Ensoniment of Plato's *Symposium*: Theory and Engineering Design for Symposia

Abstract

The ensoniment of Plato's *Symposium* begins as a free, open source software project hosted on Sourceforge.net called Symposia. Concurrent text to speech processes pronounce the Socratic dialogue as if it was spoken by ten speakers arranged as in the implied virtual setting of the story. Venturing beyond the typical subvocalized expectations of consumers of alphabetic, visual characters in written texts, the potential sonic environments this software sound system might produce challenge ocularcentric conceptions of human subjectivity, and invite rethinking posthuman, cybernetic, program-using and program-writing identity.

What does an ensoniment of Plato's *Symposium* mean? In *The Audible Past: Cultural Origins of Sound Reproduction*, Jonathan Sterne introduces the term 'ensoniment' for the organization of sound, as 'enlightenment' is for the scientific organization of primarily visual phenomena. "As there was an Enlightenment, so too was there an 'Ensoniment'. A series of conjunctures among ideas, institutions, and practices rendered the world audible in new ways and valorized new constructs of hearing and listening" (2). Sterne looks beyond technological artifacts to social practices codifying sound and listening techniques. His account of sound reproduction media transcends discrete devices to whole networks of knowledge systems, power relations, and cultural norms, detailing the development of specialized listening techniques that went hand in hand with other associated cultural developments like canning, embalming, capitalism, and individual bourgeois consumption. As his analytical history extends into the twentieth century, it stops short at the verge of *synthesis*, that is, the mechanical and electrical production of sounds that is not based on playback of recorded phenomena. He mentions mechanical inventions of 1770-1790, references to experiments by Alexander Graham Bell, and the craft of Johannes Faber, "an astronomer with failing eyesight, [who] built a talking automaton called the *euphon* in the early 1840s. It consisted of a simulated torso and head

(dressed like a Turk), a bellows, an ivory reed, and a keyboard the keys of which corresponded to various states of the mouth" (77).[1] Then systematic sound studies shifted from the vocal tract and states of the mouth, to the ear, in the range of twenty to twenty thousand hertz vibrations detectable by tympanic mechanisms akin to the human ear, and, according to Sterne, has remained so conceptualized. Thus, I argue, Sterne's analytical approach is bounded and is overdetermined by this model for sound reproduction, helped by its locus of tympanic (otic) modeling, overtones, and additive vibrations of which the speaking voice is a subset, and does not further broach speech synthesis. The ear listens, it does not originate sound like the human vocal tract. Further mention of sound synthesis is relegated to a footnote, in which Sterne posits Thaddeus Cahill's telharmonium as the link between Helmholtz's synthesizer and modern tone-combining sound synthesis devices (366n70). He makes no further mention of speech synthesis, perhaps because "Helmholtz's celebrated synthesizer, which used tuning forks and resonators to create vowel sounds, was modeled on the ear" (77).

I am remediating the term 'ensoniment' to apply to the programmed, synthetic, mechanical sound production of artifacts that have hitherto existed as *visual* codes, virtual sonic realities, made real by humans reading written texts. While Sterne critiques a language detection criterion as part of the biased 'audio-visual litany' distorting sound studies, when it comes to text to speech synthesizer software design, some kind of symbolic sequence of symbols signifying written speech for the input must be prominent in the processing to yield sound as synthetic speech. In response to blatantly ocularcentric approaches to electronic literature in the digital humanities that focus on visual texts, I am proposing a software project that summons audible

1 The interesting digression into the gender, racial, and cultural implications of information technologies to which N. Katherine Hayles appeals in *How We Became Posthuman* and *My Mother Was a Computer* is beyond the scope of this text. The fact that Faber's talking automaton was "dressed like a Turk" ought to be recalled for future studies.

presentations of visually encoded texts. It remediates visual texts associated with the classical Greek Plato's *Symposium* via text to speech computer software to create auditory environments never uttered by humans, except at the site of the mythical antecedent to the narrative.

Plato's *Symposium* involves ten discrete speakers whose spatial position can be roughly determined so that a group of people could arrange themselves and dramatically recite it so that a fixed or roving listener who is either one of the speakers or a silent, perhaps invisible auditor would hear the combined voices in the omnidirectional, simultaneous, many-at-once auditory field. If presented in a three dimensional virtual reality environment, this phenomenological presence would be generated by simultaneous text to speech functions running concurrently, producing the speech emanating from the position of each speaker in the dialogue. In it a virtual listener could navigate the auditory field over the duration of its performance. What would that listener hear if immersed in this sonic space? Taking off from the default, one speaker at a time rendering of the original text, what kinds of alternatives are possible? How might our notions of subjectivity be tested or perhaps mutate by playing such scholarly language games? If we are going to go to the trouble to create an artificial sonic environment as a workplace, testbed, laboratory, or playground in order to pursue such questions, we ought to take into consideration the accrued insights from sound studies, digital media studies, and electronic literature studies to put an edge on an otherwise accidental essay on the technical specifications, requirements, and overall engineering and integration design of a free, open source software project hosted on Sourceforge.net.

Theory: Speech Sound Production

Beyond Sterne there are many accounts of modern histories of speech synthesis. I will mention J. C. R. Licklider's famous 1960 piece “Man-Computer Symbiosis” as a representative instance seeking to establish a cultural demand for automated text to speech synthesis by

electronic computers for the benefit of business executives and especially military commanders. Licklider explains that "it is easy to overdramatize the notion of the ten-minute war, but it would be dangerous to count on having more than ten minutes in which to make a critical decision. As military system ground environments and control centers grow in capability and complexity, therefore, a real requirement for automatic speech production and recognition in computers seems likely to develop" (81). In a move by which Kittler later takes credit for explaining. but Sterne refutes, technological change is determined by military concerns – Heraclitean "war is the father of all" – cultural demand for speech synthesis and recognition increases because military commanders need to make quick decisions. To Licklider and most subsequent artificial intelligence researchers and information system designers, speech recognition poses a much more complicated and perhaps intractable problem for machine cognition to accomplish than the reverse, text to speech synthesis: "For real-time interaction on a truly symbiotic level, however, a vocabulary of about 2000 words, e.g., 1000 words of something like basic English and 1000 technical terms, would probably be required. That constitutes a challenging problem. In the consensus of acousticians and linguists, construction of a recognizer of 2000 words cannot be accomplished now" (81). Curiously, the subtle, context-sensitive cues implied in natural language that confound mechanical analysis of tympanically transduced sound samples are the very components of the human experience that sound reproduction technologies fail to reproduce: interiority, spontaneity, intention, 'soul'. The medium, not the content, is the message in the sense of what grounds all discernable content. Kittler recognizes this exasperating state of affairs with media, writing in *Gramophone, Film, Typewriter*:

> Understanding media – despite McLuhan's title – remains an impossibility precisely because the dominant information technologies of the day control all understanding and its illusion. . . . Our media systems merely distribute the words, noises, and images people

> can transmit and receive. But they do not compute these data. They do not produce an output that, under computer control, transforms any algorithm into any interface effect, to the point where people take leave of their senses. At this point, the only thing being computed is the transmission quality of storage media, which appear in the media links as the content of the media. (xli-2)

The conundrum is that human understanding operates *via* media, so must employ media in addition to its internal resources when thinking about itself and media, yet does not really understand the range of potentials of each medium, its affordances, as well as its limits, all of those internal factors that affect its operation. Thus media evolve around an empty, ideal instance that is never actualized, that nonetheless forms the basis upon which all other expressions operate. As Sterne explains,

> it was the context of reproducibility itself that mattered; the specifics of speech and voice itself did not even really matter. The inside of sound was transformed so that it might continue to perform a cultural function. . . . Recording is a form of exteriority: it does not preserve a preexisting sonic event as it happens so much as it creates and organizes sonic events for the possibility of preservation and repetition. . . . If the past is, indeed, audible, if sounds can haunt us, we are left to find their durability and their meaning in their exteriority. (306; 333)

The historical narrative of sound reproduction that Sterne relates provides evidence of design decisions being made again and again, whether intentionally by marketers or indirectly by consumer response, that affect the nature of subsequent auditory culture itself, producing listeners, for example, who do not hear the static in a gramophone, or who do expect commercial advertising to interrupt a radio broadcast. Speech synthesis, in the same way, should be expected to reflect similar design decisions. The cultural origins of speech synthesis, hinted at by

Licklider, reflect battlefield pragmatism rather than artistic simulation. On the one hand, computer speech synthesis must be doubly cursed from the start, employing the simplest of algorithms to yield acceptable sonic data to no-frills decision makers who do not really care what the synthesized speech sounds *like* as long as its communicative function is optimized. On the other hand, the designers of speech synthesis software have the opportunity to intentionally manipulate the exteriority of the sounds they generate in order to simulate the interiorities that are stripped by sound reproduction reduced to the common function of the tympanic mechanism of additive vibrational frequencies.

Roland Barthes' philosophical writings about “the grain of the voice” and listening provide insight into audible nuances that speech synthesis could strive to produce, given the appropriate social inclination motivating the engineering design requirements guiding software developers. Barthes makes this distinction:

> The *pheno-song* (if the transposition be allowed) covers all the phenomena, all the features which belong to the structure of the language being sung, the rules of the genre, the coded from of the melisma, the composer's idiolect, the style of the interpretation: in short, everything in the performance which is in the service of communication, representation, expression. . . . The *geno-song* is the volume of the singing and speaking voice, the space where significations germinate 'from within language and in its very materiality'. . . . the whole of musical pedagogy teaches the not culture of the 'grain' of the voice but the emotive modes of its delivery. (182-183)

While referring to song, these terms seem applicable to speech in general. In *Discourse Networks, 1800/1900*, Kittler provides insight into the cultural origins of a particular kind of expressive reading that is practiced by nearly everyone today, known as 'subvocalized', phonetic reading, in which, as N. Katherine Hayles explains, quoting Kittler in *My Mother Was a*

Computer, "reading functions as 'hallucinating a meaning between the letters and lines'" (4). The passage continues, "these practices gave 'voice' to print text, particularly novels – and the voice most people heard was the same voice that taught them to read, namely, the mother's, which in turn was identified with Mother Nature and a sympathetic resonance between the natural world and human meaning" (4). With respect to written literature, such as Plato's *Symposium*, both the original Greek and more evidently its annotated translations strive to encode a pheno-song via stylistic conventions so that the reading evokes the desired communicative and expressive objects, whereas the geno-song emerging during performance is dependent solely upon the improvisation of the speaker or these effects of subvocalizations during silent reading. Programming a computer to generate audio speech from alphabetic source text leaves room for attention to pheno-song and geno-song details, in addition to allowing a default reading to prevail.[2]

Theory: Phenomenology and Subjectivity

A recurring theme of sound studies is that vision has been the preferred sense since the ancient Greeks, and therefore audition is distorted by 'ocularcentric' biases (Keller and Grontkowski; Levin). Don Ihde, writing about the phenomenology of the auditory field in *Listening and Voice: A Phenomenology of Sound*, notes the key differences between the two:

> The auditory field as a shape does not appear so restricted to a forward orientation. As a field-shape I may hear all around me, or, as a *field-shape*, sound surrounds me in my embodied positionality. . . . My auditory field and my auditory focusing is not isomorphic

2 There is a tradeoff between the imaginary auditory effects achieved by subvocalization during silent reading, and the foreclosure of this literary mode of reading by the concrete manifestation in sound synthesis. I later explain that while germinating significations from geno-song nuances may not be extensive with formant synthesis, mixing multiple voices may provoke other alterations to the standard listener subjectivity.

> with visual field and focus, it is *omnidirectional*. . . . The auditory field surrounds the listener, and surroundability is an essential feature of the field-shape of sound. (74-75)

In addition to its omnidirectional surroundability, the auditory field is continuous and penetrating, contrary to the episodic nature of visual images whose presence can be annihilated by averting one's gaze or closing one's eyes.

Synthesizing a single voice conceals design details of how the auditory field is presented to the listener; indeed, silently reading printed texts forces the reader to imagine – *hallucinate*, per Kittler – the virtual reality in which sonic events transpire. Ensounding Plato's *Symposium* is no exception. This complex narrative occurs in a number of encapsulated soundscapes: first, the outer dialogue between Apollodorus and his companion, who is also in the position of the reader of the text; within this frame, Apollodorus recounts a meeting with his friend Glaucon, who bids him to recount the famous party years ago at Agathon's where Socrates and the other symposiasts gave speeches in praise of Love; and within this frame, the putative word for word account of the events by Aristodemus, who accompanied Socrates. Aristodemus narrates the back and forth conversation of the primary participants, who are Agathon, Phaedrus, Pausanias, Eryximachus, Aristophanes, Socrates, and finally Alcibiades, who bursts in with a group of drunken revelers after the other speeches have been given. Clues in the text suggest a probable spatial disposition of the speakers comfortably reclining in pairs on couches along a large table; the location of Aristodemus, the silent listener and recorder of the event, can only be guessed. A range of likely arrangements, however, can be readily illustrated (Figure 1), if we accept the implications of Eryximachus' uttering “each of us in turn, going from left to right, shall make a speech” (177D) and then vary the possible arrangements of the dining area, all consistent with the passage that refers to Phaedrus delivering the first speech “because he is sitting first on the left hand” (177D).

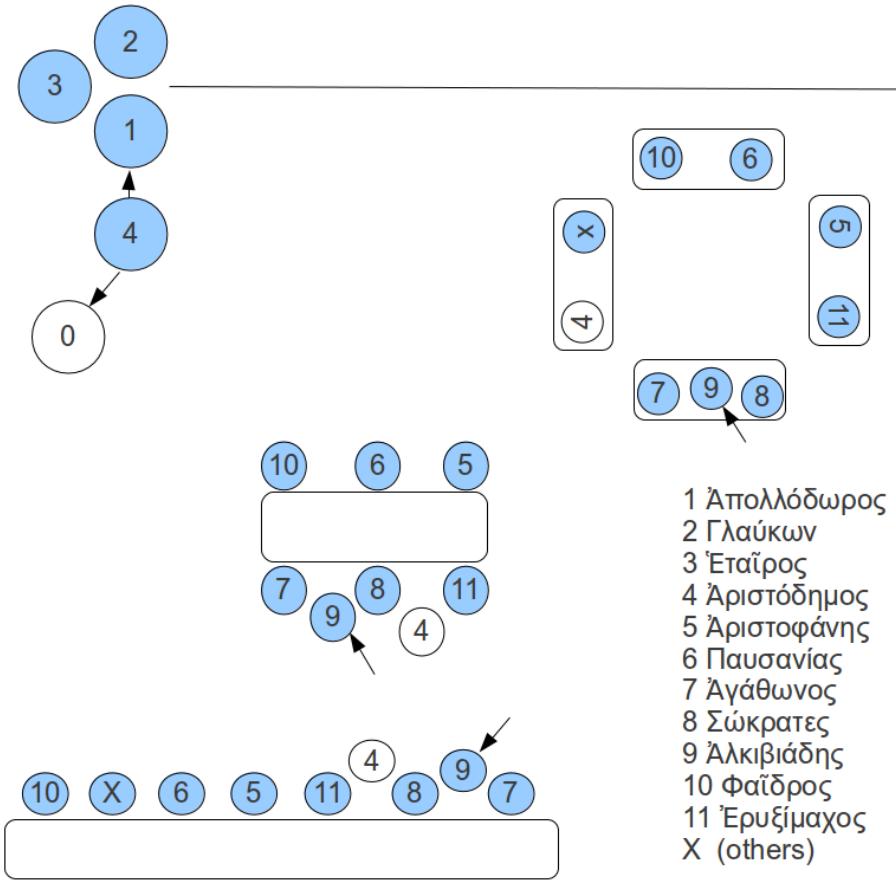


Figure 1: Various interpretations of the positions of the participants at Agathon's symposium, along with the enframing conversation between Apollodorus and the companion repeating his early walk with Glaucon.

From the perspective of ordinary reading, these organizations of the auditory field, if any thought is given to it, and also the pheno-song and geno-song characteristics of each speaker, must be deliberately fantasized. In reading we are not likely to hallucinate much more than a single voice in our heads; we do not imagine the omnidirectional auditory field described by Ihde. Indeed, on the one hand, it is by literary license that Aristodemus was able to hear each speech clearly and record it accurately, if he was indeed seated in a particular location for the duration of the event. On the other hand, an audio presentation of the event could be synthesized as if from the perspective of a listener spatially positioned (at least virtually) so that each speaker is perceived

to emanate from the appropriate distant, directional location. However, like the overall qualities of sound production and reproduction in general, such an expectation is culturally situated.[3]

Theory: Programming New Media

Humanities scholars, even those who work in the digital humanities, seldom engage in large scale computer programming projects. Trace Reddell identifies a few who do administer software projects that are sonic experiments in "The Social Pulse of Telharmonics: Functions of Networked Sound and Interactive Webcasting." He cites John Cage as a pre-computer explorer whose work inspires Peter M. Traub's "Bits and Pieces" (1999) and Freeman and Skeet's <earshot> (2002) projects.

> In the thesis accompanying the "Bits and Pieces" installation, Peter Traub claims numerous aesthetic precedents but few technical ones. . . . These scripts allow Traub's computer to make network connections, download web pages, parse HTML, and

3 Typical interpretations of the *Symposium* focus on concepts of love developed in the individual speeches, such as found in Wikipedia and countless introductions to English and surely German, French, and perhaps older Latin translations. They mostly miss its careful study of character position within virtual realities implied by the text, and give little thought to its complex framing of narratives. For example, the tension in the outer frame between Apollodorus, the putative narrator, and a generic reader, you, the companion, substituting for Glaucon in the current repetition of the text, and the special reader who responds in the voice of the companion claiming to be somewhat crazy. Perhaps we can better conceive the text from the position of media specific analysis of virtual realities by hearing it instead of just reading it. As the Figure 1 illustrates, consider what it would sound like in different configurations of the participants. In the long row Aristodemus would be the farthest from Phaedrus, whose speech is less faithfully recorded in the narrative Aristodemus recounts. Other configurations would produce different 'actual' audible fields had the event been recorded with two microphones representing the position of his head in the virtual physical environment, like in immersive first person games. There is the added complexity of remembering this retelling thematically occurs while two walk for hours for an early critique of intellectual tools (Greek alphabetic writing).

> download audio files. . . . Traub situates his installation in the historical and aesthetic contexts of collage and indeterminate art. Especially important are John Cage's spontaneous audio pieces generated through the manipulation of radio volumes and tunings, such as “Imaginary Landscape No. 4” (1951) and “Radio Music” (1956). (223)

Cage, however, did not work with high level, object oriented programming languages for manipulating the low cost, high speed, high capacity, networked computer systems currently available. Reddell discusses Traub's Perl programs operating on HTML and other files in the “Bits and Pieces” software powered audio system. Notable about this program are its automated Internet searching and retrieval of audio files that are randomly loaded into a sound synthesis program, eventually leveraging “server access, file deletion, file corruption, and other forms of broken linkage” (224). The <earshot> program by Andi Freeman and Jason Skeet provides a richer user interface than “Bits and Pieces” for creating audio mixes from culled web data. A third program Reddell mentions is “Constellations” by Atus Tanaka, whose programmed system “provides visitors with a custom webmixing interface installed on several computer kiosks distributed throughout the gallery” portrays a sonic device dependent on a visual interface. As the description continues, on account of the design that “the screen depicts . . . a planet's size on the screen equating to its volume in the mix, the media interface allows the player to combine multiple audio streams by navigating through the celestial field“ (227-228). Redell concludes:

> Using an interface more like a video game than a musical instrument – as well as taking more literally than Freeman and Skeet the emphasis on spatialization as a metaphor for sound mixing - “Constellations” is performed by an audience with little prior experience necessary to begin mixing sound sources. The navigational metaphor is immediately familiar, and the results of such telharmonic flights are almost instantly gratifying. (228)

Symposia may likewise deliver an automated mix of original Greek, multiple translations in various languages, commentaries by scholars throughout the centuries interjected during intonation of certain passages, and other transformations, in addition to offering ways for listeners to navigate the auditory field, perhaps inhabiting the bodies of one of the speakers for a transformed auditory "in bone" experience.

Although this arrangement seems isomorphic to a dramatic production, our acceptance for particular technological implementations is dependent on the cultural origins of sound reproduction. What I am proposing to do with concurrent high speed relational database driven processes in one virtual sonic environment could also be done via multiple physical systems spatially arranged with a listener in the appropriate place. Picking up where Sterne's *Audible Past* concludes, Rick Altman provides a condensed history of movie theater sound in "The Sound of Sound" based on the premise that "the location of speakers is remarkably indicative of contemporary presuppositions about sound . . . designed to match cinema sound to current standards of how sound should sound" (68). Like Sterne, he argues that technological changes do not in themselves change listening habits, although they certainly create the conditions for new ones to evolve. Altman concludes with some speculations about future iterations of computer sound systems, asking "Will CD-ROM-equipped computers need center speakers for talking books or voice-illustrated encyclopedias? . . . Will they feature FM connections to surround speakers, so that video games will feel truly wrap-around?" (71). Picking up where Altman ends, in "A Computer Environment For Surround Sound Programming," Whittleton and Corkerton describe the "Sound Storm" software system for achieving theater-quality effects with a specially equipped personal computer. It uses a visual interface in which a joystick controls the location of each sound emitter on a two dimensional stage area representing a four channel surround sound system (front left, front right, rear right, rear left). "The system is an advanced

visual programming environment that can in simple terms be considered as a computerized joystick which allows sounds to be moved around an audience" (8/1). To perform Plato's *Symposium* the motion feature of Sound Storm would hardly have to be used; the sound stage wound be laid out like Figure 1, Phaedrus on the far left and Agathon on the far right. The placement of the listener – as Aristodemus or as one of the participants – on the sound stage grid would inform the computer system to yield the surround sound effects. When Whittleton and Corkerton wrote their article, surround sound was not a standard feature of the typical personal computer. Indeed, Sound Storm as a technology is an 'early adopter' in that the authors, writing in 1994, anticipate a future consumer market where "the requirements for surround sound to be incorporated into products such as Compact Disks (CD), Videos, TV and radio broadcasts, Laser Disks (LD), Compact Disk Interactive (CDi), games, and multimedia, is becoming common place" (8/1). In hindsight, not all of these predictions panned out; surround sound has become a standard part of home theaters and computer games. What was formerly done on expensive, proprietary hardware and software can now, I believe, be done with generic personal computers and freely available, community developed software using a system like the one described in this document.

As will be further articulated in the proposed design, I recommend a sound-enabled, first person perspective game 'engine' for the ensoniment of Plato's *Symposium* in place of a special purpose, proprietary product like Sound Storm. Only in recent years have virtual reality engines and their cousins, immersive video games, begun to present the auditory field as "essentially invasive, resonant, vibratory, and immersive" for, as Steve Goodman paraphrases Erik Davis' seminal essay on acoustic cyberspace, "contemporary conceptions of virtual reality were trapped in a visual model of space . . . whereby the self transcends space, is detached from it, surveys it panoptically, as a disembodied vision machine" (114). The project requires sound generation

that supports multiple, spatially distinct sound sources 'influencing' (as the result of computation) the auditory field at another spatially distinct position.

Theory: Posthuman Permutations

According to Reddell, "Using the computer's formal interface to communicate with others through sonified data clusters, mutual webmixers fashion a kind of cultural chronotope, a shared situation or happening" (229). Likewise, a massive, multiplayer, online, role-playing game version symposia becomes an example of doing cultural chronotope with scholarly texts where regulation of ambiance becomes social responsibility, with threads of asynchronous scholarly debates emanating from other audio channels. Let machines support the sensible and intelligible boundary of not only the ocularcentric shimmering signifiers but also the noisy cacaphony of concurrent voices but producing that which no single or group of humans could or would want to speak (ensound, as in ensoniment). Electronic literature, philosophizing along this boundary of human and machine, Hayles argues, instantiates posthuman, cyborg subjectivity that has long outgrown liberal, humanist subjectivity.

> The subjectivity performed and evoked by this text differs from traditional print novels in subverting, in a wide variety of ways, the authorial voice associated with an interiority arising from the relation between sound and mark, voice and presence. Overwhelmed by the cacophony of competing and cooperating voices, the authority of voice is deconstructed and the interiority it authorized is subverted into echoes testifying to the absences at the center. (*Electronic Literature* 186)

Hayles portrays the boundary of human and computer as a Deleuzean war machine that attacks the interiority of the voice. With a slight adjustment audio reality can be made quasi-comprehensible again by limiting the number of significant (that is, perceptible within the auditory field) concurrent voices. It should be no surprise that Whittleton and Corkerton

"consider the optimum number of moving sounds to be three or possibly four as it is very difficult for the human brain to make sense of more than this" (8/2). Three or more moving or stationary *voices* are even harder to intelligibly follow the sounds, as anyone trying to listen to multiple conversations at a noisy party will attest. This limitation of human comprehension is exploded by computer systems that can process hundreds of multiplexed messages in real time, and electronic literature often exploits the many-at-once presentation of sounds and images to challenge the singular focus of traditional, liberal humanist subjectivity that many believe evolved through long acculturation with print technologies (Hayles). High speed data data processing and dynamic audio synthesis of multiple, intermixed tracks also permits manipulation of "memory glitches in which the distinction between past, present, and future becomes blurred." While Goodman is describing viral branding campaigns, he also compares this phenomenon of *deja entenu* to "accidentally stumbling across an original track when you are much more acquainted with its sampled riffs or vocal phrases populating another piece of music" (150). In a similar fashion the ensoniment of Plato's *Symposium* may present voices speaking a familiar text in different languages or translations that seem uncannily familiar to those who have read a print version.

In such ways, according to Hayles, posthuman subjectivity disrupts the literary, humanist, analog subject with alien processes that create rifts in the putative continuity soul, its likewise putatively simple ligature to body. Sonic experiments in particular may upset the audio-visual litany that so troubles Sterne, for "sound is always defined by the shifting borders that it shares with that vast world of not-sound phenomena. Sound is not the result of transhistorical interior states of the body or the subject" (343). Douglas Kahn refers to the 'polyglot' practice by a group reciting the same poem "'L'amiral cherche une maison a louer' (The Admiral is looking for a house to rent) . . . in German, English, and French (as well as in nonsense words, vocables,

singing, and whistling), moving in and out of relations of translation, by Richard Huelsenbeck, Marcel Ianco, and Tristan Tzara at the Cabaret Voltaire on 29 March 1916" (49). By complicating a simple, monolithic reading of a classical print text with a multiplicity of concurrent, competing sounds and voices, the symposia project tests the continuity of the conscious subject while paying homage to the noises of the early twentieth century avante-garde. These speculations could go on indefinitely; simply put, the spirit of the project is to create a new sonic environment in which to conduct experiments that test the limits of the traditional, literary subject. Therefore, I will turn to software engineering design requirements of symposia.

Design: Speech Synthesis

From the initial discussion of speech sound production, a set of requirements can be enumerated for the text to speech solution for the ensoniment of Plato's *Symposium*. At its core is a means of text to speech synthesis of both ancient Greek and English. Speech synthesis systems employ a number of basic methodologies that bear a striking resemblance to their earlier mechanical counterparts described earlier. Concatenative synthesis strings together recorded samples of actual human speakers, often divided into small units stored in a database. Formant synthesis eschews real speech samples for rule-based, additive synthesis of generated tones. Articulatory synthesis techniques model the human vocal tract. While the former seems more akin to the tympanic, ear-based, vibratory model of sound, where rule-based combinatory time and frequency domain mixing occurs within an overall sonic envelopment defining volume, pitch, and velocity parameters, the latter hearkens back to models of sound inspired by the vocal tract. Various commercial software applications including those built into Microsoft Windows operating systems implement these techniques. In the spirit of Reddell's final appeal to open source webmixing, artwork that includes practical information about how to duplicate and transform the medium to encourage civic involvement rather than mere consumerism, I will

impose as a selection criterion that all components of the symposia project are free, open source products, whose program source code is publicly available, part of community-based projects in which anyone may participate or adapt for their own purposes.

Anticipating a multi-threaded, object-oriented programming model and combinatory mixing of text to speech processes, the eSpeak project (http://espeak.sourceforge.net/) has been selected. According to its project web site, it is a formant-based program written in the C language capable of text to speech synthesis dozens of languages from of Afrikaans to Welsh. Critical for symposia, there is also provisional support of Ancient Greek and Latin (see http://espeak.sourceforge.net/languages.html). Variable voice attributes pertinent to generating variations of the default intonation to achieve pheno-song and geno-song characteristics include pitch (base and range), formant ("systematically adjusts the frequency, strength, and width of the resonance peaks of the voice"), echo, tone, flutter ("adds pitch fluctuations to give a wavering or older-sounding voice"), roughness, voicing, consonants, and breath (see http://espeak.sourceforge.net/voices.html). Language attributes can also be adjusted and tuned, including the phonemes, dictionary files, conditional dictionary rules, special phoneme replacements, relative stress lengths and amplitudes of vowels, as well as the character set used to parse input text. Thus, from the humble default voice experiments reviewed in this text, it is anticipated that considerably more complex refinements can be made for the ensoniment of Plato's *Symposium*.

The important eSpeak API function calls for basic text to speech synthesis are defined in its C header file *speak_lib.h*, which must be included in the *symposia.cpp* source. They are espeak_Initialize(), espeak_SetVoiceByProperties(), espeak_Synth(), and espeak_Synchronize(). The sample code shown causes the opening sentence to be spoken in ancient Greek or English, depending on the value of the language parameter passed to the ensound() function:

```
#include "speak_lib.h"
int ensound(int rate,int volume,int pitch,int range,char * voice,char * language,bool block)
{
   if(strcmp(language, "grc") == 0)
      initial_utterance = "δοκῶ μοι περὶ ὧν πυνθάνεσθε οὐκ ἀμελέτητος εἶναι.";
   else if(strcmp(language, "en-us") == 0)
      initial_utterance = "Concerning the things about which you ask to be informed I believe
                           that I am not ill-prepared with an answer.";
   sample_rate = espeak_Initialize(AUDIO_OUTPUT_PLAYBACK, 0, data_path, 0);
   espeak_error = espeak_SetVoiceByProperties(&voice_select);
   espeak_error = espeak_Synth(initial_utterance.c_str(),(int)initial_utterance.length()+1,
                               0, espeak_position_type, 0, synth_flags, NULL, NULL);
   if(block)
      espeak_Synchronize();
}
```

As noted above, further tuning of eSpeak to achieve pheno-song and geno-song effects is beyond the scope of this proof of concept. Additional native eSound API parameters (set by the function espeak_SetParameter()) that regulate the velocity (speed) of synthesis, volume, characteristic pitch, and pitch range include espeakRATE, espeakVOLUME, espeakPITCH, espeakRANGE. These is also espeak_SetVoiceByProperties(), which allows selection of one of the many simulated voices coded into the eSpeak library. For now, the default voice will suffice to demonstrate various many-at-once speaking phenomena to be featured in symposia.

Design: Surround Sound Audio Environment

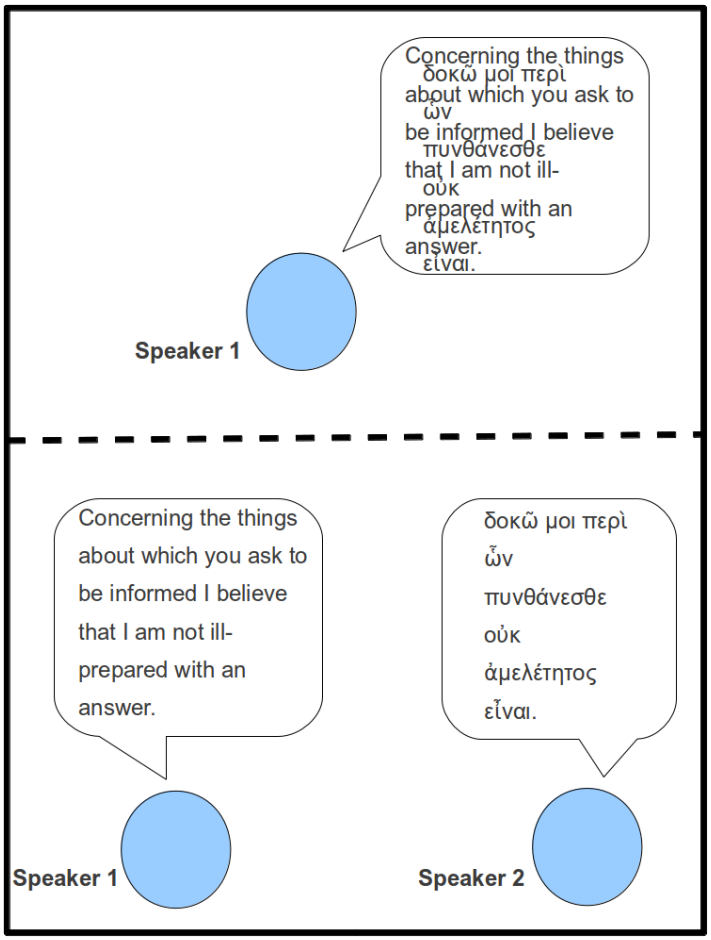
As a proof of concept I coded the *symposia.cpp* program to optionally produce two concurrent voices synthesizing the same text in two independently specified languages by forking the main process originated from the command line. Both voices emit from the same loudspeaker.

```
   pid_t * pids = new pid_t[MAX_VOICES];
   for(int i = 0; i < MAX_VOICES; pids[i] = 0, i++);
   pids[0] = fork();
   speakers++;
   if(pids[0] == 0)
   {
      ensound(rate,volume,pitch,range,voice,first_language,true);
      exit(0);
   }
   if(second_language)
   {
      usleep(1000000);
     pids[1] = fork();
      speakers++;
      if(pids[1] == 0)
      {
         ensound(rate, (volume/4)*3, pitch, range, voice, second_language, true);
         exit(0);
      }
```

```
}
for(int i = 0; i < speakers; i++)
{
   int status;
   while (waitpid(pids[i], &status, 0) == -1);
   if(!WIFEXITED(status) || WEXITSTATUS(status) != 0)
   {
      fprintf(stderr, "Process #%d PID [%d] failed\n", i, pids[i]);
      exit(1);
   }
}
```

The result was difficult to understand regardless of the offset spacing between words spoken by each voice. It mixes the two speech syntheses in the same acoustic space, as if they were uttered from the same mouth. **It is possible this is a flaw only of formant synthesis where all voices liquidate into the same grain.** This technical limitation of personal computer desktop sound systems for which surround sound has not become a cultural expectation the way it has in the home theater might also reflect the ocularcentric bias of virtual reality that Davis and Goodman argue against, or as a minor extension, a visual plus *literary* bias of the one-text-at-a-time that reading practice demands.[4] Two voices coming from the same speaker, like two pages superimposed upon one another, do not multiplex well; multiple speakers seem like a better approach, as Figure 2 illustrates.

4 A fascinating foray into the philosophical debate over the one-at-a-time bias for visual reading versus the many-at-once potential of listening can be read from Heidegger's *What Is Called Thinking?* in which he mulls over the differences between the Being of beings as *legein* and *noein*. Joining this contemplation I have also come across an interesting passage in Aquinas, "*intellegere unum solum, scire vero multa*" that invites comparison between the many-at-once nature of image perception versus the unitary, sequential perception required for reading as a basis of print-based intelligence.



This image shows the jumbled perception of voices when coming from the same speaker. Ensoniment of Plato's *Symposium* when fully implemented has ten speakers. Basically, making a play on Christopher Small's work, symposia requires a place for *sounding*.

Design: Programming New Media

The next figure contains screen shots showing two distributed speakers started in synchrony on separate computers each with at least one loudspeaker drivable from software. Note the values of the current time in each instance. In the following shell session command line transcripts, the symposia executable is run on two computers (ubuntu and alkibiades) within a few seconds of each other while the clocks on both read the same minutes value. This causes

dual voices to begin speaking together on the separate loudspeakers, one in ancient Greek, the other in English.

```
ubuntu@ubuntu:~/src/symposia$ ./symposia
Usage: symposia [speaker [rate [volume [pitch [range [voice [language1 [language2]]]]]]]]
use -1 for default integer values, 'default' for voice
time is 00:26:13
sleeping [47] timeout = (60 - seconds) = (60 - 13) = 47
ensound: initial utterance (length 96): δοκῶ μοι περὶ ὧν πυνθάνεσθε οὐκ ἀμελέτητος εἶναι.
```

```
jbork@alkibiades:~/src/symposia$ ./symposia -1 -1 -1 -1 default en-us
Usage: symposia [speaker [rate [volume [pitch [range [voice [language1 [language2]]]]]]]]
use -1 for default integer values, 'default' for voice
time is 00:26:15
sleeping [45] timeout = (60 - seconds) = (60 - 15) = 45
ensound: initial utterance (length 109): Concerning the things about which you ask to be
informed I believe that I am not ill-prepared with an answer.
```

Leveraging the affordances of free, open source software, in particular related to the standard GNU time library and NTP implementations, the logic required to start the speakers in the correct temporal sequence from the perspective of any listener is easily coded. Relying on synchronized time information, both programs can make the simple computations I wrote in this early version of the C/C++ program *symposia.cpp*:

```
   /* 20111206 wait to start until agreed upon time hoping NTP operative */
   /* start based on calculating number of seconds until before the next minute */
   /* assuming all speakers are started during the same minute on the synchronized clocks */
   struct tm now_tm;
   time_t now;
   int seconds, timeout;
   time(&now);
   localtime_r(&now, &now_tm);
   seconds = now_tm.tm_sec;
   timeout = 60 - seconds;
   printf("time is %02d:%02d:%02d\n", now_tm.tm_hour, now_tm.tm_min, now_tm.tm_sec);
   printf("sleeping [%d] timeout=(60-seconds)=(60-%d)=%d\n", timeout, seconds, timeout);
   sleep(timeout);
(ellipsis)
      ensound(rate,volume,pitch,range,voice,first_language,true);
```

To know how to do this with the “broken-down time . . structure tm,” the built in man page (GNU manual page) utility system program accomplishes a practical mini lesson in C programming. In addition to adding #include <time.h> earlier in the source code logical arrangement, use time(), localtime_r(), and the operator-() (subtraction) function overloaded for int (integer) data type operands. Here is the output of “man localtime_r” on a GNU BASH command line request:

```
Broken-down time is stored in the structure tm which is defined in <time.h> as follows:

    struct tm {
        int tm_sec;         /* seconds */
        int tm_min;         /* minutes */
        int tm_hour;        /* hours */
        int tm_mday;        /* day of the month */
        int tm_mon;         /* month */
        int tm_year;        /* year */
        int tm_wday;        /* day of the week */
        int tm_yday;        /* day in the year */
        int tm_isdst;       /* daylight saving time */
    };
```

It is simply subtracting the current value of seconds (tm_sec) from 60, assuming the programs were started within same clock minute, as did occur in the cases recorded above (**00:26:13** and **00:26:15**). The first process sleeps 47 and the second 45 to start at within one second of each other to produce the many-at-once effect via distributed discrete loudspeakers instead of coming out of one.

Further coding is in order to handle a new argument (parameter) for the speaker number value (1-9) to produce the auditory field depicted in Figure 1 for the ten units required for the full implementation of this live version of the project. The heart of the system is a means to trigger each distributed speaker to synthesize the appropriate text fragments over the course of the running of the main control program. Therefore, the next major development will be to virtualize the physical speakers into a virtual reality environment or game engine that generates a navigable space such that, at a given listening position, the omnidirectional auditory field of concurrent voices is mixed to produce four or six channel surround (or one more if including a subwoofer) sound effects. Some of the free, open source projects that may prove suitable for this include: Simple Direct Media Layer multimedia library (http://www.libsdl.org/), OpenAL 3D audio API (http://connect.creativelabs.com/openal), OGRE 3D Graphics Engine, Id Tech / Quake game engine, and others.[5] The concurrent eSpeak audio streams will be integrated into these applications via call-back functions or other supported interfaces. The bulk of the custom

5 I thank my colleague Ron Weidner from Toptech Systems for these suggestions drawn from his programming experience developing 3D games.

programming will be to produce these audio streams from various encodings of Plato's Symposium, as well as other texts like commentaries, introductions for scholarly translations, and other resources to create a new form of sonic electronic literature in the spirit of the installations Reddell surveyed. Radical re-encoding of the raw input texts into a markup language like Speech Synthesis Markup Language (SSML) or the Text Encoding Initiative (TEI) will call for careful attention to subtleties of audition and speech that have only begun to be touched in the theory section of this work, and would become a fundamental part of the project philosophy. Please visit the project web site (http://symposia.sourceforge.net) for future updates.

Works Cited

Altman, Rick. "The Sound of Sound: A Brief History of the Reproduction of Sound in Movie Theaters." *Cineaste* 21.1-2 (1995): 68-71. Print.

Barthes, Roland, and Stephen Heath. *Image, Music, Text*. New York: Hill and Wang, 1977. Print.

Davis, Erik. "Acoustic Cyberspace." 1997. Web.

Goodman, Steve. *Sonic Warfare: Sound, Affect, and the Ecology of Fear*. Cambridge, Mass: MIT Press, 2010. Print.

Hayles, N. Katherine. *Electronic Literature*. Notre Dame: University of Notre Dame Press, 2008. Print.

Hayles, N. Katherine. *My Mother Was a Computer: Digital Subjects and Literary Texts*. Chicago: University of Chicago Press, 2005. Print.

Ihde, Don. *Listening and Voice: Phenomenologies of Sound*. New York: SUNY Press, 2007. Print.

Kahn, Douglas. *Noise, Water, Meat: A History of Sound in the Arts*. Cambridge, Mass: MIT Press, 1999. Print.

Keller, Evelyn Fox and Christine R. Grontkowski. "The Mind's Eye." *Discovering Reality: Feminist Perspectives on Epistemology, Metaphysics, Methodology and the Philosophy of Science*. Eds. Sandra Harding and Merrill Hintikka. Dordecht, Holland: Reidel Publishing, 1983. Print.

Kittler, Friedrich. *Discourse Networks, 1800/1900*. Stanford, Calif: Stanford University Press, 1990. Print.

Kittler, Friedrich. *Gramophone, Film, Typewriter*. Trans. Geoffrey Winthrop-Young and Michael Wutz. Stanford: Stanford University Press, 1999. Print.

Levin, Michael. "Introduction." *Modernity and the Hegemony of Vision*. Ed. David Michael Levin. Berkeley: University of California, 1993. Print.

Licklider, J. C. R. "Man-Computer Symbiosis." *The New Media Reader*. Ed. Noah Wardrip-Fruin and Nick Montfort. Cambridge, Mass: MIT Press, 2003. Print.

Plato, , Harold N. Fowler, W R. M. Lamb, and Robert G. Bury. *Plato: With an English Translation*. London: W. Heinemann, 1917. Print.

Reddell, Trace. "The Social Pulse of Telharmonics: Functions of Networked Sound and Interactive Webcasting." *Cybersounds: Essays on Virtual Music Culture*. Ed. Michael D. Ayers. New York: Peter Lang, 2006. Print.

Sourceforge.net. *eSpeak text to speech*. 2011. Web.

Sterne, Jonathan. *The Audible Past: Cultural Origins of Sound Reproduction*. Durham: Duke University Press, 2003. Print.

W3C. *Speech Synthesis Markup Language Version 1.0*. 2011. Web.

Whittleton, D. and T. Corkerton. "A Computer Environment For Surround Sound Programming." *IEEE Colloquium on Workstations Moving into the Studio*. London, 24 Nov 1994.