Hearing, the Determining Factor for High-Fidelity Transmission*

HARVEY FLETCHER

In three parts—Part 3

FROM THE ARCHIVES OF BELL TELEPHONE LABORATORIES

This is perhaps the first authoritative study of the requirements for ideal systems for the transmission of speech and music. Much of our present-day knowledge and practice stems from this article, which presents conclusions derived from measurements of hearing on more than 500,000 people during the World's Fairs in 1939 and 1940.

If loudness were the criterion of judged quality of reproduced speech, the effect of limiting the band has been measured. The loss in loudness due to frequency-band limitation depends upon the level at which the speech is reproduced. In Table VI the values are given for two loudness levels. The upper half of the table gives values for an initial loudness of 215,000 loudness units corresponding to a loudness level of 110 decibels and the lower half of the table gives values for an initial loudness of 4000 loudness units corresponding to a loudness level of 59 decibels. When it is realized that under the most favorable circumstances the ear can detect loudness differences only when they are greater than 3 or 4 per cent then it is seen that the elimination of frequencies above 5000 or below 100 would not be detected as any loudness difference. It would be useful if we had a relation such as that exhibited in Tables IV and V obtained from judgment tests of the artistic qualities of the speaking voice. From all these data then it is seen that the frequency range can be considerably more restricted for transmitting speech than for transmitting music, before serious impairment results.

TABLE IV

MUSIC—JUDGED QUALITY

Judged Quality	High-Pass Filter Cutoff	Low-Pass Filter Cutoff
100	40	15,000
97	70	12,000
93	80	9,000
90	90	7,800
85	100	6,500
80	120	5,600
70	140	4,800
60	180	4,000
50	220	3,600
40	270	3,000
30	325	2,500
20	500	1,700
10	850	850

TABLE V

SPEECH—ARTICULATION

Articulation	High-Pass Filter Cutoff	Low-Pass Filter Cutoff
per cent		
98	40	15,000
98	100	12,000
98	250	7,000
96	570	5,000
94	720	3,900
90	960	3,100
80	1,500	2,300
70	1,920	1,970
60	2,300	1,700
50	2,600	1,500

It was seen that the range from the peak value of the loudest phonetic sound to the faintest was 56 decibels but for usual conversation this is reduced to about 40 decibels. One would not expect any degradation in transmitted speech until the intensity level range is decreased below 40 decibels. Measurements on the articulation of conversational speech for smaller ranges have been made.[11] These were given in terms of the level of the speech above threshold and masking of the noise. The noise used produced a uniform masking between 250 and 10,000 cps, dropping off at either side of these limits. First let us consider conversational speech of men and assume the listener wishes to hear it at levels he would obtain if he were 2½ feet from the speaker instead of 20 feet. The values in Table I would then be raised 18 decibels. From these values then we must deduce the level of the speech above the standard threshold if we are to apply the data referred to above.

Experiments in our laboratories have shown that the threshold level, for observers with acute hearing of conversational speech which is undistorted, is at a long root-mean-square level of 5 decibels. Now the long root-mean-square level for speech was found by Dunn and White to be 10 decibels below the maximum root-mean-square level in ¼-second intervals. Consequently at 2½ feet from the speaker the men's conversational speech will be 63 decibels above threshold. The articulation values given in Table VII are taken from *Fig.* 148 of "Speech and Hearing."[9] The noise levels corresponding to the various values of masking were obtained directly from the curve in *Fig.* 3 of the paper, "Relation between Loudness and Masking."[4] In computing the range the highest level was considered to be the maximum root-mean-square value in ¼ second, namely, 78 decibels. If we used long root-mean-square values of both speech and noise these ranges would all be reduced 10 decibels. These range values will of course depend upon the spectrum of the noise but the values given here are enough to show that when the range of a system for single-frequency tones is 40 decibels from 100 to 7000 cps very little

[11] Loc cit., p. 298.

[9] Harvey Fletcher, "Speech and Hearing," D. Van Nostrand Company, New York, N. Y., 1929, p. 75.

TABLE VI

SPEECH—LOUDNESS

Loudness	High-Pass Filter Cutoff	Low-Pass Filter Cutoff
per cent		
215,000 = 100	40	15,000
98	100	5,000
90	200	3,500
80	300	2,100
70	450	1,500
60	600	1,100
50	800	800
4,000 = 100	40	15,000
98	200	4,000
90	500	2,500
80	770	2,000
70	950	1,700
60	1,125	1,450
50	1,250	1,250

TABLE VII

CONVERSATION SPEECH—SPEAKER 2½ FEET FROM LISTENER

Level of speech above threshold	= 63 db
Long root-mean-square intensity level	= 68 db
Peak level in ⅛-second intervals exceeded 5 per cent of the time	= 85 db
Root-mean-square level in ⅛-second intervals exceeded only 1 per cent of the time	= 78 db

Articulation Masking		Noise Level	Range
per cent			
97	0	10	68
95	10	36	42
88	20	46	32
77	30	57	21
55	40	67	11
28	50	77	− 1

or no distortion is produced when speech is transmitted. However, if we wish the noise from the system to be inaudible, the levels must be below the curve of *Fig.* 3. This means a range of 48 decibels, that is, from 78 to 30 decibels in the frequency region where the maximum speech levels occur.

Multiple Channel Transmission

The third way in which the quality of the transmission can be improved if there are no economic deterrents is by using more than one channel for the transmission. We have found that the quality of reproduced music is very much improved by using two or three channels. As stated above, such transmission makes it possible to produce apparent motion of the sound and provides much greater possibilities in dramatic productions.

Some preliminary tests to determine in a quantitative way the increased quality due to using just two channels instead of one have been made in our laboratories. In one of these a dramatic skit used very simple program material and was designed to tie together smoothly a number of sounds rich in high frequencies. It opened with a man's voice dictating a letter to a woman on right. Then the man walked from right to left and back and engaged in a short conversation. The typewriter started and the man walked to the center and made a phone call. After another conversation between the man at center and the woman at right, the man walked to the left jingling keys and opened a steel cabinet at left, and so on.

For the single-channel condition the plot was the same but the props were shifted in position and the action was adjusted to the restricted space. For each judgment the observers listened first to the single-channel full-frequency-range version, and then with the smallest possible interval to the filtered two-channel version.

In the two-channel system filters were introduced and a number of observers indicated which system they preferred. In Table VIII are shown the results. When the cutoff for the low-pass filter was somewhere between 5500 and 7000 then one half the group preferred the two-channel system with filters and the other half preferred the single-channel

TABLE VIII

ONE-CHANNEL VERSUS TWO-CHANNEL TRANSMISSION

Low-Pass Cutoff	Speech and Noise Per Cent Preference	
	One Channel Full Range	Two Channels Limited Range
15,000	3	97
11,000	10	90
8,000	27	73
7,000	39	61
5,500	71	29
3,800	85	15
High-Pass Cutoff		
125	3	97
250	19	81
500	47	53

which transmitted all frequencies from 40 to 15,000 eps. Also it is seen that according to these tests the two-channel system filtering all frequencies below 500 was considered as good as the wide-band single channel. Similar tests were made in a very preliminary way on a 45-piece orchestra during a boadcast and the single channel was found to be equivalent to a two-channel with a low-pass cutoff of about 5000 cycles. These

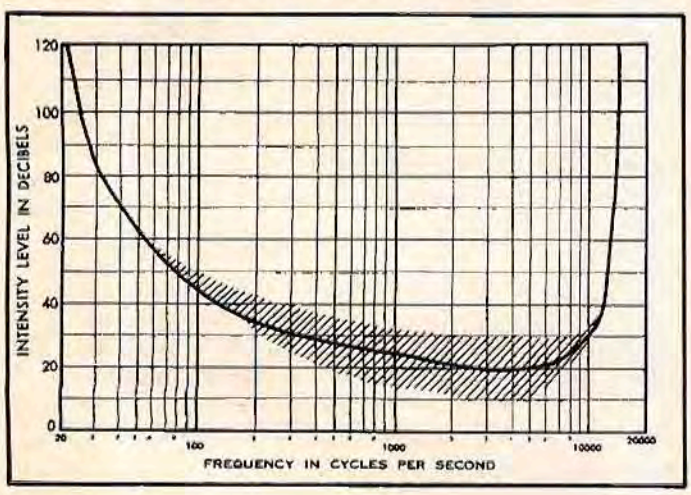

Fig. 3. Hearing limits for pure tones—typical listener in typical residential room noise.

TABLE IX

Low-Pass Cutoff	Preference	
	Binaural	Diotic
	per cent	per cent
8,500	68	32
5,500	58	42
4,510	70	30
3,750	50	50
2,850	42	58

tests are very preliminary and are only indicative of what more accurate tests might show. No direct judgment tests of a quantitative nature have been made between a two- and a three-channel system but comparison between them indicates that three channels are definitely better. This is particularly true when the sound is reproduced from a stage into a large hall.

Although head receivers are seldom used as receivers to listen to music, it may be interesting to describe some tests made in our laboratories to test the judged quality of music transmitted by a binaural system versus that transmitted by a diotic system. In these tests the listener had a pair of high-quality headphones clamped on his ears. He could listen under condition *A* (binaural) where two high-quality circuits and microphones were used to transmit music to him; condition *B* (diotic) where one channel and one microphone were used to transmit to his pair of headphones. In the diotic system the full-range of frequencies from 40 to 15,000 eps was transmitted. In the binaural system low-pass filters were introduced to eliminate a part of the upper range. Twenty-five engineers took part in the preference tests. The Philadelphia Orchestra, playing various selections, was used as the music for the test. The results are shown in Table IX. The percentage of persons preferring the binaural system with limited range is shown under the title BINAURAL. These results show that a binaural system which transmits a frequency band greater than 3750 eps was preferred to a diotic band transmitting the entire range. In other words, if a band 8000 eps wide or greater is available for a transmission channel, this indicates that better results may be achieved by using two channels binaurally than one diotically. It is rather remarkable that one half of the observers preferred a binaural system having all frequencies above 3750 eps eliminated rather than a single-channel diotic system transmitting all frequencies between 40 and 15,000 eps.

Before definite conclusions can be drawn as to the improvement of one versus two versus three channels for various purposes, much more data must be collected. It may not be economically feasible to use more than one channel in

(*Continued on page 53*)

HEARING

(*from page 36*)

broadcasting but it should be kept in mind when considering what improvements are possible in the quality of the reproduced sounds.

In conclusion then, it is seen that for an ideal system, one which is determined by the capabilities of hearing, the maximum and minimum intensity levels and frequencies will be determined by *Fig.* 3 and two or three channels should be used in the transmission. With such a system there will be no limitation upon the type of material used in the broadcast. For economic reasons we may back away from these ideal requirements. Although the best quality cannot be obtained with a frequency range lower than 14,000 to 15,000 cps, economic necessity may require a compromise to a somewhat lower frequency limit. The ideal volume range for producing a facsimile of such music is 65 decibels. How much this volume range can be reduced below this value without producing serious impairment has not been determined in a quantitative way. There is no doubt that considerable improvement in quality will also result by going to two or more channels instead of one, but whether such improvement is worth the additional cost must be decided for each kind of service in which the transmission system is used.