The "Objective" and Subjective View in Music Transcription
Author(s): Nazir A. Jairazbhoy
Reviewed work(s):
Source: *Ethnomusicology*, Vol. 21, No. 2 (May, 1977), pp. 263-273
Published by: University of Illinois Press on behalf of Society for Ethnomusicology
Stable URL: http://www.jstor.org/stable/850947
Accessed: 19/07/2012 11:25

THE "OBJECTIVE" AND SUBJECTIVE VIEW IN MUSIC TRANSCRIPTION

Nazir A. Jairazbhoy

"Ethnomusicology could never have grown into an independent science if the gramophone had not been invented" (Kunst 1959:12). Prior to the invention of the phonograph, scholars of non-Western music were obliged to transcribe music by ear in the field; the resulting notations, "however well-intended, usually fell short in every respect–i.e., both rhythmically and as regards pitch" (ibid.). Early scholars often transcribed music during festivals and rites through all the accompanying confusion and without the benefit of repeats.[1] The many difficulties encountered in the transcription of unknown traditional musics, and the fact that conventional methods of transcription failed to convey the style and manner of music performance, led some to attempt to learn the songs and to use their own renditions as illustrations of the exotic music traditions–sometimes with very strange results. Nettl (1964:100) refers to the apocryphal story of the anthropologist who learned an Australian aboriginal song and practiced it on the long voyage back to Europe, making such changes in it (presumably unconsciously) that it was transformed into a German folk song.

In addition to providing a check on field transcriptions, the phonograph obviously added new dimensions to ethnomusicological studies. Kunst (op. cit.) draws attention to one of these dimensions, musical style, which could now begin to be investigated in comparative terms. The early cylinder, disc, and film[2] recorders naturally suffered from poor fidelity and often from the perishable nature of the raw stock, which could not stand the strain of many repeats. Transcribing from these early recording devices can hardly have been an easy task.

With the invention of magnetic wire and tape recorders, the discipline of ethnomusicology received a further stimulus in the 1950s. These machines were compact, light, portable, and many were even within the financial reach of casual tourists, with the result that the music of many previously unknown traditions became available in the West. For the transcriber, the advantages of tape over disc or cylinder are too obvious to need enumeration. However, while tape recorders have continued to improve in every respect and significant collections of recorded materials are in existence, there has been no major breakthrough during the past twenty years as far as the transcriber is concerned.

Great hope was (and often still is) placed on automatic transcription devices which, it was thought, would perhaps replace the subjective (and therefore fallible) ear of the aural transcriber. After the exciting initial efforts by Dahlback (1958) and the Seeger Melograph, culminating in Model C at the end of the last decade, progress in these directions seems to have come to a halt—not because the final goal had been reached, but because of the enormous complexity of music. The Melograph could only cope with a single melodic line, a phenomenon which is not too common in the natural course of events. Even so, it presented a profusion of visual data, involving tolerances much finer than the ear can distinguish, thus creating a new series of problems for the interpreter.[3] If an automatic transcriber is ever able to produce graphic displays of a full orchestra, it is evident that the problems of interpretation of the displays will be infinitely more complex.

Some of the basic premises underlying the development and use of automatic transcribers appear to have been inadequately explored. Hood (1971:21) refers to the objective display produced by automatic notation devices, while Bartók (1951:3) comments, "The only true notations are the sound tracks on the record itself." The terms "objective" and "true" surely need qualification. How true to the original was the sound that Bartók heard on the records of his period? Even with perfect electronic recording and play-back equipment, we would still need to question the validity of the above terms, since the precise content of the sounds of an individual instrument or an orchestra depends on the number of factors, such as the location of the listener or microphone(s), the acoustics of the room, etc.[4] When music is recorded, coloration is inevitably added, no matter how good the recording equipment, the acoustics of the room or the sensitivity of the recording technician, whose subjectivity is evident in the positioning of the musicians, selection and placement of microphones, and so on. There is no single optimum in the recording process, just as there is no single perfect seat in the auditorium. The performing musician himself hears his own music in quite a unique way, which is very different from the way a member of the audience hears it.[5] This is obvious when he is part of an ensemble, but a soloist also has a subjective impression of the sound of his instrument, partly because he is involved in the process of creation, but also because the sound content reaching his own ears differs from that which reaches the ears of a member of the audience. An instrumentalist first hearing a recording of his own instrument is often surprised at the strange tone quality, just as a speaker is at first surprised to hear his own recorded voice.

Thus it should be apparent that if a recording is transcribed by an automatic device, it may make a more or less "objective" statement (within its own technical limitations[6]), but only about a subjective recorded impression of an exceedingly complex original.

There also seems to have been no thorough examination of the relevance of this "objective" look at sound to ethnomusicological studies. It has been conclusively established that, owing to a non-linearity lying somewhere between the excitation of the ear by sound and the sensation of sound in the brain, harmonic and intermodulation "distortion" is added to the incoming sound signals,[7] two of the types of distortion which we are particularly careful to avoid when purchasing high-fidelity equipment. This means that we can never hear the "objective" sound of the automatic transcriber and that all our music has been structured on the basis of what we hear, the subjective and distorted rather than the "objective." Studies involving simple sine waves show that, when two tones are received simultaneously, tones of the harmonic series as well as combination tones are added by the listener.[8] It is indeed difficult to imagine just what the ear registers from a complex orchestral sound source–which tones are added, which amplified, and which are negated or masked.[9] Experiments to determine just this are now being conducted in various physics and psychoacoustic laboratories, and in the future, it may be possible for computers to add the coloration/distortion produced by the "average" ear of, for instance, a 25 year old Caucasian male. In other words, it might be more meaningful for ethnomusicological studies if the automatic transcriber were more "subjective";[10] but this is much more difficult to accomplish.

In addition to the harmonic and intermodulation distortion added by the ear, there are also other irregular responses which have an important bearing on our perception of sound. Fletcher and Munson (1933) discovered that the threshold of human hearing varies with the frequency of musical tones. Thus, if a 1000 Hz tone becomes audible at a level to be arbitrarily called zero dB, then a 30 Hz tone will become audible at 65 dB and a 10 KHz tone at nearly 10 dB (Burroughs, 1974:150). In lay language, this means that low and high tones need to be produced at greater amplitude than middle tones in order to be heard. What makes this much more complicated is that, as overall amplitude increases, the ear's sensitivity to different frequencies also changes and becomes more uniform. These figures once again relate to the average listener; each individual will show deviations, age being one of the most important factors involved.[11] It is interesting to note that modern amplifiers often have loudness contour switches to boost bass and treble frequencies at low listening levels to compensate for this aural phenomenon in a very general way.

An automatic transcriber can give the spectral content of a musical tone and may well record the presence of certain frequencies which we are unable to hear. The interpreter has the unenviable task of determining which of the frequencies are meaningful. Bergeijk (1958:101) comments, "It is humanly impossible to hear a very complicated sound as a collection of separate frequency components"; it is equally impossible for the reader to reconstruct such a spectral display in terms of audible sound without the aid of something like an automatic transcription reader–an instrument which has not yet been devised.

There is also a difference between the display of frequencies by an automatic transcriber and the pitches heard by the ear. The aural perception of frequency (i.e. pitch) is subject not only to the frequency of a tone, but also to its intensity and, to some extent, the complexity of the wave form.[12] Thus, a steady tone of 100 Hz will sound about 10% lower in pitch if its intensity is increased from 40 to 100 dB. However, the ear's response to this increase in intensity is not constant over the frequency range, and the pitch of a tone of 500 Hz is only lowered about 2% with this same increase in intensity (Tremaine 1974:20). Automatic transcribers generally measure the periodicity of the fundamental, removing the partials by means of filters. However, when the fundamental of a tone is removed or is very weak, the aural sense of the pitch of this fundamental still remains (Bergeijk 1958:133), but of course is not registered by the automatic transcriber. Wood (1962:86-87) describes two sets of phonograph records produced by Bell Telephone Company, identical except that in one set all the frequencies below a certain limit had been filtered out. The two sets apparently sounded very much the same, even though all the fundamental tones were missing in one. Presumably, between the ear and the brain, the missing fundamentals were added in the form of difference tones.

There are other possible physical interferences and interactions of which we know even less. Do the involuntary bodily functions, e.g., heart beat and respiration, have any effect on our hearing? What about the metabolic rate? Do the shapes of the resonating chambers in a human body have any bearing on what is heard? What about the phenomenon, described by Seashore (1967:103) as "sonance," where one apparently hears an average pitch, intensity, or timbre when successive waves are too rapid to be distinguished individually, as in the case of vibrato? In the present state of our knowledge, with so many unknowns, there appears to be no way in which the analyst could deduce what would actually be heard from the transcriptions of an automatic music transcriber.

In the above discussions, we have attempted to focus on the physiological aspects of hearing. It is clear, however, that once this auditory information reaches the brain, it is processed in some extremely complex manner, involving factors such as memory and conditioning. "To a degree people hear–and people see, for that matter–what they expect to hear and see" (Bergeijk 1958:70). One of the functions involved is the process of selection, which is an aspect of conditioning beginning very early in an individual's life. As a child learns to recognize the sound of his mother and other sounds meaningful to his own existence, he is also learning to ignore extraneous sounds. Many of us are accustomed to and take for granted a background level of about 40 dB, of which we seldom become consciously aware.[13] This selective faculty of the mind seems to be much more sophisticated than any electronic filtering device, since it is not limited to particular frequencies and can even function under conditions of great amplitude. An obvious example of this is an involved music listener who is completely unaware of a jet aircraft flying overhead, or one who can ignore the scratch and rumble of an old, poorly maintained disc. These two examples show the mind's ability to differentiate between musical and non-musical sounds. There is good reason to believe that selectivity is also practiced within the musical sound range. Listeners unfamiliar with Western music, for instance, have genuine difficulty "hearing" vertically, just as some Western trained listeners have difficulty hearing past a drone.[14] Most musicologists have probably experienced instances of unconscious selection, as for instance, when "hearing" (actually becoming aware of) something new in an often-heard piece of music. Even an identically repeated musical pattern need not necessarily sound monotonous if the listener's awareness shifts from one musical feature to another. Indeed, some African rhythms (described as inherent rhythms by Kubik, 1962) seem to be constructed in a manner particularly conducive to shifts in the alignment of the rhythmic image perceived by the listener.

Sounds which an individual selects to "hear" are also obviously related to the musical conventions of a particular genre. An analogy from the sphere of dance illustrates this point rather obviously. In the North Indian classical dance, Kathak, the dancer, having completed an intricate rhythmic movement, might retire to the rear of the stage, often in a rather perfunctory manner. Within the conventions of this tradition, it is understood that the rear of the stage is a kind of private zone, as though it were out of sight of the audience. Here the dancer may slouch, wipe his brow, or converse with the musicians while he prepares for the next series of movements. Although the accompanying music continues through this "break," the attention of the audience is, for that moment, diverted off the stage. Virtually the same phenomenon is also found in North Indian classical music. In the vocal form *khyal,* the musical phrase *mukhra* generally indicates the completion of a musical idea.[15] If he should choose to, the musician may resort to the conventionally accepted private zone, where he may clear his throat, shift position, or converse with his accompanying musicians until he is ready to resume again. This kind of behavior would of course be more evident in a private "sitting" than in a formal concert. The continuity of attention to which we are accustomed in listening to Western art music does not apply here.[16]

These examples illustrate a faculty of the mind to select and focus on portions of the total sound stimulus registered by the ear. Needless to say, the automatic transcriber practices no such discrimination and it is the analyst who is obliged to use his subjective judgment both in choosing the examples to be analysed and in their interpretation. When dealing with a music tradition whose conventions are not known, it may be that the transcriptions need to be as comprehensive as possible, e.g., those of Bartók and Lord (1951). The basic problem facing the analyst, however, is to convert the "objective" record of the physical aspects of music as performed into terms of psychological experiences and responses which have musical meaning. This, according to Seashore (1967:26), involves "a vast array of principles, such as the psychophysics of hearing, musical evolution, musical knowledge and training, the limits of the organism, individual differences, health, musical environment, musical guidance, practical norms and esthetic principles." The ethnomusicologist could easily add a number of other equally important factors.

In some respects an automatic transcription can be quite deceptive because it does not take into account the adaptability of the human mind. For instance, when one first walks into a night club where a loud rock band is performing, the sound levels may seem unbearable. After a few moments or minutes, the mind adjusts to this intensity and the music does not seem quite so loud. Such an experience could well produce auditory fatigue which could affect the threshold of hearing for several hours (Bergeijk 1958:95). Similarly, a very quiet musical passage following a loud one sounds even quieter until the mind has made the necessary adjustments. There is evidently some form of volume control built into the hearing system which can only be experienced on a subjective level. However, the operation of this control is not instantaneous and is probably parallel to the kind of adjustment made by the iris of the eye as one moves from very bright light to darkness.

The mind also seems to have the ability to adapt to strange timbres and unusual musical intervals in the same way. At first exposure, the latter may sound out of tune, but with repeated hearing they begin to sound quite normal within that context. The "thin" tone of old phonograph records may be disconcerting at first, but this feeling does not last very long. Thus "objective" measurements in dB or cents give no real indication of what one experiences in any specific context.

The automatic transcription also takes no account of the extraordinary constructive abilities of the human brain. When listening to a familiar music tradition, our "hearing" is influenced by very specific memories, as well as musical syntax, both of which induce feelings of anticipation, frustration, suspense, and resolution. These feelings result from an interaction between what we expect to hear and what we actually do hear. How real is what we expect to hear? Is the mind actually able to create the impression of musical sounds based on memory and syntax, sounds which are derived from, but not actually present in the external auditory stimuli? Bergeijk (1958:69) describes an experiment which seems to indicate that this is indeed so. A group of subjects was familiarized with a piece of choral music through repeated hearings. The music was then played to each of them against a loud background of white noise. The volume of the music was then gradually lowered to zero level and the subjects were asked to indicate the precise moment when the music had disappeared. All the subjects heard the music continue for a considerable time after it had actually been turned off.[17]

The reader can easily test the driving power of memory and rhythm by singing a familiar song and stopping suddenly short of the end. It is not surprising that this mental faculty is exploited in some musics, for instance, in the North Indian *tihai* (a thrice repeated rhythmic figure) or the cadential phrase, *mukhra,* which is sometimes left incomplete to be concluded by the mind of the listener.

The purport of the above discussions is to show that there is, and perhaps always will be, a large gap between what an automatic transcriber would "hear" and what an experienced listener of a particular musical idiom might "hear."[18] Perhaps automatic transcribing devices can be adapted to accommodate the physical characteristics of our hearing system; it will, however, require a great deal of experimentation by psychoacousticians on a cross-cultural level before we can train our automatic devices to hear from a cultural point of view.

Seeger (1958:187-88) recognized that there must be a subjective element in music as a means of communication and recommended that, for the foreseeable future, both notation and the automatically transcribed graph should be employed concurrently. Yet, in advancing the value of the Melograph, he was perhaps inordinately critical of aural transcription, particularly that based on Western notation:

> In employing this mainly prescriptive notation as a descriptive sound-writing of any music other than the Occidental fine or popular arts, we do two things, both thoroughly unscientific. First we single out what appears to us to be structures in the other music that resemble structures familiar to us in the notation of the Occidental art and write these down, ignoring everything else for which we have no symbols. Second, we expect the resulting notation to be read by people who *do not carry the tradition of the other music.* The result, as read, can only be a conglomeration of structures part European, part non-European, connected by a movement 100% European. To such a riot of subjectivity it is presumptuous indeed to ascribe the designation "scientific" (Seeger 1958:186-87).

Much of this is less meaningful today than it was nearly twenty years ago when it was written. Students of ethnomusicology are delving so deeply into the music of other cultures that many are now virtually bi-musical, i.e., they both think and perform very much like the indigenous musicians. They are beginning to approach one of the goals of ethnomusicology—the ability to identify not only those structures "familiar to us in the notation of the Occidental art" (which is in itself a legitimate aspect of the field), but to identify and be able to isolate structures recognized by the native musicians themselves. There are many instances in the literature of our field to show that our scholars are not "ignoring those elements for which we in the West have no symbols," but creating new symbols where it seems necessary.[19] Seeger's second point also has less bearing now, since, with the availability of a variety of recordings and the growing number of scholars and students acquiring expertise in different musical traditions, an increasing number of readers do, in fact, carry a great deal of the tradition being transcribed.

Equally, in view of the previous discussions, we may be justified to question the value of the "objective" or external view of the automatic transcriber and even the scientific validity of a machine whose results cannot be verified, since no device exists which can reconvert the graphs into sound. Descriptive music writing seems, in any case, to have an ambivalent nature; to be comprehended by the reader, it must be treated as a prescriptive transcription, and since musicologists have far less experience converting such complicated graphs into sound, it would seem to be unlikely that these graphs would be "heard" with any more consistency among musicologists than are aural transcriptions.

Thus it will be clear that an automatic transcription should not be thought of as a replacement for aural transcription. They perform different but equally justifiable functions. The primary value of automatic transcriptions would be to throw light on what we do *not* "hear," what we change in the process of "hearing," or what we take for granted. They can also provide an insight into some of the extremely subtle elements of music which we cannot readily distinguish aurally, but which might nevertheless influence our perception of the music on a subconscious plane. They can provide us with acoustical standards against which we could compare the effects of auditory and cultural responses. There is little doubt that automatic transcriptions, with their detailed, external view of music, will eventually help us to understand some of the physiological and cultural processes of man. However, when the subject of study is concerned with the psychological or communicational aspects of music within a culture, aural transcriptions by a trained ethnomusicologist who has steeped himself in that culture may well be far more meaningful.[20]

NOTES

1. Helen Roberts (1931:111), however, points out some of the advantages of "longhand" transcriptions made in the field.

2. One of the innovations of the 1930s was a recorder produced in Germany called the Tefiphon. It cut grooves along the length of a continuous band of 35 mm film. The bands varied in length from 2 m. to 20 m., the latter permitting continuous

recordings of more than an hour. A machine of this type was used by Dr. A. A. Bake for his field work in India during 1937-38. The quality of the recordings was perhaps better than would be expected; however, the film tended to become brittle over the years and consistent tracking in the grooves was often a serious problem. Nevertheless, tape copies of this valuable collection were made in 1968 by A. A. Dick at the School of Oriental and African Studies, University of London, where the originals and a set of copies are housed. A second set is in the UCLA Ethnomusicology Archives.

3. A comparison of automatic and aural transcription of single melody lines is found in George List (1974).

4. Hood (1971:117) draws attention to the work of Vern Knudsen, which indicates that even elements such as temperature and humidity are factors in the registration of loudness.

5. Hood obviously does not subscribe to this view since he describes a series of recording tests of koto music in which he asked the performer "to choose the recorded sound that, in his judgment, came closest to reality" (1971:263).

6. For example, Charles Seeger states in his notes accompanying the Melograph of the Thai lullaby analyzed in List (1974:373), that the visibility of accuracy of the stylus line used is gauged at 20 cents.

7. There is some disagreement among acousticians as to whether the harmonic and intermodulation distortion (i.e., the addition of difference and summation tones) are caused by the asymmetry of the ear-drum or by some other feature in the auditory tract. In any case, the phenomena can easily be demonstrated as follows: a frequency or signal generator (which theoretically produces pure sine waves and has no overtones) tuned, for instance, to 200 Hz is heard to produce beats when a second tone of slightly more or less than 400 Hz is sounded against it. Since there is no tone of 400 Hz in the external sound spectrum (this can be checked with an oscilloscope) against which beats could be produced, this tone must be produced in the process of audition. The demonstration of intermodulation distortion is perhaps even more convincing. Two tones which are harmonically unrelated, for instance, one of 200 Hz and the other of 330 Hz, will produce beats when heard against a third tone of slightly more or less than 130 Hz, the frequency difference between the two tones. Similarly, beats will also be heard against a tone of slightly more or less than 530 Hz, the sum of the two tones. Since the difference and summation tones are not to be seen on an oscilloscope, it is deduced that they too are added by the hearing system.

8. The tones created by the ear are discussed in some detail by Jeans (1968:231ff).

9. Masking is described by Josephs (1967:45-47). In brief, it is the phenomenon of a weak sound becoming inaudible in the presence of a stronger sound. It is a characteristic of the human hearing system which is not registered by electronic devices. The degree of masking depends on both the frequency and amplitude of the masked and masking tones. Bergeijk (1958:93) gives an example, originally taken from the work of Harvey Fletcher, which indicates just how difficult it would be to estimate the results of this phenomenon when hearing an orchestra. If three tones of 400, 300, and 2,000 Hz are produced simultaneously at 50, 10, and 10 dB, respectively, the 300 Hz tone would be masked, i.e., only the 400 and the 2,000 Hz tones would be heard. If, however, the loudness of all three tones were to be increased by 30 dB, then it would be the 2,000 Hz tone which would be masked, rather than the 300 Hz tone.

10. Feld (1976:311) draws attention to a similar view in the field of visual communication held by De Heusch, who argues that "inasmuch as film objectivity and truth are illusory, the best ethnographic approach is one taking advantage of the sophisticated subjectivity of the film-maker/ethnographer."

11. The deterioration in hearing due to age, referred to as presbycousis, is specially significant in the higher frequency ranges and may amount to as much as a 60 dB loss at frequencies above 8,000 Hz for the age group between 60 and 75 (Seashore 1967:80). However, high frequency loss due to age is smaller for women than men, but low frequency loss is greater (Tremaine 1974:25).

12. There are some complex sounds which have no definable frequency but give the impression of having pitch.

13. The average dwelling is said to have a noise level of about 40 dB above the threshold of hearing (Tremaine 1974:11).

14. It is interesting to speculate whether a culturally conditioned listener of Indian music hears a drone continuously, intermittently, or perhaps not at all after the first few moments.

15. In North Indian instrumental music, the *mohra* serves virtually the same function.

16. It may be that the Indian listener is able to maintain his level of interest for extremely long periods precisely because he is given periodic respite through much of a performance.

17. Indian sitar and surbahar players sometimes appear to exploit this mental capacity of the listener as they manipulate the dying sounds in lengthy string deflections which, to the uninvolved observer, may seem to go somewhat beyond auditory experience.

18. The difference between the two is similar to that distinguished in linguistic theory between "phonetic" and "phonemic" sounds. These are described by Nettl (1964:104) as "the former being the study of speech sounds as they occur and the latter being concerned with those distinctions among speech sounds which produce, in a given language, distinctions in meaning." Since the terms phonetic and phonemic have special reference to speech, it might be useful to coin new terms, for instance, "audetic" and "audemic," to refer to parallel concepts in the music or audio range.

19. For example, *sori* (♭̸) and *koron* (ᚹ), which are now quite commonly used to designate the slightly sharp and flat tones in Persian music.

20. Nettl (1964:128) considers automatic transcription to be "ideal," but finds justification for aural transcription as a means of personal discovery of music and also indicates its value in publications for consumption by the non-specialist. These are undoubtedly important considerations for recommending aural transcription; however, it is the frequently held notion that automatic transcription represents some kind of an ideal which has prompted the writing of this paper.

REFERENCES CITED

Bartók, Béla and Albert B. Lord
1951 Serbo-Croatian Folk Songs. New York: Columbia University Press.

Bergeijk, Willem A. Van, John R. Pierce, and Edward E. David
1958 Waves and the Ear. New York: Doubleday and Company, Inc.

Burroughs, Lou
1974 Microphones: Design and Application. New York: Sagamore Publishing Company.

Dahlback, Karl
1958 New Methods in Vocal Folk Music Research. Oslo: Oslo University Press.

Feld, Steven
1976 Ethnomusicology and Visual Communication. ETHNOMUSICOLOGY 20:293-325.

Hood, Mantle
1971 The Ethnomusicologist. New York: McGraw-Hill, Inc.

Jeans, James
1937 Science and Music. New York: Dover Publications, Inc. (reprinted 1968)

Josephs, Jess J.
1967 The Physics of Musical Sound. Princeton: D. Van Nostrand Company, Ltd.

Kubik, Gerhard
1962 African Music. Journal of the African Music Society. 3(1).

Kunst, Jaap
1959 Ethnomusicology. The Hague: Martinus Nijhoff.

List, George
1974 The reliability of transcription. ETHNOMUSICOLOGY 18:353-77.

Nettl, Bruno
1964 Theory and Method in Ethnomusicology. Glenco: The Free Press.

Roberts, Helen
1931 Suggestions to field-workers in collecting folk music and data about instruments. Journal of the Polynesian Society 40:103-28.

Seashore, Carl E.
1938 Psychology of Music. New York: Dover Publications, Inc.
(reprinted 1967)

Seeger, Charles
1958 Prescriptive and descriptive music writing. Musical Quarterly 44:2.

Tremaine, Howard M.
1974 Audio Cyclopedia. Indianapolis: Howard W. Sams and Company, Inc.

Wood, Alexander
1963 The Physics of Music. London: Methuen and Company, Ltd.