Simultaneity and Polyphony in Speech Based Audio Art

Eric Somers

Department of Performing and Visual Arts
State University of New York
Dutchess Community College
Poughkeepsie, NY 12601
somers@sunydutchess.edu

ABSTRACT

The author discusses the use of simultaneous speech in the design of radiophonic sound compositions. He begins with a discussion of speech in pre-literate societies and shows how simultaneity is a characteristic of speech prior to the influence in writing and print, and discusses the how the linearity of printed communications has tended to eliminate aspects of spoken words that cannot be duplicated in print. The author then discusses the design of radiophonic sound projects, by himself and others, which utilize the ability of the ear to keep track of multiple simultaneous spoken narratives and which may more closely simulate thought patterns of eye entered cultures. Finally, he relates how the differences between eye-culture perception and ear-culture perception might impact the design projects related to the sonification of data.

1. INTRODUCTION

Living in a society which has been shaped by an emphasis on visual culture, it is easy to overlook forms of sound communication for which there is not a visual equivalent. As a sound artist, I am interested in producing sonic forms which exploit the unique properties of sound perception and communication.

One of these is the so-called "cocktail party" effect: the ability of the ear to "tune in" to multiple simultaneous conversations in a crowed room. There is no equivalent ability in the perception of written text. A reader would not be able to simultaneously read text from the two columns on this printed page, for example. Visual perception is linear. Even when a person seems to take in a complex visual image "all at once" a careful analysis of eye movements will show that the image must be scanned bit by bit by the eye in order for all elements to be seen.

2. SPEECH PRACTICE IN NON-PRINT CULTURES

2.1 Sounded speech in oral cultures

If one were to ask most people today which sense organ they would rather lose, the eye or the ear, most would choose deafness over blindness. But in cultures with no written language -- I will call them “ear cultures” since the term "illiterate" seems to carry pejorative connotations -- many people would choose blindness first. In those cultures the ear is the main means of word communication between people.

Ong [1], Carpenter [2], McLuhan [3], Schwartz [4] and others have described a number of characteristics of speech communication in those societies that bear striking differences from the roles of speech in eye cultures -- the term I will use for cultures that heavily communicate with writing and printing.

One characteristic is the lack of a strong sense of temporal placement in terms of historical events. There is no sense of "time line" as in eye cultures because it is written text, with its linear organization, that has given us a sense of past, present and future. In addition, written words can be stored, re-examined and tested against other stored data to develop a sense of structure and organization.

In ear cultures a body of quasi-historical oral stories does evolve and are repeated and learned by troubadours, but as the stories are passed down -- and contrary to a number of myths that have evolved about ear cultures -- the events and characters in the stories are altered, and a strong sense of the exact dates and times of events is lost. [5]

Speech in ear cultures consists of three inseparable components: diction (in its original meaning of "choice of words"), inflection (tessitura and loudness) and body position and movement (apparent, of course, only to sighted persons). [6]

Prior to the invention of writing and printing, there was no way that diction could be separated from the other two elements.

In a spoken culture it is not uncommon for more than one person to speak at the same time. Indeed, the act of "listening" often involves speaking. More about this below.

2.2 Eye dominance brought about by writing and printing

The written word preserves only the diction or choice of words. Spoken intonation and inflection as well as body gestures, position and movement in speech are not preserved.

According to McLuhan, Ong and others (previously cited), as societies became oriented to the printed word -- especially after the printing press made low cost books and periodicals available to large numbers of people -- there evolved an interaction between the printed and spoken word. People considered "refined" or "educated" were taught that in delivering a spoken text, one should not convey much more than can be conveyed in print. People were taught to have a "well modulated" voice free of great pitch or loudness variation, and to stand still while speaking in public or to others. Large gestures were considered "vulgar."

Although one might imagine that one would have to travel to a very remote place to observe evidence of the differences between the use of speech in an ear culture and an eye culture, this is not necessarily true. There is a tendency for preserving "old ways" in the practice of religion. If one compares the

religious practice of cultures that were fairly recently ear-only cultures with the practice of religion dominated for centuries by print, the differences between the use of sounded speech become strikingly apparent.

Many African-American churches are based on similar gatherings from slave times when most African-Americans were not taught to read or write. If one attends such a service today (or watches one on television), one can observe speech practice based on oral tradition. The preacher in many of these services uses a very inflected form of speech, sometimes shouting, sometimes almost whispering, sometimes singing or chanting. These widely varying changes in tonality and loudness are accompanied by broad gestures and considerable movement of the body. When the choir sings, the singers do not stand still but usually sway gently to the cadence of the music.

Most important, the congregation of "listeners" shows their appreciation and understanding of the preacher's message by verbally adding comments (e.g. "Amen") during the pastor's delivery. Thus listening itself becomes a process of speaking simultaneously with the preacher.

Contrast that with, say, the religious practice of Anglican (Episcopal) worship. In this service the priest delivers his sermon with a very controlled voice having limited variation in pitch and loudness. He stands still while speaking, using few gestures, Most important, the listeners never speak out loud while listening. Indeed, if they did they might be asked to leave the service for being "disruptive." The Anglican choir sings without swaying, and everyone waits his or her turn before speaking or acting during the service.

The Anglican service, like most everything else in eye cultures, seems to have been influenced by eye dominance in two ways: The method of speaking involves verbal presentation of only that aspect of speech which is preserved in the written word, namely diction. Inflection and body movement are restricted. Similarly the entire service in linear. Events happen according to a time line in which each participant says or does something at a specific time after waiting his/her "turn." There is much less simultaneity than in the African-American service based on ear culture practice.

Lest one think the example above is characteristic only of Western cultures, another example of the vocalized and gestural "listening" can be seen in the Korean film, *Chunhyang* by Im Kwon Taek. This film features narration by a Pansori artist, a traditional troubadour-like performer who relates an epic story in highly expressive chant-like speech accompanied by a single drum used mostly for emphasis. Although the story the Pansori artist is telling is ultimately dramatized in the film, the Pansori narration continues and at various emotionally intense portions of the film, the image shifts from the historical recreated to simply showing the Pansori artists performing for a modern audience in a theatre. As with the African-American church service, the audience makes verbal exclamations while the performer is speaking, and at a few emotionally intense moments some audience members virtually stand up and make dance-like body motions while listening to the performer. Again we see the close connection between body movement and sound, as well as speech simultaneity between the listener and the performer.

The purpose of this historical review is to point out that the ear and the eye promote very different ways of thinking and perceiving. In eye cultures the modes of thinking and perceiving related to the eye are often given dominance with eye values imposed on aural communication. Speech communication is often linear and consists of a series of sequential, not simultaneous, actions on the part of speakers and listeners. In ear cultures there is much more simultaneity and less temporal orientation.

3. POLYPHONY IN MUSIC AND SPEECH ART

3.1 Musical polyphony as a basis for speech simultaneity

We have noted above the unique ability of the ear to perceive simultaneous sounds. This is most evident in music where different instruments are playing different notes simultaneously. Yet this might not be considered a good example, because in tonal music the harmonic relationships which exist between the notes are actually producing, at any moment, a more complex sound from several simpler ones. Thus it can be argued that three simultaneous notes are not three sounds, but a single complex chord synthesized from the sounds of three instruments playing together. However, Wishart [7] and others have noted that polyphony can be followed even in musical lines which have no natural harmonic relationship to each other. Other factors, such as timbre, pitch range (tessitura) and spatial separation still permit the listener to follow the individual simultaneous melodic streams.

Polyphony in choral music also introduces the possibility of following texts in which the same words are not sung by each voice part at the same time. Sometimes this difference is a matter of offset, one part of the choir singing "Hear my prayer, Oh Lord . . ." While another vocal group begins the same text staggered in time so that they may be singing "Here my prayer" while the first part is singing "Oh Lord." Yet there also exists polyphonic choral music in which different voice parts are singing entirely different texts (e.g. some Tudor choral music), though it is much less common.

What happens when these same notions of polyphony and simultaneity are applied to radiophonic speech art? Some of the earliest and most celebrated work in this area were three programs created for the Canadian Broadcasting Corporation (CBC) by the noted pianist, Glenn Gould [8]. In the first, *The Idea of North*, Mr. Gould crossfades between different voices describing their experiences of living in the northern part of Canada. More than one voice can be heard simultaneously throughout much of the piece, although Mr. Gould makes it pretty clear which is the "main" speaker at any point by making that voice the loudest. When it first aired, the program generated immediate controversy for departing from the traditional radio format of sequential speech. Nevertheless, he went on to produce two more such documentary programs in the same style.

Although talked about in terms of musical polyphony in light of Mr. Gould's musical background, the form did not follow traditional choral polyphony as much as one might at first think.

In all of the research done about the ability of the ear to follow multiple simultaneous conversations, the so called "cocktail party problem " identified by Cherry [9] in 1953, there seems to be little doubt that by dividing the attention of the listener, the ability to closely follow any one stream is diminished.

In choral music that situation is usually compensated for by introducing a high degree of repetition and redundancy into the text. Thus, key lines are repeated lest the listener miss some of the text while attending to another voice part. In *The Idea of North*, less of this redundancy is present. Thus many listeners perceived the background voices to be a distraction while

attempting to follow the "main" speaker. Since the Gould radio pieces were monophonic, there was no spatial differences to help separate different simultaneous speakers.

It can be argued that choral music, even forms where different texts are sung simultaneously by different voice parts, is not really indicative of the kind of simultaneity found in ear cultures. Polyphonic singing did not appear until after the invention of both writing and written musical notation, and did not flower into its most productive era until after the printing press and helped promote widespread literacy. The polyphony of traditional choral music is a linear one, several "lines" are sung together, but highly coordinated by a common metre and controlled by the dominance of a single conductor.

By contrast, some traditional forms of music in ear cultures have multiple non-synchronized drummers and events of indeterminate duration improvised on the spot. The modern ear-culture successor to this traditional form of music is less the multi-lined choral composition and more the techno or turntable music of the dance and hip-hop culture in which samples of melodies, beats and speech are combined in a very non-linear fashion.

3.2 Perception of speech simultaneity

There has been considerable research on the factors which permit one to follow multiple streams of speech. In his book, *Auditory Scene Analysis*, Bregman [10] summarizes work by Cherry, Dorman, Warren, Darwin and others regarding research about the ability to pick out speech from multiple speech streams, speech with other sounds, etc.

Cherry's work, augmented by other researcher following up showed that two conversations separated in perceived space were easier to follow than the same two conversations perceived monophonically. Similarly, differences in pitch and timbre between the two voices also makes it easier to separate the conversations. Bregman himself argues that what seems to be physical differences may actually be the result of more complex psychological processes of separation as the perceived differences may not actually match the physical ones. An article by Jones and Yee [11] in *Thinking in Sound*, summarizes work by Martin], Shields *et al.*], Meltzer *et al.* and Cutler indicating that differing stresses on certain words or syllables was an aid to differentiating multiple speech streams.

Clearly, this work offers clues to the success not only of polyphonic speech, but of sung polyphonic music. In such music, for example, the voice parts are separated in pitch and space, and the differing stresses on words or syllables is usually much greater than in normal speech. The research highlights the very different cognition that occurs when work are spoken rather than written or printed, and it is these very differences which can form the basis of sound art based on polyphonic speech and even suggest uses in other forms of sonic design.

4. DESIGN AESTHETICS INVOLVING MULTIPLE SPEECH STREAMS

As a sound designer, I am not so interested in testing the factors which make perception of multiple speech possible, but in creating compositions which use simultaneous speech to create a word experience that cannot be duplicated in written or printed form. Nevertheless the design factors related to creating effective multiple speech compositions seem consistent with the research cited above and with some of the techniques of polyphonic choral composition.

4.1 "Stag in a Boat"

In a suite of electro-acoustic compositions I created for the "Stag in the Boat" art installation by the American painter and performance artist, Ann Wilson, each movement is introduced by multiple voice streams of Ms. Wilson reading text related to the theme of that movement. As with the pieces of Glenn Gould, the sound is monophonic -- only the music which follows is stereophonic -- but the staggered voices are not varied in loudness as with the Gould works. The same speaker is doing all of the reading, so there is also no variation of timbre.

From the standpoint of perception, one can probably assume that the intelligibility is based on the high degree of redundancy (as with much polyphonic choral music), as well as differences of stress since factors of spatial separation and pitch/timbre differences are not present.

But even in this simple use of spoken polyphony, there is obviously an aesthetic or design reason for choosing to use multiple voice streams and certainly the result of doing so producing an extremely different end result from having Ms. Wilson simply read the text in a single linear voice. It is this difference which makes me interested in simultaneity in text based audio art. I mentioned above that the polyphony of *The Idea of North* was found to be distracting and confusing by some listeners. But can simultaneity create an experience which is substantially richer than a linear one?

The cynic might say the multiple speech streams are simply an attempt to be "arty" or even to be pretentious. I suppose early composers of polyphonic choral music might have received some of the same skepticism by others still staying with simpler monophonic forms. Even though intelligibility is possible (for some of the reasons cited above) in polyphonic speech and singing, nobody argues that it is enhanced with these techniques.

The aesthetic underpinnings which make simultaneity powerful in an auditory world may come from a re-examination of speech patterns in ear (non-literate) cultures. Also, ear-centric design thinking may help in designing more effective uses of sonification.

It an eye centered culture where linear thinking and linear perception is a norm, it is easy to forget that the multiple experience world of the ear culture is the norm for hearing. Before a newborn baby is visually aware of its surroundings, it has auditory awareness. Sounds in the natural world are rarely isolated, but come mixed with a variety of environmental sounds which may include multiple speech but certainly include speech (from a parent) with other sounds. The infant responds not so much by listening quietly as by making its own noises in response to what it hears.

Similarly, linear thinking and reasoning is something which has to be learned. The natural thought pattern is one of collage, in which various ideas and perceptions seem to be assembled in one's mind and impressions created from them.

A textual collage of polyphonic speech more closely replicates this natural process of information intake and processing, than does a formal talk which is usually based on a written text. Indeed, the word "lecture" itself is derived from a middle English form referring to the act of reading. Thus linear speech is really a product of an eye centric culture.

The mistake made by Glenn Gould, if indeed it is a mistake, was to present what was essentially long linear monologues in a polyphonic form. Referring again to the turntable artist, we

might consider “sampling” and repetition a better model for spoken polyphony.

4.2 “Hey, Boboaca!”: Deconstructing print into aurality

Most recently I have composed a sound composition in which a short, one page, written text is deconstructed into a fully aural process in which ideas are communicated by non-linear simultaneous speech and other sounds. The attempt is to mimic complex thought processes which produce multiple thought when hearing a story.

The premise is that linear thought processing, the kind required for me to write this paper, for example, are learned. One’s own internal communication consists of a series of disjointed ideas and images, some of which are repeated over and over in one’s head while others come and go and intermix. It is my belief that the use of speech in ear cultures more closely follows this pattern rather than a linear begin-end-middle model.

Some of the characteristics of oral narrative identified by Ong [21] include "aggregative rather than analytic," "redundant," "Close to the human lifeworld," and "empathetic and participatory." I became interested in exaggerating each characteristic to produce a sound collage in which an original story gradually reveals itself in the context of people relating personal experiences as well as narrating portions of an original text in a quasi musical polyphonic fashion.

The result is a radiophonic deconstruction, *Hey Boboaca*, in which I have taken a one-page short story by the Romanian-American writer, Lucia Cherciu and recreated it as a sonic collage of elements of the story (read by nine different readers) and well as side comments by some of the readers that relate incidents in their own lives to what is being told in the story. Part of this mirrors one's internal experience in listening to someone tell a story, where one’s own experience interjects ideas and images in one’s mind while listening. Indeed sometimes one even shares one’s own experience with the person telling the original story and with other listeners. Ong has said "Spoken words are always modifications of a total situation which is more than verbal. They never occur alone, in the context simply of words." [13]

In my piece, some fragments of the text re-occur in a repetitive fashion (think again of sampler art) but often by different readers. Other lines are heard only once or twice. (Aren’t some phrases more memorable when hearing something being read?).

It is not clear that a listener, hearing “Hey Boboaca!” would be able to retell only the author’s original story without any of the side information added by the various readers. But the notion of single authorship is again a product of print culture where texts are preserved, catalogued, and ordered. Oral stories are passed on and modified without as much of a sense of a single author. Only when something is written down is it often ascribed to an “author.” Ong points out theories that the poems ascribed to "Homer" are hardly original with him that "Homer stitched together prefabricated parts." [14]. Again, the similarity to modern "sampler" art forms cannot be ignored.

Thus “Hey Boboaca” is almost an attempt to “unwrite” Ms. Cherciu’s story and re-create it as an unfocused oral legend. The intent of the piece – and only listeners can judge its success – is for the listener to “find” the story in a much broader context and to think less about the original story as an object, but more about speech as a result of experience. The use of auditory simultaneity, collage construction and redundancy are exaggerated so as to create an acute awareness that the piece is sound art, not text to be visualized as if one were reading it from a book.

5. IMPLICATIONS FOR THE DESIGN OF SONIFICATION

In the course of creating *Hey Boboaca*, as well as in my experience with prior text pieces, I have become acutely aware of the difference between auditory experience and visual experience.

Of course, we must remember that literacy changes to a certain degree the oral (or “aural” – Ong tends to think of the speaker and use the expression “oral” while I relate to the listener and use the form “aural”) patterns of a culture. Cultures without writing were (and are) oral/aural cultures. For a great period after the invention of writing, the written word was the only way information could be stored outside the human being. Today video and audio recording permit the gathering of “oral histories” and the passing along of oral (or aural) traditions in non-text forms. Thus I consider modern culture not eye culture nor ear culture, but media culture, an integration of some of the values of both.

McLuhan, Swartz and Carpenter (see references below), among others, have written about the occurrence of oral patterns of thought in modern media culture. I submit that a major error is thinking that information we hear can be, or will be, treated the same in our thinking as information we gather with our eyes. Ong (who has greatly influenced my view of ear culture) reminds us that the eye separates and organizes while the ear integrates and internalizes.

It strikes me that a number of sonification efforts – attempts to present data sets in sonic form so as to understand relationships differently from visual representations of the data – seem focused on attempting to better understand organization through sonification. I submit that a more worthy goal of sonification, would be to use sonification as a tool of experience more than analysis.

Let me offer an example from the field of music. If one hears a difficult atonal piece of music for the first time, it can seem formidable. There is even a doubt that the listener could hear the same piece a second time, several weeks later, and identify it again. If one wants to understand its construction, the fastest way is to examine the music in printed form, the score. Yet such an analysis still removes the music from the realm of direct experience. If the listener played the work over and over again (and we remember than repetition is an important aspect of learning in ear cultures), then gradually the hearer would come to think of the work as “theirs” in the sense that they would recognize it and fit it into the totality of their musical experience (in a much more direct way than studying the score).

The experience of hearing Cherciu’s short story presented in a highly simultaneous and sonically integrated fashion will strike the typical short story reader as difficult and forbidding. Yet its very simultaneity will cause juxtapositions of expressions and visualizations (derived from the words) than the linear form may never induce. The story will be internalized as experience, not viewed as an external object apart from one’s own body.

Recent thinking in the field of education supports the notion that addressing a variety of senses enhances the learners understanding of a given subject. It is recognized that learning is enriched by bodily and sensory experience that go beyond reading, writing and listening to lectures.

Thus the greatest potential for sonification may be not to make our ears into another form of our (very analytical) eyes, but to try and understand what can be understood through assimilating complex sounds through repetition and deconstruction ("looping" as it were). Certainly this is an area that perceptual psychologists might explore. Being an artists and not a scientist, I do not have a good experimental model at this time, but as a sound designer, I have come to realize that internalizing simultaneous spoken text provides a vastly different understanding of the words than is achieved by either reading words on paper or hearing them presented in traditional linear spoken form.

6. REFERENCES

[1] W. J. Ong, *Orality and Literacy. Methuen*, London and New York, 1982.
[2] E. Carpenter, *Oh, What a Blow That Phantom Gave Me!*, Holt, Rinehart and Winston, New York, 1972.
[3] M. McLuhan, *Understanding Media: The Extension of man*. McGraw-Hill, New York, 1964.
[4] T. Schwartz, *The Responsive Chord*, Doubleday Anchor, New York, 1973.
[5] W. J. Ong, *op. cit.* pp. 57-67.
[6] W. J. Ong. *op. cit.* 47, 101-102.
[7] T. Wishart, *On Sonic Art*, Harwood Academic Publishers, 1998. pp. 109-125.
[8] G. Gould, "The Idea of North" on the 3-CD set *Solitude Trilogy*, on CBC Records' Perspective Series, PSCD 2003-3. It also includes Gould's "The Latecomers" and "The Quiet in the Land."
[9] E.C. Cherry, "Some experiments on the recognition of speech with one and with two ears." *Journal of the Acoustical Society of America*, 25, 975-979.
[10] A.S. Bregman, *Auditory Scene Analysis*, MIT Press, Boston, 1994.
[11] M. Jones, W. Yee, "Attending to auditory events: the role of temp[oral organization, *Thinking in Sound*, S. McAdams, E. Bigand, ed., Oxford, 1993, 2001.
[12] W. J. Ong, *op. cit.* pp. 38-45.
[13] W. J. Ong. *op. cit.*
[14] W. J. Ong. *op cit*. pp. 17-19.