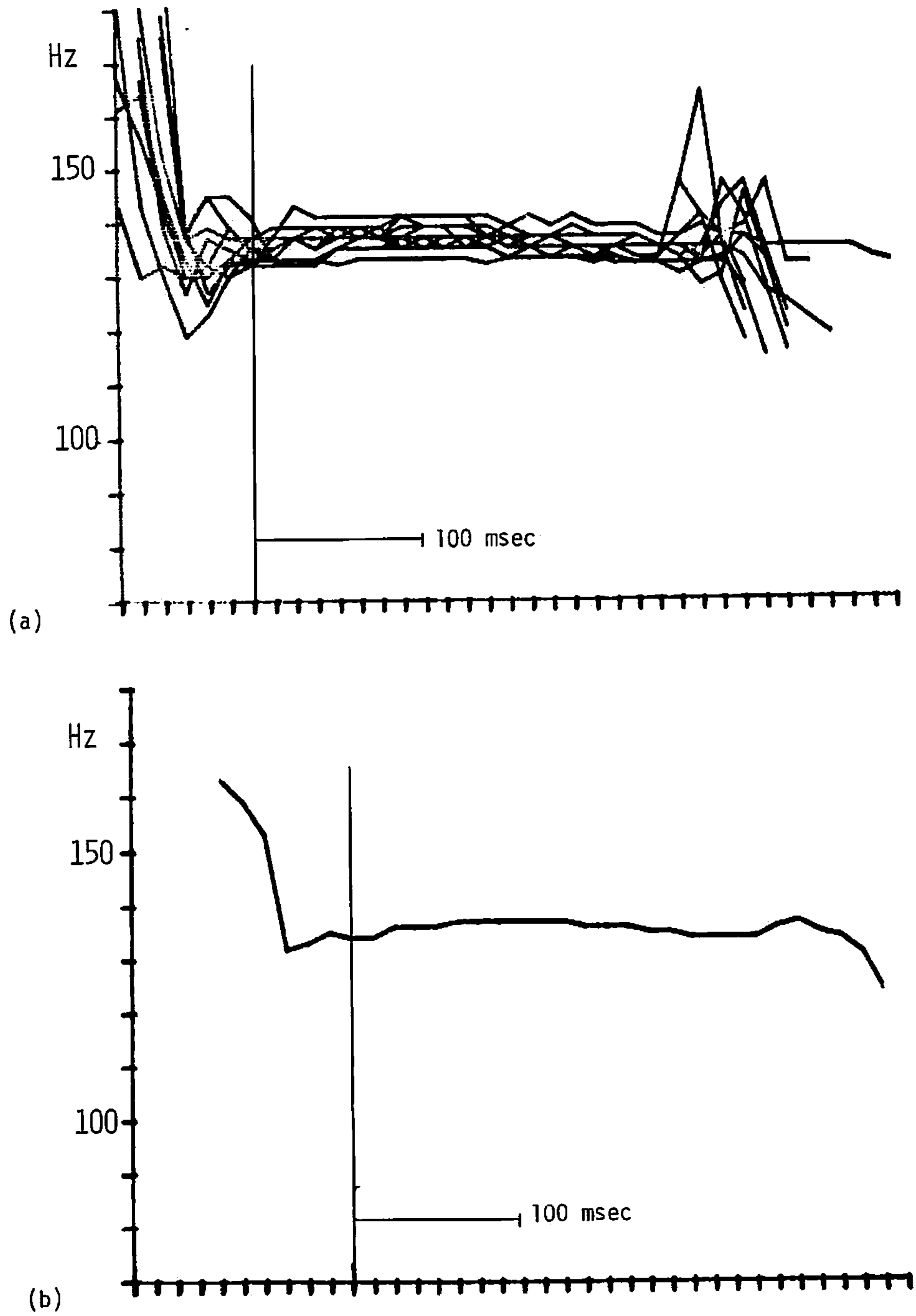
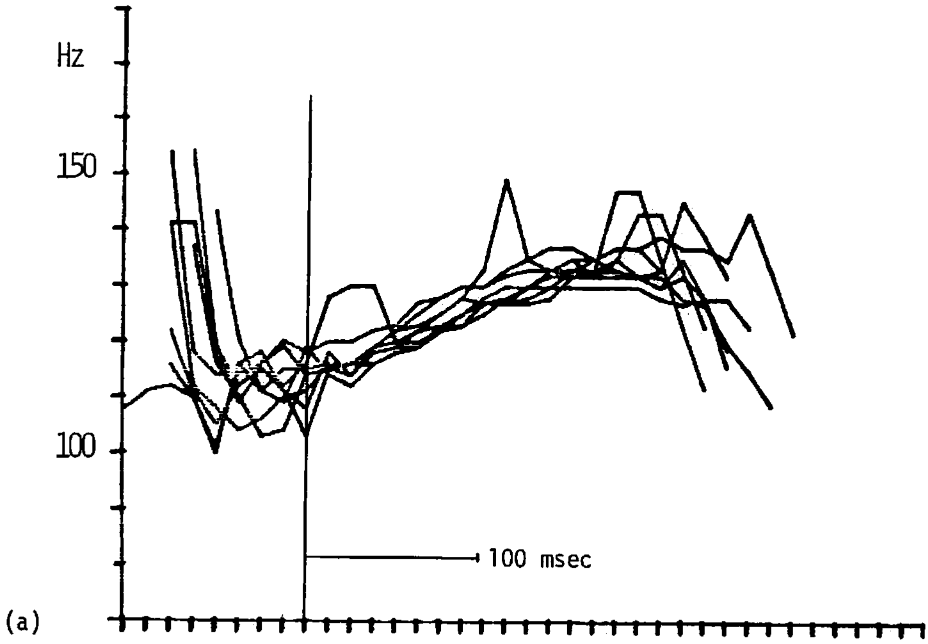


Fig. 2. <u>MAA</u>



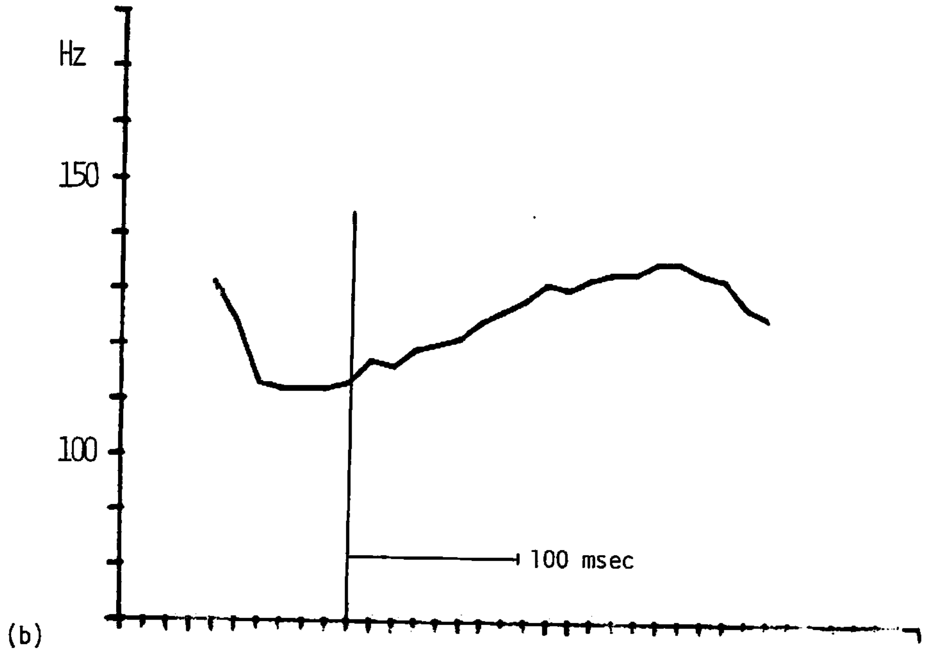


Fig. 3. mA

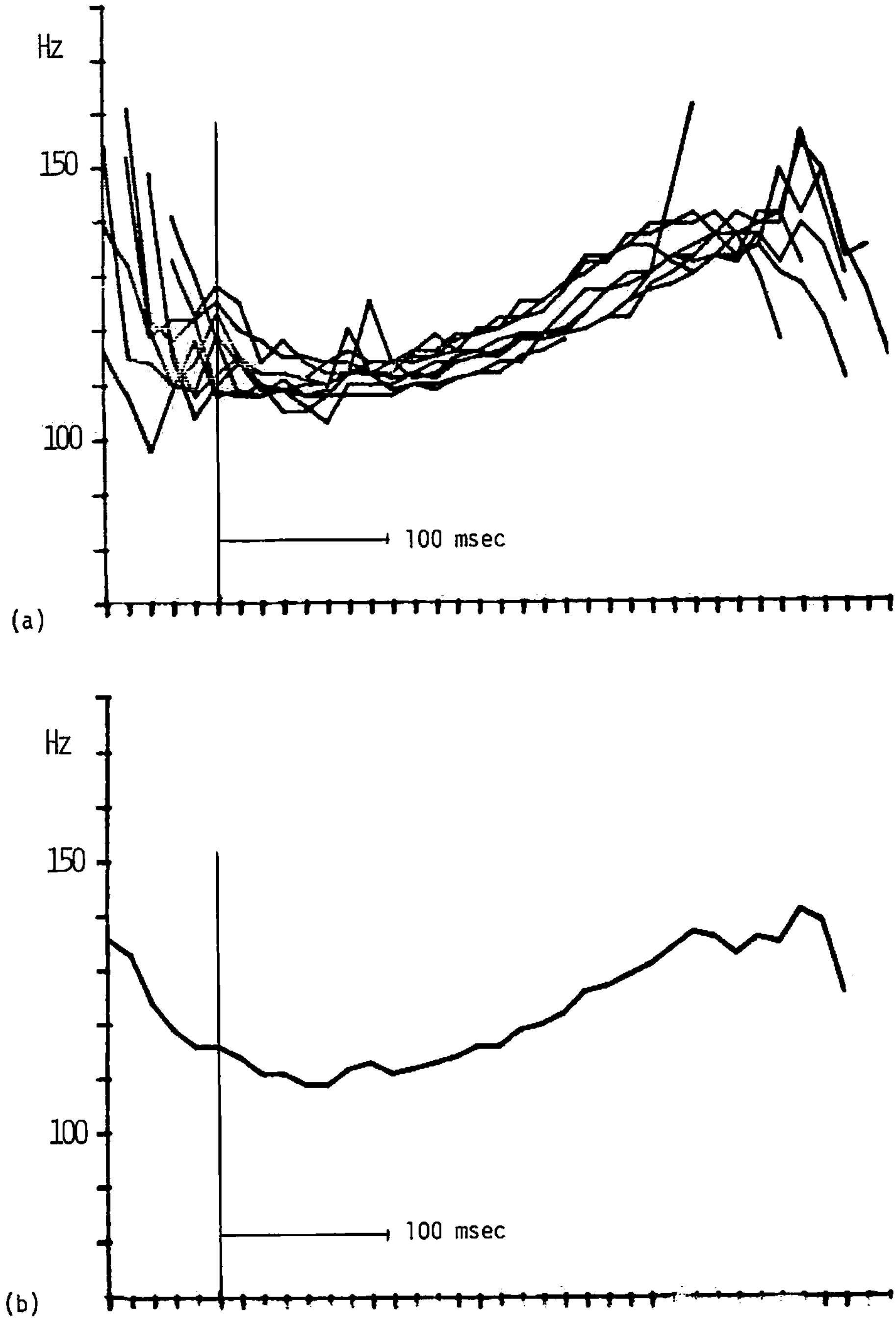
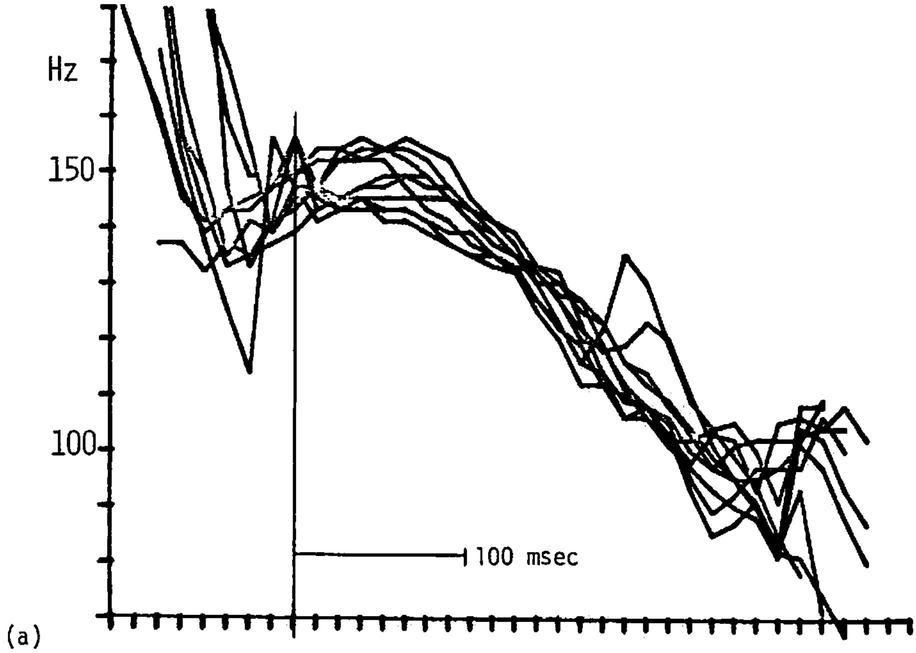


Fig. 4. maA



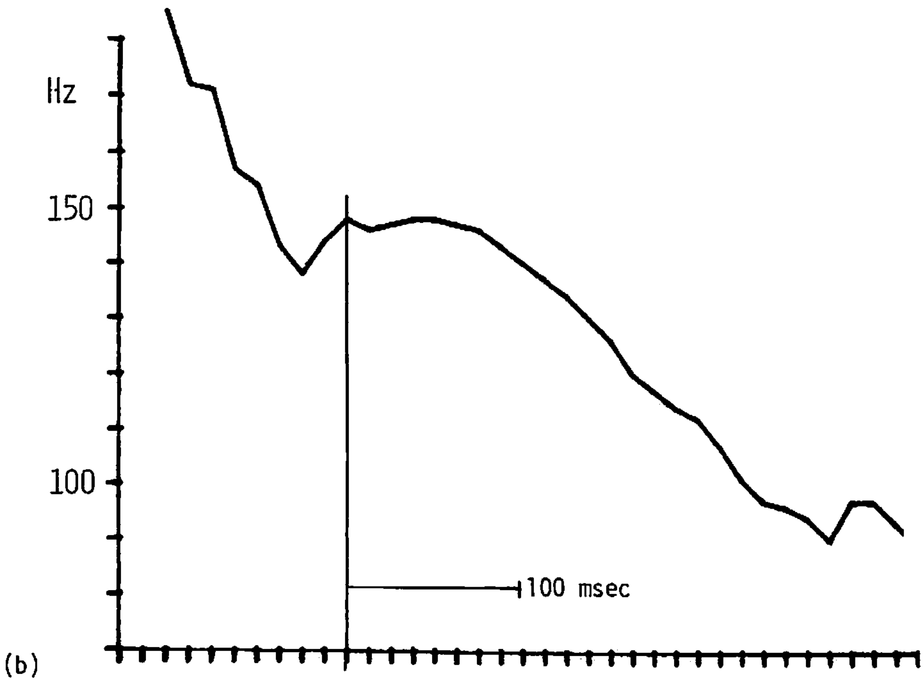


Fig. 5. MAa

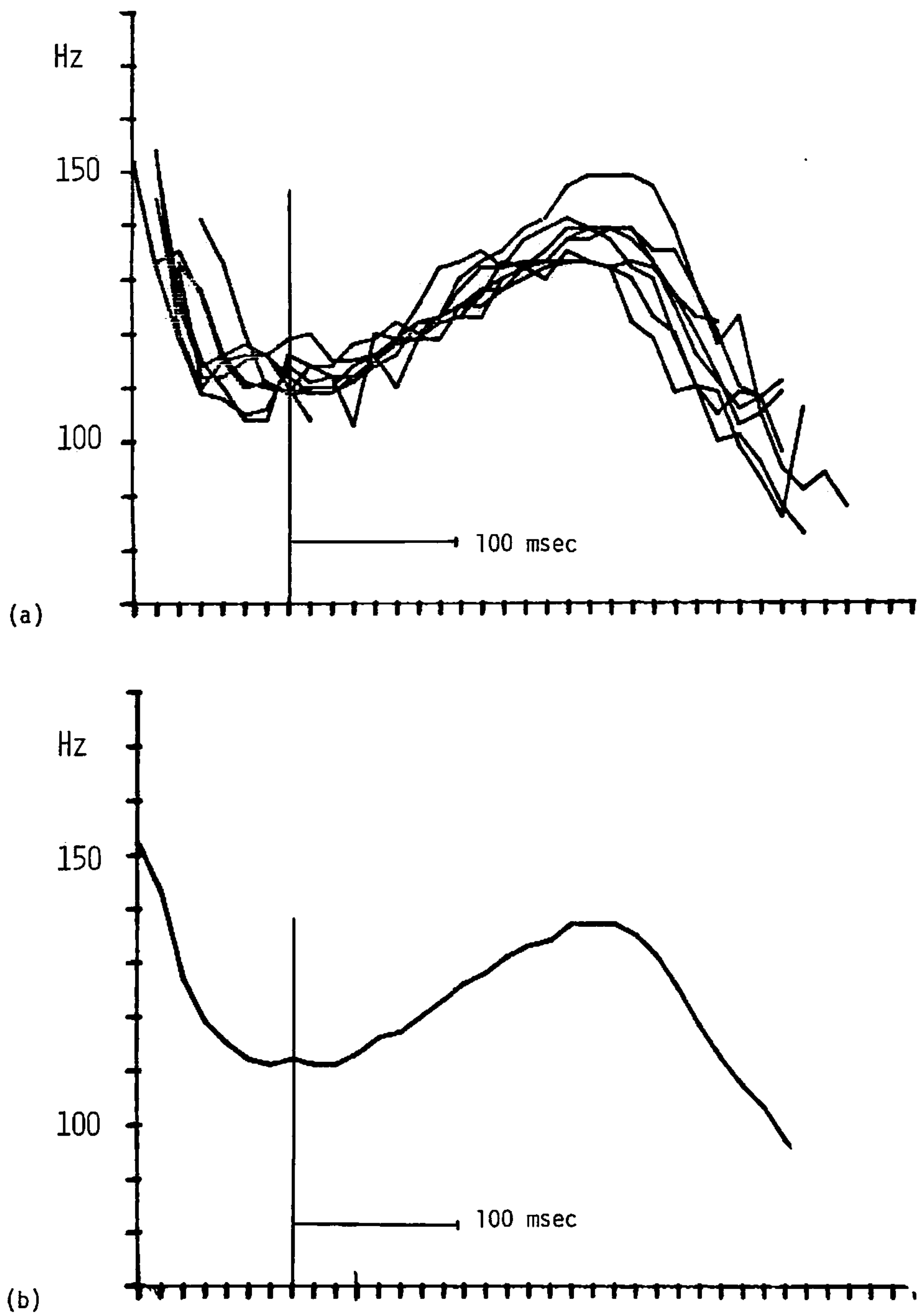


Fig. 6. <u>mAa</u>

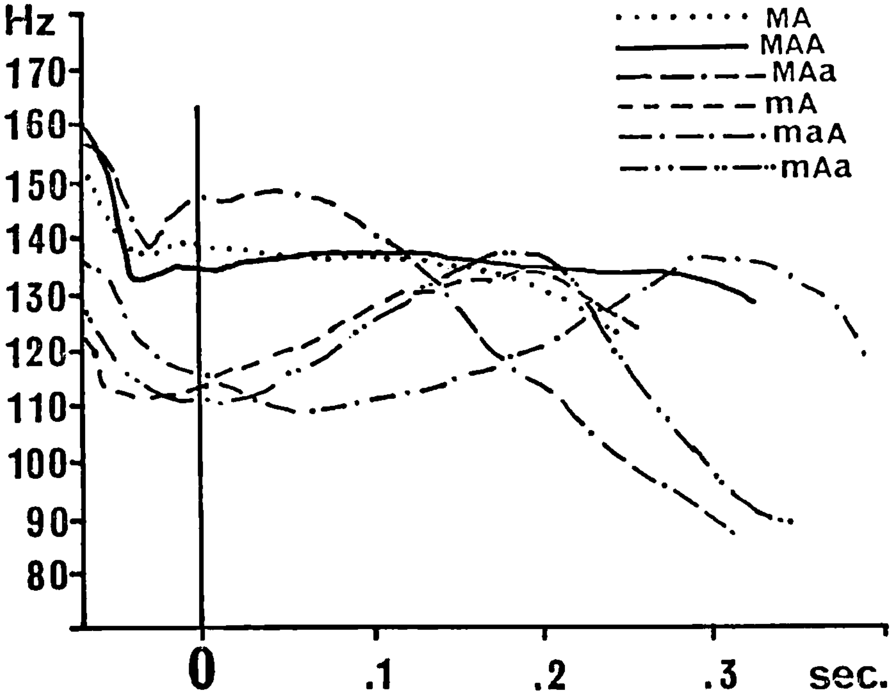


Fig. 7. The average curves plotted together. Note that the curves seem to gather toward one high and one low frequency level with or without movements between them.

References

Fujisaki, Hiroya, and Yasuyoshi Mitsui (1973); An Improved Method for Pitch Extraction of Speech in the Time Domain. Reports of the 1973 Autumn Meeting of the Acoustical Society of Japan, pp. 129-130. (In Japanese).

Fujisaki, Hiroya, and Yoshihisa Tanabe (1972); A Time-Domain Technique for Pitch Extraction of Speech. Annual Report of the Engineering Research Institute (Faculty of Engineering, University of Tokyo), Vol. 31, pp. 259-266. (A shorter version in J. of the Acoustical Society of Japan 29:7, pp. 418-419.)

Kitamura, Hajime (1974); Gendai Chibettogo no hatsuon (The Pronunciation of Modern Tibetan; in Japanese). Tokyo Univ. of Foreign Languages, Institute of the Languages and Cultures of Asia and Africa, Tokyo.

Kjellin, Olle (1974); A Preliminary Report on Tibetan Tones. Ann. Bull. RILP 8, pp. 115-123.

Noll, M. A. (1964); Short-time Spectrum and Cepstrum Techniques for Vocal-Pitch Detection. J. of the Acoustical Society of America 36, pp. 296-302.

Noll, M. A. (1966); Cepstrum Pitch Determination. J. of the Acoustical Society of America 41, pp. 293-309.

Richter, Eberhardt (1964); Grundlagen der Phonetik des Lhasa-Dialektes. Schriften zur Phonetik, Sprachwissenschaft un Kommunikationsforschung Nr. 8. Akademie-Verlag, Berlin.

Sedláček, Kamil (1959); The Tonal System of Tibetan (Lhasa Dialect). T'oung Pao 47:3/5, pp. 181-250.

Woo, Nancy (1969); Prosody and Phonology. MIT dissertation. (Reproduced 1972 by Indiana Univ. Linguistics Club.)