The Binaural Location of Complex Sounds

By R. V. L. HARTLEY and THORNTON C. FRY

NOTE: Much has been written on the subject of the binaural location of pure tones but the case of complex sounds has received little attention in recent literature. The purpose of the present paper is to bring the discussion of complex sounds abreast of that relating to pure tones. Those who wish to acquaint themselves with the work on pure tones will be interested in reading the theoretical work of the authors and the experimental studies carried out by G. W. Stewart and students working under his direction. This work has been reported in various papers, most of which have appeared during recent years in the *Physical Review* and the *Physikalische Zeitschrift*.

A resume of the present paper is given by the authors in their concluding paragraph.—*Editor.*

THE need of determining the location of enemy submarines and aeroplanes during the war brought into use practical methods for locating a sound source which depend upon differences between the sound waves reaching the two ears. This stimulated a general study of the phenomena involved in binaural sound location. The foundation for this study had already been laid in the work of Lord Rayleigh and others, who, following more or less in his footsteps, had accumulated a considerable amount of information of both theoretical and experimental sorts. Of this information almost all that was of a theoretical nature and a considerable portion of the experimental kind dealt only with the location of *pure* tones, the more complicated and in some respects more important problem of *complex* sounds being almost entirely neglected. Such advances as were made in the theoretical aspects of the problem during the war were subject to the same restriction so that even to-day no comprehensive theory has been advanced which adequately covers the problem of the location of such sounds as occur in every-day life, and in the practical applications of binaural methods. However, the results obtained with pure tones can be made to throw considerable light upon the problem, and it is primarily from this standpoint that the following discussion is written.

It may be well at the outset to review some of the outstanding differences between the observed phenomena in the two cases. The accuracy of location is much less for pure tones, as is also the sense of definiteness of the sound image. The location of pure tones is almost wholly binaural as is evidenced by the inability of persons deaf in one ear to locate such a tone. With complex sounds not only is the location by binaural effects more accurate and definite, but also the observer is not dependent on these alone. Persons who are deaf in one ear can locate familiar complex sounds almost as well as those with normal hearing.

Practically all theories of sound location start from the assumption that the listener subconsciously observes certain sound characteristics which depend upon the position of the source and forms a judgment of where the source must be by comparing these characteristics with information which he has stored up as a result of his past experience with cases in which the position of the source was known. In order to fix the position of the source he must assign to it three coördinates such as its distance and some two angles which define its direction. To do this he must be able to observe at least three independent properties of the sound which are functions of the position of the source. If fewer than three are available some difficulty in location is certain to arise. If more than three are available there is the possibility of a number of simultaneous independent determinations of the three coördinates.

If the sounds of every-day life were never distorted in transmission all of these determinations would yield the same set of coördinates and the only advantage which the listener would gain from the additional information available would lie in the fact that some one set might be peculiarly sensitive to slight differences in the position of the source, and therefore might lead to increased certainty on the part of the observer. Owing to reflection from the walls of buildings and the like, the sounds of every-day life seldom arrive undistorted, so that the observer must always be somewhat uncertain as to whether or not the coördinates of the sound source are actually those which he deduces from the properties of the sound wave as it reaches his ears. If enough properties are available to permit him to make two independent determinations he may use one of them to check the other, and if they agree he is justified in a feeling of increased certainty as to the accuracy of his judgment. The more independent determinations he can make the more checks he will be able to apply and consequently the more confident he will be.[1]

It should not be inferred, however, that it is only the sounds of the street which reach the observer in a distorted form. In a great many laboratory experiments the characteristics of the sounds have

[1] It is interesting to note in this connection that it is not surprising that an observer locates a complex tone with much greater certainty than a pure tone when we consider how rapidly the number of independent sets of data increases with increase in complexity of sound. We have already said that three independent properties are needed for the determination of the three coördinates of the source. Hence if only three are available, only one determination can be made and no checks are possible. On the other hand, if four are available, four groups of three each can be formed and therefore four separate determinations can be made. Similarly, 10 determinations can be made from 5 properties, 20 from 6, and 120 from 10.

been inconsistent, and in some cases they have not even corresponded to any actual source whatever. Under these circumstances, if an image is formed at all, some purely psychological factors must enter in. For pure tones it has been found possible to explain much of the experimental data obtained under circumstances such as this by assuming that the observer subconsciously judges one or more of the characteristics to be in error and applies such corrections as will make all of the data correspond to an actual source. As a criterion for determining which characteristics will be altered, it is assumed that, in general, those are chosen which require the smallest changes.

Let us now consider what characteristics are available for locating sounds of different kinds. A pure tone from a source at rest with respect to the observer has at any point only two physical characteristics which are subject to change with the position of the source. They are its amplitude and phase. Corresponding to each position of the source there is a particular amplitude and phase at each of the two ears so that a total of four properties—the loudness of the sound, the average phase, the difference in amplitude (which may conveniently be expressed as a ratio) and the difference in phase at the two ears—are available for determining the position of the source. It is inconceivable that the average phase can have anything to do with the location of the sound since it may be changed at will without altering the position of the source. The same remark applies to the loudness of the sound except in those instances where the observer is familiar with the source to such an extent as to know how loud it may be expected to be. Hence, if we restrict ourselves to the cases in which prejudicial information of this sort does not exist, we find that the observer has only two quantities from which he may deduce the position of the source. We should therefore expect that these two quantities would make it possible to locate the tone with respect to two coördinates only. This is found to be in general agreement with experiment, for most observers locate all sources of pure tones in the same horizontal plane with their heads and determine only the distance and angular departure from the median plane. If the source is more than a few yards away the intensity ratio and phase difference change very slowly with distance so that in this case even the sense of distance is not keen and a feeling of certainty exists with respect to the direction only.

In many experiments the tones at the two ears have been varied arbitrarily so as to give combinations having equal phases and unequal intensities or vice versa—combinations which cannot arise from actual physical sources in the absence of distortion. Under these conditions the observer generally corrects one to a value consistent with the other except in extreme cases where the correction required for this purpose would be inordinately large. When this occurs he may either assume both to be correct and form two images—one based on the phase difference together with a mentally supplied intensity ratio consistent with it, and the other similarly derived from the observed intensity ratio—or he may fail to have a sense of location at all.

Before considering the available characteristics of complex sounds in general let us confine our attention for a time to those which are made up of a limited number of sustained pure tones such as an organ note with its series of overtones, or a group of tuning forks. Here the number of characteristics increases rapidly with the number of component tones. For each component tone there are two quantities: intensity ratio and phase difference. In addition, at either ear alone the relative intensities of any two of the tones changes with the position of the source, owing to the diffraction of the sound waves around the head being different for different frequencies. There are therefore as many of these observable intensity ratios as there are pairs of components. Similarly, for any two tones whose frequencies are commensurable, the relative phases of the two at the same ear depend upon the position of the source.

Not all of these characteristics are capable of contributing to binaural as distinct from monaural location. In fact, only the phase differences and intensity ratios of the separate components are binaural. A man who is deaf in one ear has available all of the relations between the intensities and phases of the various components at his normal ear. That these relations do actually contribute to sound location is supported by experimental evidence. Myers [2] found that, after familiarizing himself with a complex sound, a blindfolded observer could locate its position with considerable accuracy, even when it was moved about in the median plane, but that his accuracy could be destroyed by varying the relative intensities of the components.[3] It is not surprising then, that for complex sounds the accuracy is about the same whether the location is binaural or monaural.[4] The observed failure of monaural location in the case of a

[2] C. S. Myers, *Proc. Royal Soc.*, 1914, B 88,267.

[3] It should be noticed that this effect must have been purely psychological since it could be produced without moving the source at all. It therefore lends plausibility to the assumption upon which our theory is based: that when discordant or unusual stimuli are experienced, a mental readjustment of the stimuli is made in order to render them more nearly consistent with every-day experience.

[4] As shown by the experiments of Angell and Fite upon persons deaf in one ear. *Psychol. Rev.*, vol. 8, pp. 225–246, 1911.

pure tone follows directly from the absence of other frequencies with which the pure tone may be compared.

As we are here concerned with binaural phenomena we shall confine our attention to the relative phases and intensities at the two ears. The question at once arises: does the observer actually hear the different tones separately, and if so, does he assign a location to each separately?

To what extent the listener locates each component separately depends upon the ease with which the tones can be distinguished. The experiments which bear most directly upon this point are those in which the component tones at the two ears are arbitrarily adjusted to give values of phase difference corresponding to different locations. This is done under conditions where the location of each component separately is largely determined by the phase difference. More [5] experimented with two tones, transmitting them to the ears through tubes of adjustable lengths. This permitted him to change the phase difference at the two ears while keeping the intensities substantially equal. He observed the apparent location for various settings when each tone was applied by itself and when both were applied together, using forks of 256 and 320 cycles. With the paths equal the tones combined into a chord located in the median plane and the separate components could not be heard. With a setting for which the two components separately appeared on opposite sides of the head, one component was heard distinctly by the right ear only on the right side, and the other by the left ear only on the left side. At the same time the chord was heard rather indistinctly near the median plane but tending slightly toward the side of the lower tone.

Apparently the observer does not consciously separate the chord into its components unless he is forced to do so by some inordinate discrepancy between the positions of the images formed from them. There is no evidence in the case of equal paths to show that he did or did not subsconsciously locate the separate components and find them to be in agreement. In view of the second experiment it seems probable that he did. In this latter experiment he obviously found that the two components corresponded to different locations and assigned different sources to each. At the same time his experience told him that tones which would combine to form a musical sound generally have a common source. Hence he may have concluded subconsciously that the sound waves had probably been distorted in coming from a common source and so he corrected his observations

[5] Louis T. More: *Phil. Mag.* XVIII, 1909, p. 308.

on both tones to make them consistent and arrived at an image of the chord between the other two.

Similar results were obtained with forks of 256 and 384 cycles per second, except that in general the lower tone was completely blotted out. The higher tone was usually quite distinct and definitely located. The image of the chord was nearer to the image formed when the higher component was sounded by itself than to the image formed from the lower one alone. With settings for which the directions of the tones separately were the same, whether right, left, or middle, the upper tone disappeared leaving only the chord. In experiments with forks of 256 and 512 cycles it was difficult to distinguish the separate notes. With settings for which the two separately were on opposite sides the combination was on the side of the lower fork. This can be interpreted as meaning that the octave relationship is inherently difficult to resolve, or else that tones an octave apart so generally come from a common source that the observer was unwilling to make any other assumption.

Although the explanation of these results is not yet thoroughly understood, they show very definitely that in locating complex sounds made up of pure tones the observer does within limits locate the components separately. If they agree, a single image is formed; if they do not, he may either locate the tones separately or form a single compromise image or do both.

It is in this way that the theory developed for pure tones is applied to complex sounds made up of pure tones. The next step is to extend it so as to include complex sounds in general. To do this we must picture the observer as resolving each sound into sinusoidal components locating the components separately and forming one or more images based on a combination of the apparent sources as indicated by the separate components. While it is fairly easy to effect such a resolution mathematically it is somewhat less easy to interpret the result in a manner satisfactory to our intuitive conceptions of the phenomena involved; also, granted the theoretical possibility of the resolution, there remains the question of what physical or psychological limitations there may be to its application.

In view of the fact that a really pure component tone has no beginning or end, and no fluctuations in its amplitude, it is not at once apparent how a single discrete sound such as the bark of a dog can be resolved into components of that nature. However, if enough components are available it has been established beyond question that by properly choosing their frequencies, amplitudes, and phases, a combination may be arrived at in which the algebraic sum of all the components is zero for all instants before and after the period occupied by the sound and equal to the instantaneous value of the sound wave for instants within that period. This combination is known to mathematicians as the Fourier Integral corresponding to the wave, and the formula for the phase and amplitude of each component sinusoid is known. It is an extension of the well known Fourier series expansion used for resolving sustained periodic disturbances.

The physical interpretation of this integral may be facilitated by reviewing the steps in its evolution from the Fourier series. It is well known that if the sound in question were repeated at regular intervals the resulting periodic wave could be resolved by Fourier analysis into a series of sinusoidal components, the frequencies of all of which are integral multiples of the frequency of repetition of the sound. Successive components therefore differ in frequency by an amount equal to this frequency of repetition. Now it is not essential that the repetitions of the sound follow each other immediately. Instead, they may be separated by intervals of silence. The effect of such silent intervals is to reduce the frequency of repetition and therefore also the fundamental frequency. As a result the component frequencies are brought closer together and the number within any particular frequency range is increased.

Suppose now that the interval between repetitions is indefinitely increased. As this is done the effect of any one occurrence of the sound becomes more and more independent of the others, and in the limit when the sounds next preceding and next following the one under consideration are infinitely far removed, we have the case of a discrete sound. As this limiting case is approached the fundamental frequency becomes smaller and smaller and the component frequencies, which are multiples of it, are separated by infinitesimal frequency differences. While the amplitude of each component also decreases, the number of components increases at such a rate that the aggregate energy of all the components within a given frequency range remains finite. In this way, the distribution of the sound energy over various frequencies—that is, the "energy spectrum"—can be obtained.

It is evident, then, that when an aperiodic complex sound is resolved mathematically there results an infinity of component tones, each having a characteristic intensity and phase. If an observer were capable of an equally complete resolution he would have at his disposal an infinity of sets of data from which an infinity of images could be formed. In the absence of distortion these should all coincide.

Practically, of course, no such refinement of resolution is possible. The ability to distinguish differences in pitch varies from person to person, but the minimum intervals employed in musical composition probably give a rough measure of the normal resolving power of the ear. Even with this limitation the broad sound spectrum, such as an irregular sound produces, is capable of yielding a very large number of separable components; and hence a large number of individual images. It is this fact—that with a very complex sound the number of independent determinations of the image is limited only by the resolving power of the observer—which makes his accuracy of binaural location as well as his sense of certainty much greater for such sounds than for pure tones.

So long as the images of all the components coincide, it is of little importance how fine the resolution is, for further refinement only serves to increase the sense of certainty by adding to the volume of accordant evidence. However, when the images are not in agreement the problem is more complicated and the degree of resolution becomes important. Here also purely physical considerations cease to be adequate and psychological factors must be considered similar to those involved in the location of a pure tone for which the intensity ratio and phase difference do not correspond to any actual source. When an observer is faced with discordant results he must make some subconscious judgment. For small discrepancies such as occur in every-day experience, he probably assumes those images which depart most from the rest to be misplaced because of distortion during transmission and so either corrects or ignores them. If the discrepancies are large he may find it difficult on the ground of experience to believe that so much distortion could occur. In such an event he will most likely form several images from different components or in extreme cases lose the sense of location altogether.

Bowlker found separate images to occur experimentally both for band music, which approaches a collection of tones and for the irregular barking of dogs. He placed tubes of unequal length to his two ears thereby upsetting the normal diffraction around the head and interposing a longer path on one side than on the other. Obviously, the distortion produced in this manner is of a type not likely to be met in every-day life and affects different frequencies in widely different fashions. He reports that when listening to "a band of three or four instruments played in the open—the notes will be found to be scattered over a wide range, most being to the side of the short tube, some being in front and some being to the side of the long tube. In listening with such a pair of tubes to two dogs furiously barking the effect is at first quite alarming—one seems to be in the middle of a pack of dogs some of which are rushing viciously at one's throat."

An illustration of failure to form any image is found in a phenomenon observed in the use of binaural compensators for determining the direction of submarine sounds. The sound is picked up by two submarine telephone transmitters and led to the ears through independent paths. By adjusting the lengths of the paths the image can be shifted from side to side and for practical purposes the setting of the instrument is made by bringing the image exactly to the middle. A fairly definite sound image is formed, but observers report that part of the sound does not merge into this sound image and move in response to the adjustment, but instead appears as a diffuse background of noise.[6] This may be explained on the assumption that, while the images formed from most of the sound components agree sufficiently well that the observer corrects them to a single position, certain components are so distorted by resonance effects inherent in the apparatus that their images are scattered more or less at random. The lack of agreement among any considerable number of these prevents the formation of a second image and causes the sense of diffusedness.

As the distortion becomes still more extreme we should expect the experimental results to depend more and more upon the observer's power of resolution, for as the distortion is progressively increased a condition must finally be reached where the positions of the images are appreciably different for two components whose frequencies are so nearly alike as to make their recognition as separate tones difficult if not impossible. This condition actually occurred in an experiment of Baley's with a sound consisting of a mixture of sustained tones. Its effect on the listener is interesting from the standpoint of subconscious readjustment of discordant data.

Baley's [7] experiment consisted in applying a number of sustained tones to one ear of a musically trained observer and a number of different tones to the other ear, and testing his ability to assign them to their proper sides. So long as the intervals between the tones were fairly large, the observer never failed to locate them correctly. Considering the entire stimulus as a complex sound we may think of the observer as locating the tones individually and finding them to fall definitely into two groups whose images are located one at each ear.

[6] This interesting phenomenon was called to our attention by Mr. Richard D. Fay of the Submarine Signalling Corporation who tells us that it has been noted by a large number of observers.

[7] Stephan Baley: *Zeit. f. Psychol. u. Physiol.*, v. 70, 1914, p. 347.

However, when he used six tones which were separated from each other by a single tone interval, the separate components could not be distinguished and a painful sensation was produced. The observer was apparently faced with the situation that to make the observed intensity ratios and phase differences correspond to a single source would involve extremely large corrections in the observed data. On the other hand, his power of tone resolution was insufficient to separate the components and assign them to different sources. It is not surprising, then, that the difficulty manifested itself by painful sensations. While this illustration is taken from an extreme condition of laboratory experiment and may appear to have little bearing on the every-day location of sounds, it is really significant because of the manner in which it illustrates the importance of psychological factors in all cases in which the sound waves are distorted.

Resumé

In the foregoing discussion an attempt has been made to bring out the main features involved in extending the theory of the binaural location of pure tones to cover, qualitatively at least, the location of complex sounds. It has virtually been assumed that the latter involves three processes: first, the resolution of the sound into its component tones; second, the independent (generally subsconscious) location of each separate component; and third, the formation of a conscious judgment of the position of the source based on the locations of the individual images. The greatly increased amount of data available when the sound is complex has quite different effects on the final result according as the different images do or do not coincide. If they do, the accuracy of location and the sense of certainty are increased. If they do not, confusion arises, subconscious corrections are called for, and the final result is likely to depend very considerably on the psychological processes and individual prejudices of the particular observer.