Physical Factors*

By J. C. STEINBERG and W. B. SNOW

In considering the physical factors affecting it, auditory perspective is defined in this paper as being reproduction which preserves the spatial relationships of the original sounds. Ideally, this would require an infinite number of separate microphone-to-speaker channels; practically, it is shown that good auditory perspective can be obtained with only 2 or 3 channels.

ABILITY to localize the direction, and to form some judgment of the distance from a sound source under ordinary conditions of listening, are matters of common experience. Because of this faculty an audience, when listening directly to an orchestral production, senses the spatial relations of the instruments of the orchestra. This spatial character of the sounds gives to the music a sense of depth and of extensiveness, and for perfect reproduction should be preserved. In other words, the sounds should be reproduced in true *auditory perspective.*

In the ordinary methods of reproduction, where only a single loud speaking system is used, the spatial character of the original sound is imperfectly preserved. Some of the depth properties of the original sound may be conveyed by such a system,[1] but the directional properties are lost because the audience tends to localize the sound as coming from the direction of a single source, the loud speaker. Ideally, there are two ways of reproducing sounds in true auditory perspective. One is binaural reproduction which aims to reproduce in a distant listener's ears, by means of head receivers, exact copies of the sound vibrations that would exist in his ears if he were listening directly. The other method, which was described in the first paper of this series, uses loud speakers and aims to reproduce in a distant hall an exact copy of the pattern of sound vibration that exists in the original hall. In the limit, an infinite number of microphones and loud speakers of infinitesimal dimensions would be needed.

Far less ideal arrangements, consisting of as few as two microphone-loudspeaker sets, have been found to give good auditory perspective. Hence, it is not necessary to reproduce in the distant hall an exact copy of the vibrations existing in the original hall. What physical

* Second paper in the Symposium on Wire Transmission of Symphonic Music and Its Reproduction in Auditory Perspective. Presented at Winter Convention of A. I. E. E., New York City, Jan. 23–26, 1934. Published in *Electrical Engineering,* January, 1934.

properties of the waves must be preserved then, and how are these properties preserved by various arrangements of 2- and 3-channel loudspeaker reproducing systems? To answer these questions, some very simple localization tests have been made with such systems. Perhaps attention can be focused more easily on their important properties by considering briefly the results of these tests.

Localization Afforded by Multichannel Systems

In Fig. 1 is shown a diagram of the experimental set-up that was used. The microphones, designated as *LM* (left), *CM* (center), and *RM* (right), were set on a "pick-up" stage that was marked out on the floor of an acoustically treated room. The loud speakers, designated as *LS*, *CS*, and *RS*, were placed in the front end of the auditorium at the Bell Telephone Laboratories and were concealed from view by a curtain of theatrical gauze. The average position of a group of twelve observers is indicated by the cross in the rear center part of the auditorium.

The object of the tests was to determine how a caller's position on the pick-up stage compared with his apparent position as judged by the group of observers in the auditorium listening to the reproduced speech. Words were uttered from some 15 positions on the pick-up stage in random order. The 9 positions shown in Fig. 1 were always included in the 15, the remaining positions being introduced to minimize memory effects. The reproducing system was switched off while the caller moved from one position to the other.

In the first series of tests, the majority of the observers had no previous experience with the set-up. They simply were given a sheet of coordinate paper with a single line ruled on it to indicate the line of the gauze curtain and asked to locate the apparent position of the caller with respect to this line. Following these tests, the observers were permitted to listen to speech from various announced positions on the pick-up stage. This gave them some notion of the approximate outline of what might be called the "virtual" stage. These tests then were repeated. As there was no significant difference in results, the data from both tests have been averaged and are shown in Fig. 1.

The small diagram at the top of Fig. 1 shows the caller's positions with respect to the microphone positions on the pick-up stage. The corresponding average apparent positions when reproduced are shown with respect to the curtain line and the loudspeaker positions. The type of reproduction is indicated symbolically to the right of the apparent position diagrams.

With 3-channel reproduction there is a reasonably good corre-

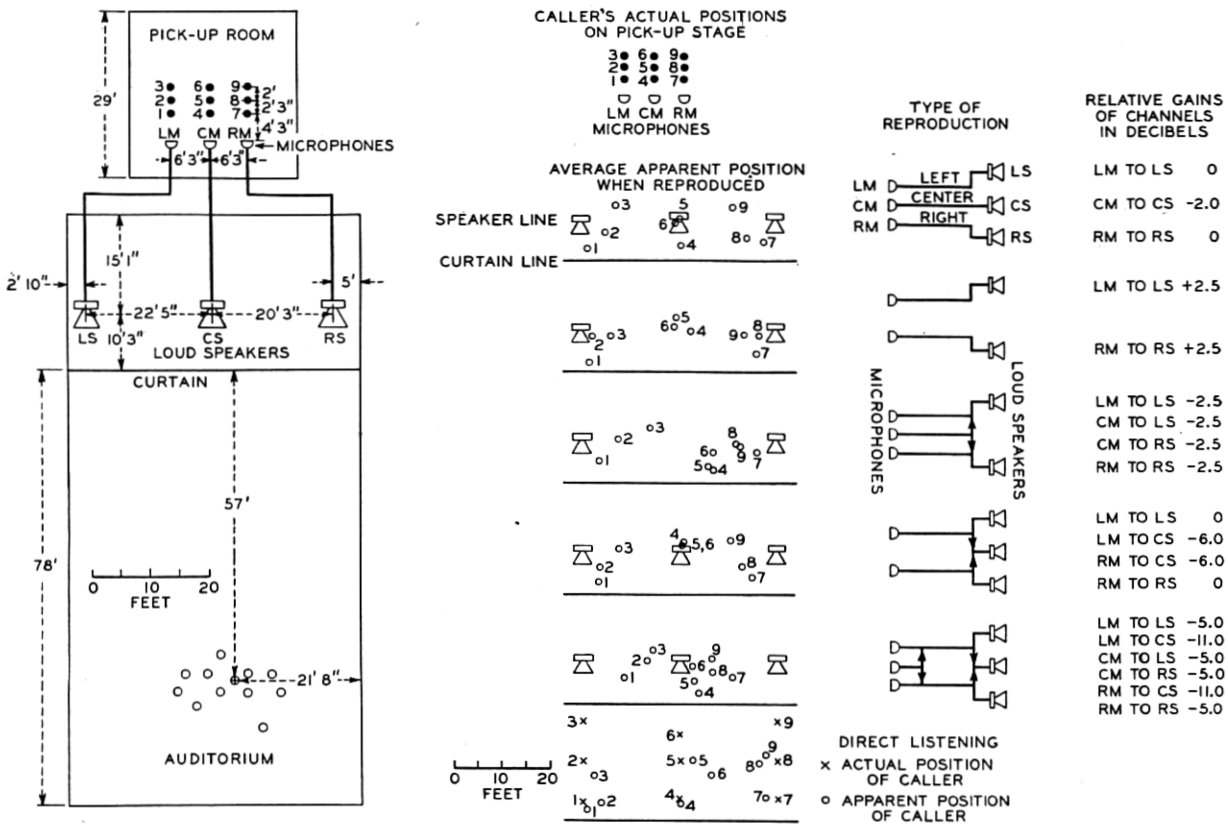

Fig. 1—Diagram of arrangement (left) for sound localization tests and (right) the results obtained.

spondence between the caller's actual position on the pick-up stage and his apparent position on the virtual stage. Apparent positions to the right or left correspond with actual positions to the right or left, and apparent front and rear positions correspond with actual front and rear positions. Thus the system afforded lateral or "angular" localization as well as fore and aft or "depth" localization. For comparison, there is shown in the last diagram the localization afforded by direct listening. The crosses indicate a caller's position in back of the gauze curtain and the circles indicate his apparent position as judged by the observers listening to his speech directly. In both cases, as the caller moved back in a straight line on the left or right side of the stage, he appeared to follow a curved path pulling in toward the rear center; e.g., compare the caller positions 1, 2, 3, with the apparent positions 1, 2, 3. This distortion was somewhat greater for 3-channel reproduction than for direct listening.

The results obtained with the 2-channel system show two marked differences from those obtained with 3-channel reproduction. Positions on the center line of the pick-up stage (i.e., 4, 5, 6) all appear in the rear center of the virtual stage, and the virtual stage depth for all positions is reduced. The virtual stage width, however, is somewhat greater than that obtained with 3-channel reproduction.

Bridging a third microphone across the 2-channel system had the effect of pulling the center line positions 4, 5, 6, forward, but the virtual stage depth remained substantially that afforded by 2-channel reproduction, while the virtual stage width was decreased somewhat. In this and the other bridged arrangements the bridging circuits employed amplifiers, as represented by the arrows in Fig. 1, in such a way that there was a path for speech current only in the indicated direction.

Bridging a third loud speaker across the 2-channel system had the effect of increasing the virtual stage depth and decreasing the virtual stage width, but positions on the center line of the pick-up stage appeared in the rear center of the virtual stage as in 2-channel reproduction.

Bridging both a third microphone and a third loud speaker across the 2-channel system had the effect of reducing greatly the virtual stage width. The width could be restored by reducing the bridging gains, but fading the bridged microphone out caused the front line of the virtual stage to recede at the center, whereas fading the bridged loud speaker out reduced the virtual stage depth. No fixed set of bridging gains was found that would enable the arrangement to create the virtual stage created by three independent channels. The gains used in obtaining the data shown in Fig. 1 are indicated at the right of the symbolic circuit diagrams.

Factors Affecting Depth Localization

Before attempting to explain the results that have been given in the foregoing, it may be of interest to consider certain additional observations that bear more specifically upon the factors that enter into the "depth" and "angular" localization of sounds. The microphones on the pick-up stage receive both direct and reverberant sound, the latter being sound waves that have been reflected about the room in which the pick-up stage is located. Similarly, the observer receives the reproduced sounds directly and also as reverberant sound caused by reflections about the room in which he listens. To determine the effects of these factors, the following three tests were made:

1. Caller remained stationary on the pick-up stage and close to microphone, but the loudness of the sound received by the observer was reduced by gain control. This was loudness change without a change in ratio of direct to reverberant sound intensity.
2. Caller moved back from microphone, but gain was increased to keep constant the loudness of the sound received by the observer. This was a change in the ratio of direct to reverberant sound intensity without a loudness change.
3. Caller moved back from microphone, but no changes were made in the gain of the reproducing system. This changed both the ratio and the loudness.

All of the observers agreed that the caller appeared definitely to recede in all three cases. That is, either a reduction in loudness or a decrease in ratio of direct to reverberant sound intensity, or both, caused the sound to appear to move away from the observer. Position tests using variable reverberation with a given pick-up stage outline showed that increasing the reverberation moved the front line of the virtual stage toward the rear, but had slight effect upon the rear line. When the microphones were placed outdoors to eliminate reverberation, reducing the loudness either by changing circuit gains or by increasing the distance between caller and microphone moved the whole virtual stage farther away. It is because of these effects that all center line positions on the pick-up stage appeared at the rear of the virtual stage for 2-channel reproduction.

It has not been found possible to put these relationships on a quantitative basis. Probably a given loudness change, or a given change in ratio of direct to reverberant sound intensity, causes different sensations of depth depending upon the character of the reproduced sound and upon the observer's familiarity with the acoustic conditions surrounding the reproduction. Since the depth localization is inaccurate even when listening directly, it is difficult to obtain sufficiently accurate data to be of much use in a quantitative way. Because of this inaccuracy, good auditory perspective may be obtained with reproduced sounds even though the properties controlling depth localization depart materially from those of the original sound.

Angular Localization

Fortunately, the properties entering into lateral or angular localization permit more quantitative treatment. In dealing with angular localization, it has been found convenient to neglect entirely the effects of reverberant sound and to deal only with the properties of the sound waves reaching the observer's ears without reflections. The reflected waves or reverberant sounds do appear to have a small effect on angular localization, but it has not been found possible to deal with such sound in a quantitative way. One of the difficulties is that, because of differences in the build-up times of the direct and reflected sound waves, the amount of direct sound relative to reverberant sound reaching the observer's ears for impulsive sounds such as speech and music is much greater than would be expected from steady state methods of dealing with reverberant sound.

For the case of a plane progressive wave from a single sound source, and where the observer's head is held in a fixed position, there are apparently only three factors that can assist in angular localization: namely, phase difference, loudness difference, and quality difference between the sounds received by the two ears.

In applying these factors to the localization of sounds from more than one source, as in the present case, the effects of phase differences have been neglected. It is difficult to see how phase differences in this case can assist in localization in the ordinary way. The two remaining factors, loudness and quality differences, both arise from the directivity of hearing. This directivity probably is due in part to the shadow and diffraction effects of the head and to the differences in the angle subtended by the ear openings. Measurements of the directivity with a source of pure tone located in various positions around the head in a horizontal plane have been reported by Sivian and White.[2] From these measurements, the loudness level differences between near and far ears have been determined for various frequencies. These differences are shown in Fig. 2 from which, using the pure tone data given, similar loudness level differences for complex tones may be calculated. Such calculated differences for speech are shown in Fig. 3.

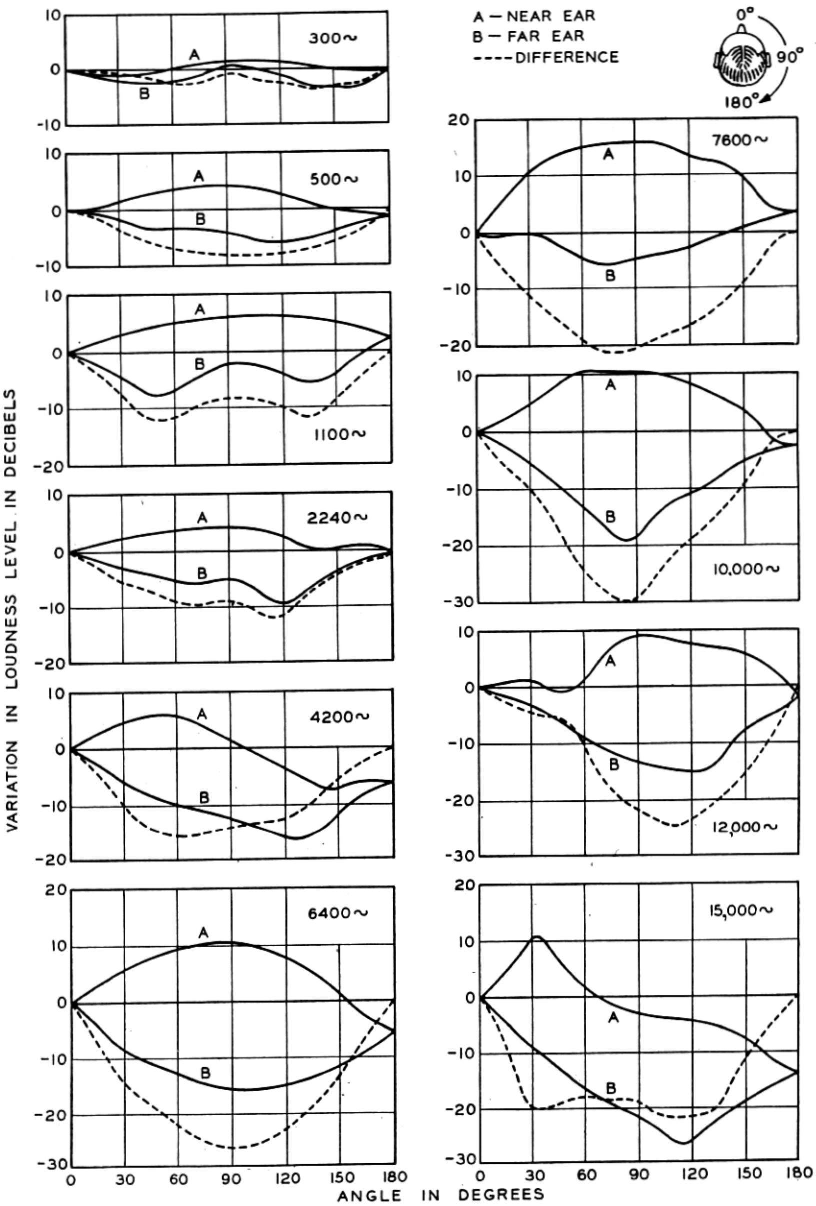

Fig. 2—Variation in loudness level as a sound source is rotated in a horizontal plane around the head.

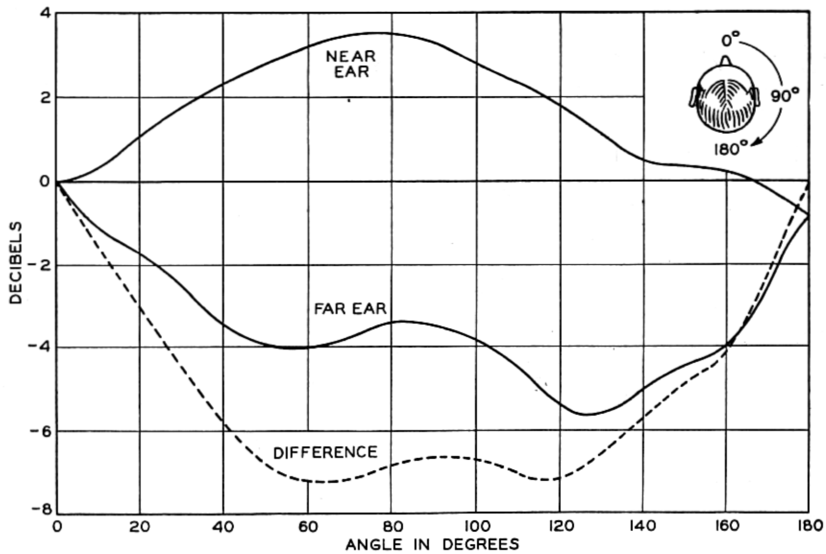

Fig. 3—Variation in loudness as a speech source is rotated in a horizontal plane around the head.

As may be inferred from the varying shapes of the curves of Fig. 2, the directive effects of hearing introduce a frequency distortion more or less characteristic of the direction from which the sound comes. Thus the character or quality of complex sounds varies with the angle of the source. There are quality differences at each ear for various angles of source, and quality differences between the two ears for a given angle of source. In Fig. 4 is shown the frequency distortion at the right ear when a source of sound is moved from a position on the right to one on the left of an observer. It is a graph of the "difference" values of Fig. 2 for an angle of 90 degrees. Frequencies above 4,000 cycles per second are reduced by as much as 15 to 30 decibels. This amount of distortion is sufficient to affect materially the quality of speech, particularly as regards the loudness of the sibilant sounds.

Reference to the difference curve of Fig. 3 shows that if, for example, a source of speech is 20 degrees to the right of the median plane the speech heard by the right ear is 3 db louder than that heard by the left ear. A similar difference exists when the angle is 167 degrees. Presumably, when the right ear hears speech 3 db louder than the left, the observer localizes the sound as coming from a position 20 degrees or 167 degrees to the right, depending upon the quality of the speech. If this be assumed to be true, even though the difference is caused by the combination of sounds of similar quality from several sources, it should be possible to calculate the apparent angle.

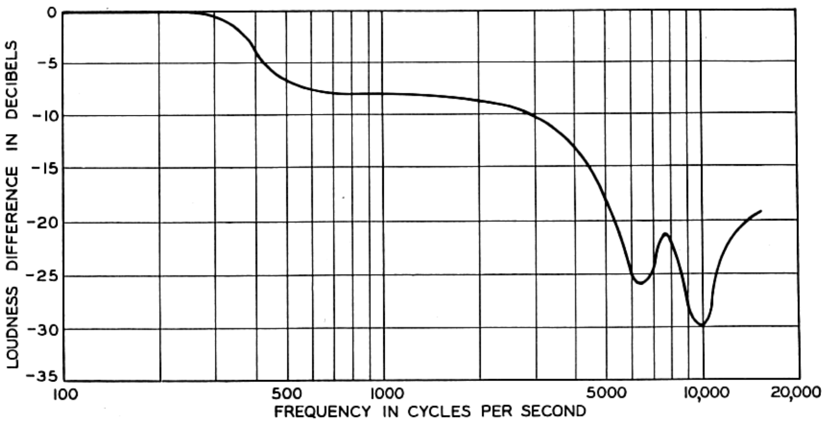

Fig. 4—Loudness difference produced in the right ear when a source of pure tone is moved from the right to the left of an observer.

Loudness Theory of Localization

Upon this assumption the apparent angle of the source as a function of the difference in decibels between the speech levels emitted by the loud speakers of the 2- and 3-channel systems has been calculated. Each loud speaker contributes an amount of direct sound loudness to each ear, depending upon its distance from, and its angular position with respect to, the observer. These contributions were combined on a power basis to give a resultant loudness of direct sound at each ear, from which the difference in loudness between the two ears was determined. The calculated results for the 2- and 3-channel systems are shown by the solid lines in Fig. 5. The y axis shows the apparent angle, positive angle being measured in a clockwise direction. The x axis shows the difference in decibels between the speech levels from the right and left loud speakers. The points are observed values taken from Fig. 1. The observed apparent angles were obtained directly from the average observer's location and the average apparent positions shown in Fig. 1. The speech levels from each of the loud speakers were calculated for each position on the pick-up stage. This was done by assuming that the waves arriving at the microphone had relative levels inversely proportional to the squares of the distances traversed. By correcting for the angle of incidence and for the known relative gains of the systems, the speech levels from the loud speakers were obtained.

A comparison of the observed and calculated results seems to indicate that the loudness difference at the two ears accounts for the greater part of the apparent angle of the reproduced sounds. If this is true, the angular location of each position on the virtual stage results from a particular loudness difference at the two ears produced by the speech coming from the loud speakers. When three channels are used a definite

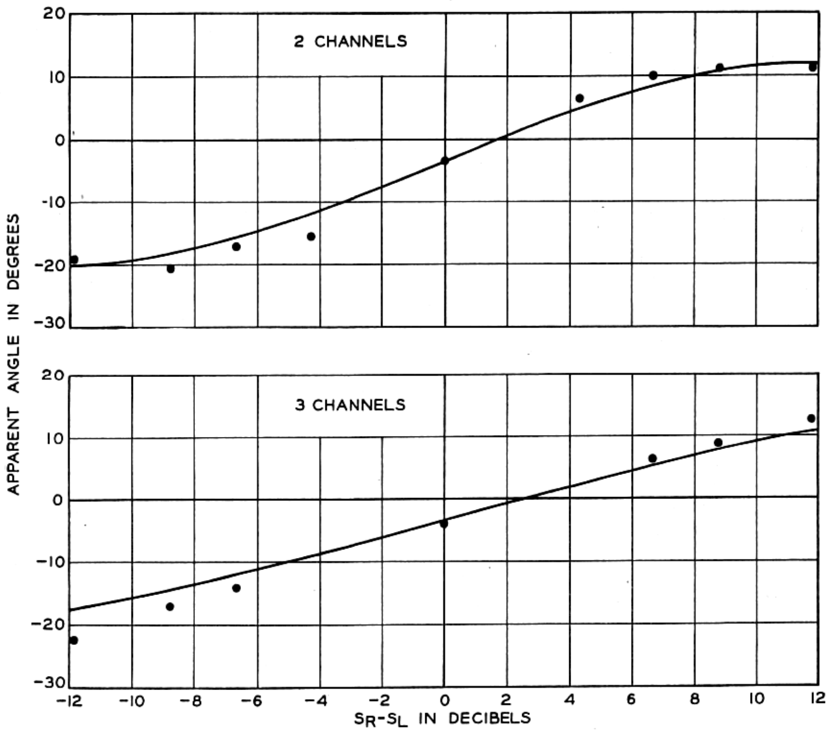

Fig. 5—Calculated and observed apparent angles for 2- and 3-channel reproduction.

set of loud speaker speech levels exists for each position on the pick-up stage. To create these same sets of loud speaker speech levels with the 3-microphone 3-loud speaker bridging arrangement already discussed, it would be necessary to change the bridging gains for each position on the pick-up stage. Hence it could not be expected that the arrangement as used (i.e., with fixed gains) would create a virtual stage identical with that created by 3-channel reproduction. However, with proper technique, bridging arrangements on a given number of channels can be made to give better reproduction than would be obtained with the channels alone.

Experimental Verification of Theory

Considerations of loudness difference indicate that all caller positions on the pick-up stage giving the same relative loud speaker outputs should be localized at the same virtual angle. The solid lines of Fig. 6 show a stage layout used to test this hypothesis with the 2-channel system. All points on each line have a constant ratio of distances to

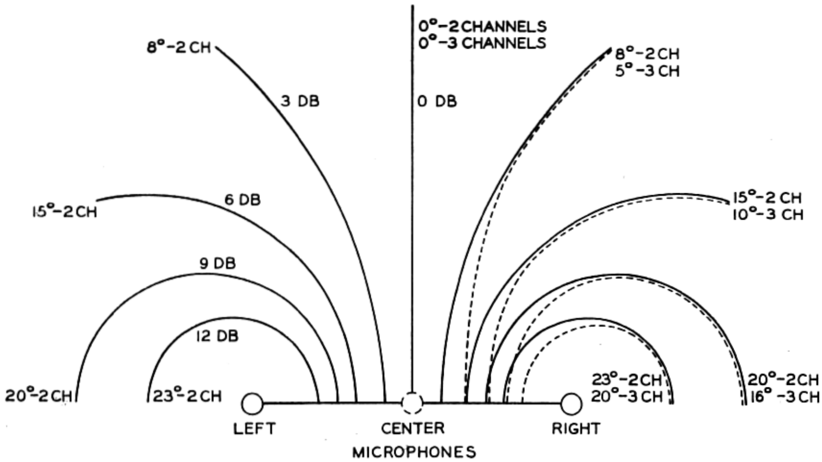

Fig. 6—Pick-up stage contour lines of constant apparent angle.

the microphones. The resulting direct sound differences in pressure expressed in decibels and the corresponding calculated apparent angles are indicated beside the curves. The apparent angles were calculated for an observing position on a line midway between the two loud speakers but at a distance from them equal to the separation between them. The microphones were turned face up at the height of the talker's lips to eliminate quality changes caused by changing incidence angle. It was found that a caller walking along one of these lines maintained a fairly constant virtual angle. For caller positions far from the microphones the observed angles were somewhat greater than those computed. For highly reverberant conditions, the tendency was toward greater calculated than observed angles. Reverberation also decreased the accuracy of localization.

A change of relative channel gain caused a change in virtual angle as would be expected from loudness difference considerations. For instance, if the caller actually walked the left 3-db line, he seemed to be on the 6-db line when the left channel gain was raised 3 db. Many of the effects of moving about the pick-up stage could be duplicated by volume control manipulation as the caller walked forward and backward on the center path. With a bridged center microphone substituted for the two side microphones similar effects were possible and, in addition, the caller by speaking close to the microphone could be brought to the front of the virtual stage.

For observing positions near the center of the auditorium the observed angles agreed reasonably well with calculations based only upon loudness differences. As the observer moved to one side, however, the virtual source shifted more rapidly toward the nearer loud speaker than was predicted by the computations. This was true of reproduction in the auditorium, both empty and with damping simulating an audience, and outdoors on the roof. Computations and experiment also show a change in apparent angle as the observer moves from front to rear, but its magnitude is smaller than the error of an individual localization observation. Consequently, observers in different parts of the auditorium localize given points on the pick-up stage at different virtual angles.

Because the levels at the three microphones are not independent, and because the desired contours depend upon the effects at the ears, a 3-channel stage is not as simple to lay out as a 2-channel stage. For a given observing position, however, a set of contour lines can be calculated. The dashed lines at the right of Fig. 6 show four contours thus calculated for the circuit condition of Fig. 1 and the observing position previously mentioned. The addition of the center channel reduces the virtual angle for any given position on the pick-up stage by reducing the resultant loudness difference at the ears. Although the 3-channel contours approach the 2-channel contours in shape at the back of the stage, a given contour results in a greater virtual angle for 2- than for 3-channel reproduction.

Similar effects were obtained experimentally. As in 2-channel reproduction, movements of the caller could be simulated by manipulation of the channel gains. From an observing standpoint the 3-channel system was found to have an important advantage over the 2-channel system in that the shift of the virtual position for side observing positions was smaller.

Effects of Quality

If the quality from the various loud speakers differs, the quality of sound is important to localization. When the 2-channel microphones were so arranged that one picked up direct sound and reverberation while the other picked up mostly reverberation, the virtual source was localized exactly in the "direct" loud speaker until the power from the "reverberant" loud speaker was from 8 to 10 db greater. In general, localization tends toward the channel giving most natural or "closeup" reproduction, and this effect can be used to aid the loudness differences in producing angular localization.

Principal Conclusions

The principal conclusions that have been drawn from these investigations may be summarized as follows:

1. Of the factors influencing angular localization, loudness difference of direct sound seems to play the most important part; for certain observing positions the effects can be predicted reasonably well from computations. When large quality differences exist between the loudspeaker outputs, the localization tends toward the more natural source. Reverberation appears to be of minor importance unless excessive.

2. Depth localization was found to vary with changes in loudness, the ratio of direct to reverberant sound, or both, and in a manner not found subject to computational treatment. The actual ratio of direct to reverberant sound, and the change in the ratio, both appeared to play a part in an observer's judgment of stage depth.

3. Observers in various parts of the auditorium localize a given source at different virtual positions, as is predicted by loudness computations. The virtual source shifts to the side of the stage as the observer moves toward the side of the auditorium. Although quantitative data have not been obtained, qualitative data on these effects indicate that the observed shift is considerably greater than that computed. Moving backward and forward in the auditorium appears to have only a small effect on the virtual position.

4. Because of these physical factors controlling auditory perspective, point-for-point correlation between pick-up stage and virtual stage positions is not obtained for 2- and 3-channel systems. However, with stage shapes based upon the ideas of Fig. 7, and with suitable use of quality and reverberation, good auditory perspective can be produced. Manipulation of circuit conditions probably can be used advantageously to heighten the illusions or to produce novel effects.

5. The 3-channel system proved definitely superior to the 2-channel by eliminating the recession of the center-stage positions and in reducing the differences in localization for various observing positions. For musical reproduction, the center channel can be used for independent control of soloist renditions. Although the bridged systems did not duplicate the performance of the physical third channel, it is believed that with suitably developed technique their use will improve 2-channel reproduction in many cases.

6. The application of acoustic perspective to orchestral reproduction in large auditoriums gives more satisfactory performance than probably would be suggested by the foregoing discussions. The instruments near the front are localized by every one near their correct positions.

In the ordinary orchestral arrangement, the rear instruments will be displaced in the reproduction depending upon the listener's position, but the important aspect is that every auditor hears differing sounds from differing places on the stage and is not particularly critical of the exact apparent positions of the sounds so long as he receives a spatial impression. Consequently 2-channel reproduction of orchestral music gives good satisfaction, and the difference between it and 3-channel reproduction for music probably is less than for speech reproduction or the reproduction of sounds from moving sources.

References

1. "Some Physical Factors Affecting the Illusion in Sound Motion Pictures," J. P. Maxfield. *Jour. Acous. Soc.*, July, 1931.
2. "Minimum Audible Sound Fields," L. J. Sivian and S. D. White. *Jour. Acous. Soc.*, April, 1933.