PREPRINT NO. 288

THE FREQUENCY SPECTRA OF SEVERAL LANGUAGES AS A CRITERION FOR THE FREQUENCY RESPONSE OF LANGUAGE LABORATORIES

by
John G. McKnight
Ampex Corporation, Audio Division
Sunnyvale, California

PRESENTED AT THE
15th ANNUAL MEETING
OCT. 14 - OCT. 18, 1963

AN AUDIO ENGINEERING SOCIETY PREPRINT

CONVENTION PRICE . . . . . $ .25
BY MAIL TO MEMBERS . . . . $ .50
BY MAIL TO NON-MEMBERS . $ .85

THE FREQUENCY SPECTRA OF SEVERAL LANGUAGES AS A CRITERION FOR THE FREQUENCY RESPONSE OF LANGUAGE LABORATORIES

John G. McKnight
Ampex Audio, Sunnyvale, Calif.

The speech of eleven speakers of six languages was measured for peak energy in one-third octave bands, as compared to a reference level established by a vu meter. Levels as high as 0 dB at 10 kc, and -18 dB at 15 kc, were found. A tape speed of at least 7.5 ips is necessary to record all of the speakers without bandwidth restrictions, or level reduction.

INTRODUCTION

The frequency response necessary for language laboratory applications is determined by the characteristics of the particular speaker, language, and listener. This paper studies the first of these characteristics - - the frequency spectra of the speaker's voice. Native speakers only were employed; one American-English speaker and two each Portugese, Russian, Burmese, Thai and Vietnamese speakers were measured.

METHOD OF ANALYSIS

The speaker was seated in a radio studio, and read in a conversational manner. A Brüel and Kjaer Model 4133 capacitor microphone was placed at a distance of approximately one foot from the speaker's mouth. (This microphone is flat $\pm$ 1 dB up to 30 kc.) The microphone output was recorded at 15 inches per second at a level 10 dB below the normal recording level (to prevent any possibility of distortion) on an Ampex Model 354 Magnetic Tape Recorder, which was flat $\pm$ 1 dB from 100 cps to 20 kc. The spectrum analysis was performed by a technique previously described by the author[1]; briefly, all 27 one-third octave filters of a Brüel and Kjaer Spectrometer, Model 2109, were simultaneously fed from the reproduced tape recording of the speech to be analyzed. Each filter fed a circuit which would "remember" the peak signal level from that filter (90% of the true reading is obtained for signals as short as one millisecond). After the recording of 30 to 45 seconds of the language was reproduced, the retained peak signals were read out of the memory unit, and recorded on a paper chart.

The recorder/reproducer/analyzer equipment was calibrated by using a "pink" random noise signal; the calibration curve (Fig. 1) shows that the response of the system was flat $\pm$ 1½ dB from 60 to 18 000 cps.

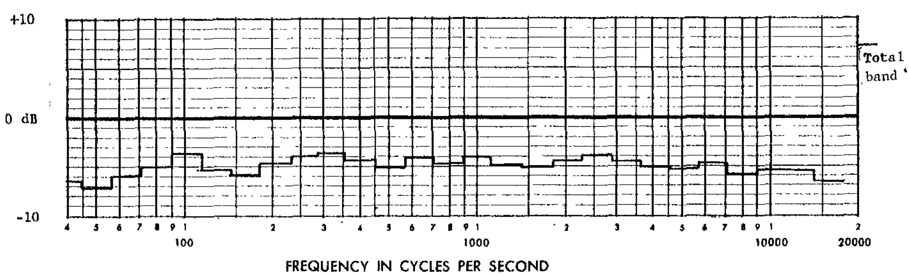

Fig. 1 Calibration curve, spectrum analyzing system. Analysis of a recording of "pink" random noise.

The actual analyses of the peak amplitudes in one-third octave bands for the languages are shown in Figs. 2 through 7. All readings are referred to the reading of "zero level" on a vu meter.

---

1. J. G. McKnight, "The Distribution of Peak Energy in Recorded Music, and Its Relation to Magnetic Recording Systems", J. Audio Eng. Soc. 7, 65-71, 80 (1959).

INTERPRETATION

The spectrum of American-English (Fig. 2) will first be examined, since it is commonly stated that a response to 3- or 4 kc is sufficient to transmit English with good intelligibility. The single male English speaker measured showed peaks of approximately +5 dB in six one-third octave bands in the 200 to 1500 cps range; but the band at 8 kc was only -5 dB, 10 kc -8 to -10 dB, and 12½ kc -15 to -19 dB. How does this come about when "English can be transmitted with good intelligibility with response to only 3- or 4 kc"? The fact is that this statement is true only for the conditions of native language speaker and listener, transmitting meaningful connected discourse, where there is considerable redundancy to supply the information missing due to the restricted frequency response.

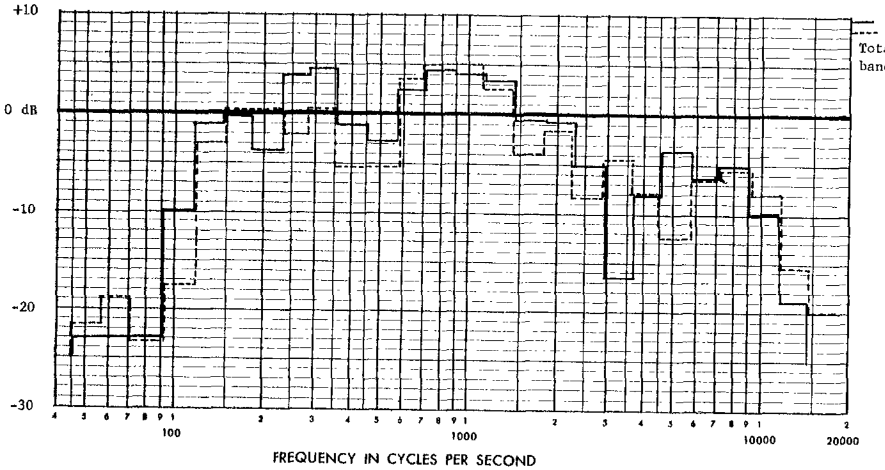

Fig. 2 Spectrum analysis of American-English. Ratio of peak energy in one-third octave bands to the vu meter reading. Male speaker; connected discourse ———— ; phonemes - - - - - -.

The male Portugese (Fig. 3) and Burmese (Fig. 4) speakers showed similar spectra. The limited data available here would suggest that there may be as much variability between speakers of a given language as there is between the various languages. This would seem reasonable since the same basic vocal apparatus is used by all speakers, with a variation from speaker to speaker having little to do with the particular language being spoken. Many of consonants having considerable high frequency content (/ f, θ, s, ʃ /) are common to, and similar in, all languages.

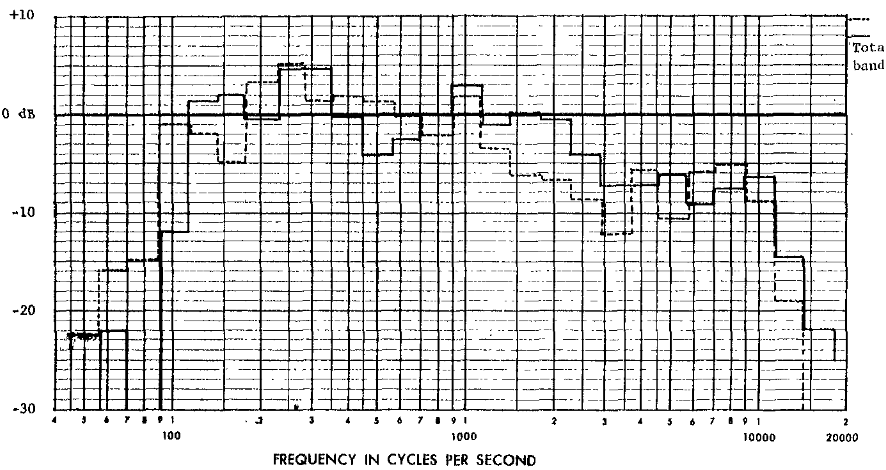

Fig. 3 Spectrum analysis of Portugese. Ratio of peak energy in one-third octave bands to the vu meter reading. First male speaker———; second male speaker ------- .

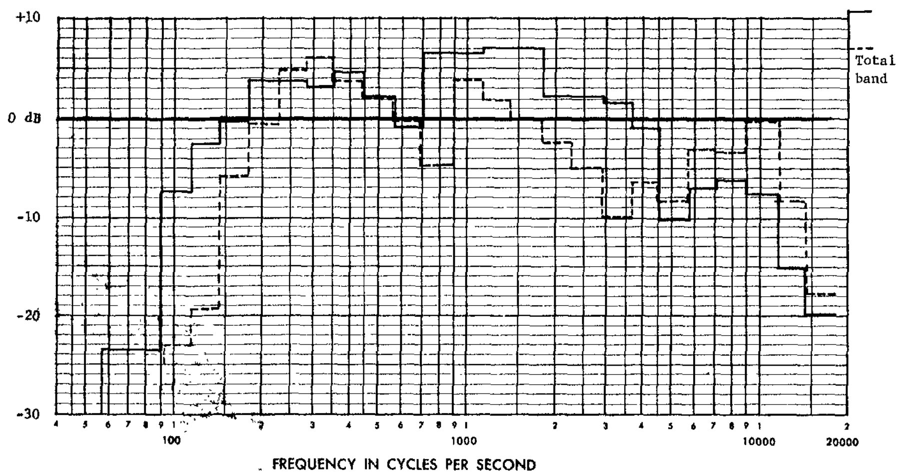

Fig. 4 Spectrum analysis of Burmese. Ratio of peak energy in one-third octave bands to the vu meter reading. Male speaker ———— ; female speaker - - - -

The female speakers had even greater high frequency energy present: the female Russian speaker (Fig. 5) had a peak of -3 dB at 10 kc, and was down 8 dB at 12½ kc, and 20 dB at 16 kc; the female Burmese speaker (Fig. 4) measured -3 dB in the bands from 6 to 9 kc, 0 dB at 10 kc, -8 dB at 12½ kc, and -20 dB at 16 kc.

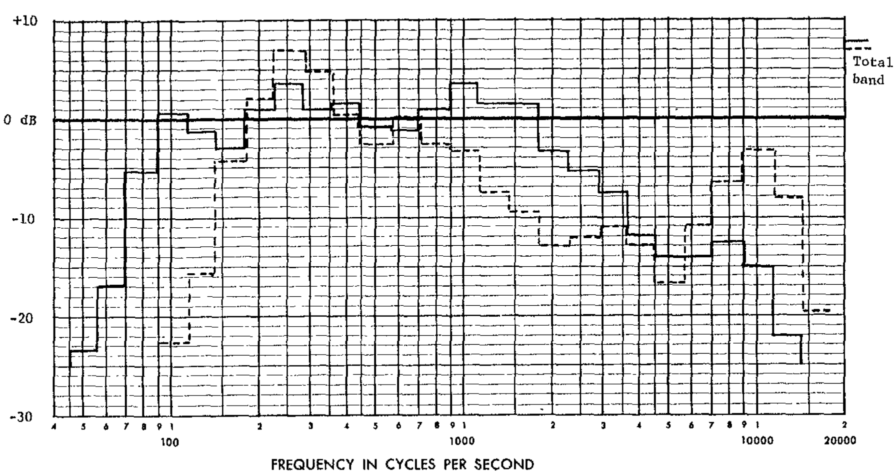

Fig. 5 Spectrum analysis of Russian. Ratio of peak energy in one-third octave bands to the vu meter reading. Male speaker ———— ; female speaker - - - -

The Thai and Vietnamese speakers showed somewhat less high frequency energy than the other speakers.

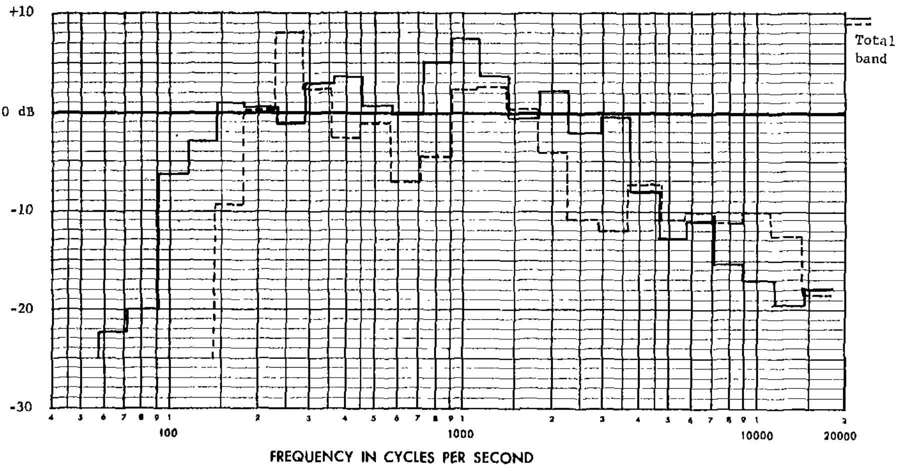

Fig. 6 Spectrum analysis of Thai. Ratio of peak energy in one-third octave bands to the vu meter reading. Male speaker ——— ; female speaker – – – – .

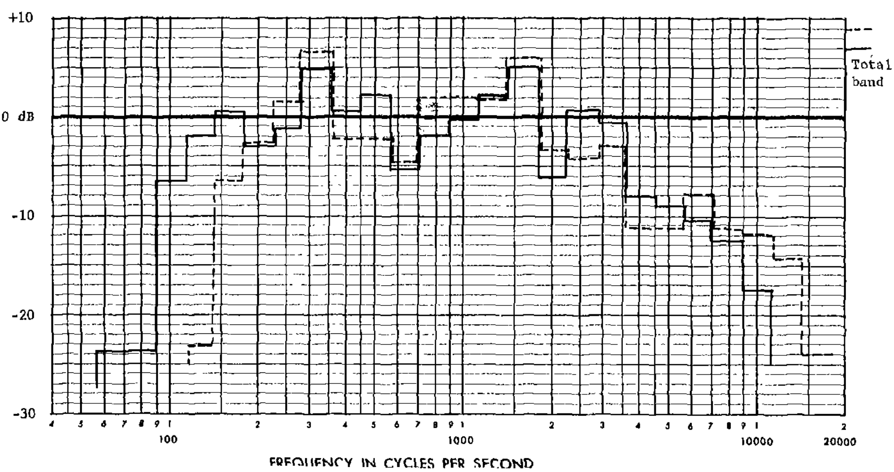

Fig. 7 Spectrum analysis of Vietnamese. Ratio of peak energy in one-third octave bands to the vu meter reading. Male speaker ——— ; female speaker – – – – .

CONCLUSIONS

A statement of the frequency response required for efficient language teaching requires knowledge which is not presently available. This paper, on the other hand, studies the amount of peak high frequency energy present in the speaking voice. (Peak energy was measured, since amplifiers and tape recorders overload at peak levels, not at the average level.)

On the basis of this small sample of American-English, Portugese, Russian and Burmese speakers, we conclude that the spoken voice includes considerable peak energy up to at least 15 kc for male voices, and up to 18 kc for female voices. The variation between speakers is at least as great as that between languages.

No measure has been made of the effect on language learning of removing these high frequencies; it may be that frequencies above 10- or 12 kc are unnecessary. In fact, presently used language laboratory microphones have little response beyond 4- to 7 kc (response rarely goes to 10 kc), so that these frequencies are not usually transmitted to the tape recorder. However, if the highest frequencies are to be fully transmitted through a system (microphone, amplifiers, tape recorder and headfones) the system must be able to transmit 12 kc at a level equal to at least -5 dB, and preferably 0dB, relative to the "zero vu" reference level. Above 12 kc the levels fall sharply, being down from 15 to 25 dB at 15 kc.

EFFECT OF ENERGY SPECTRA ON MAGNETIC TAPE RECORDERS

A boosting of the high frequencies (pre-emphasis) is necessary in magnetic recording in order to achieve flat overall frequency response with the optimum signal-to-noise ratio. The pre-emphasis, while necessary for flat frequency response, is a potential cause of increased distortion at high frequencies. The effect becomes greater as the tape speed is reduced, and/or the frequency is increased.

The exact results depend on the particular tape used, but typically a tape at 15 ips can accept full level up to the highest audio frequencies, with little distortion. At 7.5 ips, overloading and consequent distortion occur with typical tape (and, in parentheses, for the best tape presently available) when levels of 0 dB occur at frequencies above 8 (10) kc, or levels of -5dB at frequencies above 10 (12) kc. At 3.75 ips this overloading and distortion occurs when levels of 0 dB occur at frequencies above 5 (5.5) kc, or levels of -5dB at frequencies above 6.5 (7.5) kc. Due to intermodulation this distortion is apparent at the middle frequencies, not just at the high frequencies themselves; therefore, it is noticeable even if the reproducing earfones do not pass the very high frequencies.

If we compare this data on the tape recording process using "typical" tape, with the language spectral data presented above, we must conclude that in order to prevent distortion at the high and middle frequencies, either at the frequency response of a 7.5 ips recorder should be sharply restricted beyond approximately 8 kc, and that of a 3.75 ips recorder beyond approximately 5 kc; or else the recording levels should be reduced for those speakers with great amounts of high frequency energy. If the best presently available tapes are used, the full response could be accomodated at 7.5 ips, and the response restrictions at 3.75 ips would move out to 6 kc.

ACKNOWLEDGMENTS

The author is pleased to acknowledge the co-operation of the Army Language School at Monterey, California, in providing the speakers and studio facilities for the recordings from which these analyses were performed.