Composing Musical Spaces
By Means of Decorrelation of Audio Signals

Horacio Vaggione

Université de Paris VIII
Centre de Recherche Informatique et Création Musicale (CICM)
Horacio.Vaggione@univ-paris8.fr

Since the seminal work of John Chowning on the simulation of moving sound sources (Chowning 1971) there has been in computer music research a tremendous amount of work concerning perceptual sound space definition. Chowning's model implemented some complementary cues regarding localization, distance, and movement, among which we can recall (1) an interaction between spectral brightness and loudness; (2) a Doppler shift, well known technique for simulating speed and movement in terms of time-varying frequency rates; and (3) a control over the Azimut, the horizontal plane regarding physical separation between channels. Chowning himself referred at the last year's DAFx Conference (Chowning 2000) to the first aspect (the second being already well understood). I will deal myself this year with another aspect apparently related to the third point, concerning inter-channel temporal decorrelation of audio signals.

I do care specially about this subject because it constitutes for me, as a composer, an important aspect of my own concern with layered musical activity arranged in a variety of space perspectives (Vaggione 1989, 1998; Budon 2000). This aspect underlies a detailed articulation of sound objects and textures, which can be enhanced through techniques controlling degrees of temporal decorrelation of waveforms in a multi-channel setting. Moreover, decorrelation is actually embedded, in one way or another, in many sound spatialization systems (see for example Lindemann 1986). Hence one of the goals of this presentation is to uncover the role of decorrelation, besides showing my personal use in electroacoustic music composition.

Decorrelation can be realized by direct manipulation of waveforms with the help of any sound editor program featuring an extended zooming facility, or more algorithmically by means of convolution and FIR or IIR filters. Phase specification can be used for the synthesis of decorrelated signals. Offsets between channels can be straight or interpolated. They can also be controlled dynamically by means of functions stored in an array of look-up tables (Vaggione 1984).

To introduce the subject I will like to present an example taken from one of my electroacoustic pieces: *Agon*, for multi-channel tape (Vaggione 2000). I hope it will illustrate the features of the dynamic sound images which are postulated here: a multiplicity of layers belonging to different time scales, merged in a kind of virtual soundscape (Sound example 1).

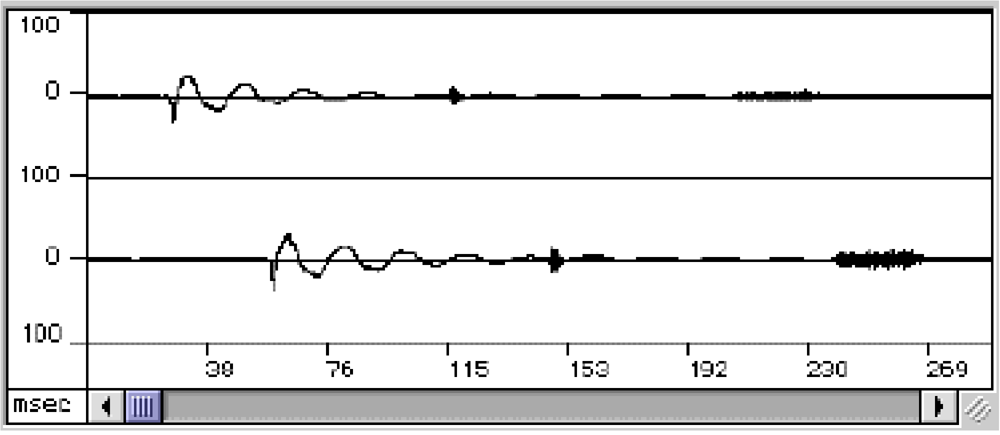


Figure 1: Final gesture of *Agon.*

Figure 1 shows the waveform display of the final gesture of the piece, where an object (composed of three different figures) is replicated in a second channel and decorrelated with a 31 msec offset. Note the slight differences with respect to phase relationships between channels, as well as to respective global amplitude. This gesture flows very quickly, during 270 milliseconds. It would not have the same musical effect without decorrelation (Sound example 2).

The next example, taken from near the beginning of *Agon*, features the same gesture but enriched by another superposed gesture that adds not only more morphological diversity, but also a richer interplay of space activity, including asymmetric inter-channel crossed movements (Sound example 3, Figure 2):

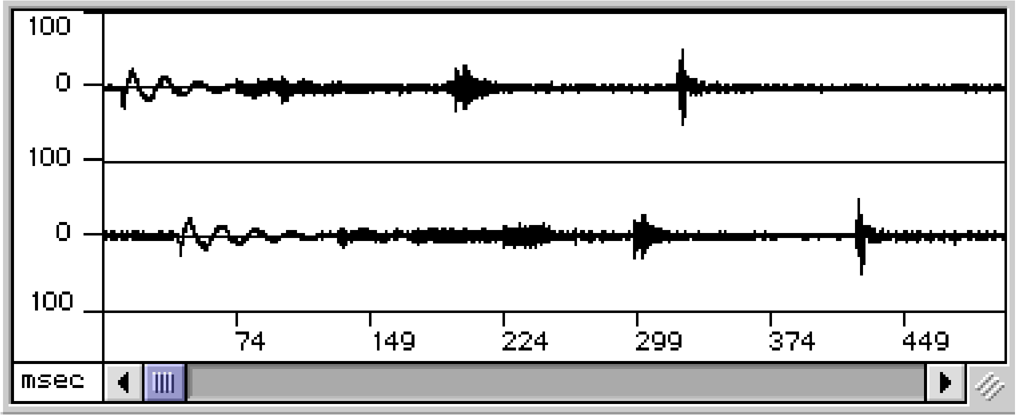


Figure 2: Another gesture from *Agon.*

Realizing the Importance of Decorrelation Through Algorithmic Sound File Mixing

I become myself aware of the importance of temporal decorrelation of audio signals for music composition while working in an Electroacoustic Studio situation. The use of performable diffusion systems based on arrays ("orchestras") of multiple loudspeakers contributed to develop this insight. I learned then to accommodate my perception to slight variations of synchronization of multi-channel mixed signals. Moreover, having used computers since the 1970s (specially the Music-N family of programs), I was confronted with the problems and advantages of pure digital sound file processing, where a "note" statement could instantiate many sound files in precise ways, but at the prize of controlling heavily the amplitudes and the phases of the sounds to be mixed, in order to avoid clipping and phase distortion. This lead me to define algorithmic mixing procedures (Vaggione 1984, Roads 1996). Paul Lansky was at that time also involved with algorithmic mixing (Lansky 1990). Control over the decorrelation of audio signals arise naturally from this algorithmic file handling where synchronization was of paramount importance. I realized then that slight deviations in the scheduling of data rows, specially when handling in the time domain many different sound files simultaneously, produced a sensation of space.

About Simultaneousness

In music, as in everything else, there is not temporal coincidences other than relative: we perceive simultaneities because our perception is not fast enough as to hear microscopic temporal differences. These time intervals are indeed acting and influencing our perception of musical facts. Speaking roughly, «any micro-temporal analysis of onset times for supposedly simultaneous attacks in musical performance would reveal asynchronisms on the order of dozens if not hundreds of milliseconds (this...is exploited by the Musical Instrument Digital Interface (MIDI) protocol in which simultaneous musical events are impossible, even in chords!)» (Roads 2001). I suppose however that this impossible simultaneousness is not only manifested at the macro scale of MIDI definition, but that is pervasive all along the human perceivable temporal range, and, moreover, that it is a positive factor in making music sounding "alive". S. MacAdams reminds that "tone onset asynchrony is a useful technique in musical practice for distinguishing certain "voices", and it is obvious that this cue is used with great versatility by many jazz and classical soloist" (McAdams 1984). McAdams cites research conducted by R. Rash (1978, 1979) describing "how asynchronization allows for increased perception of individual voices in performed ensemble music, which also may be used in "multi-voiced" instruments such as guitar and piano. Across these studies, asynchrony values in the range of 30 - 70 msec have been found to be effective in source parsing" (McAdams, op.cit.). On another register, Gerald Strang have stressed, many years ago, the necessity of incorporating "imperfection" in computer music (Strang 1970), referring mainly to the "dry, boring nature" of fixed (periodic) waveforms, something that Risset was trying at the same time to overcome by articulating microtime inside spectra (Risset 1969). We can generalize this need of "imperfection" (i.e., of decorrelation) to temporal intervals of any size.

Some Perceptual Thresholds

Our experience (see for example Green 1971, and, for an overview about the subject, Roads 2001), indicates that clicks of few milliseconds can be perceived as having already a spectral content, as well as a global intensity. But, more interesting, these clicks can be already perceived as decorrelated sources creating spatial sound images.

Let's introduce inter-channel temporal decorrelation in the most simple way: by straight waveform manipulation. Consider a single sound produced by a Spanish castagnette, lasting 150 msec (Sound example 4). In order to verify the thresholds mentioned above we can select a portion of the attack lasting 2 msec - very short, but still bearing a spectral content (Sound example 5):

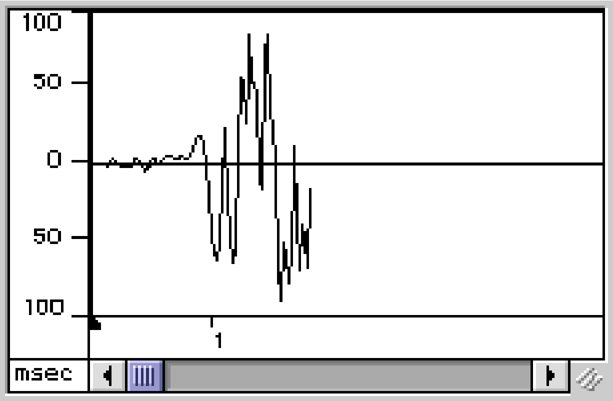


Figure 3.

and then replicate this grain in a second channel and decorrelate the two signals of 1 msec
(Sound example 6):

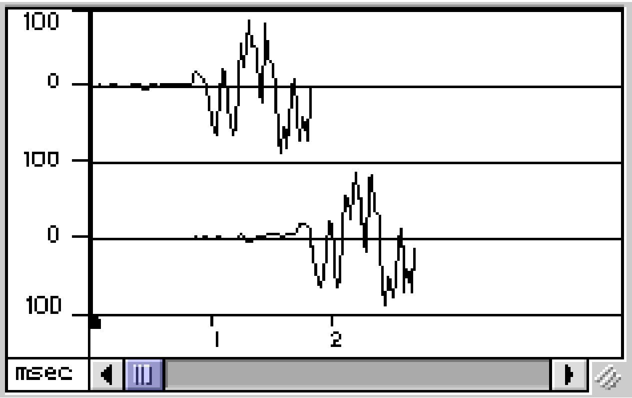


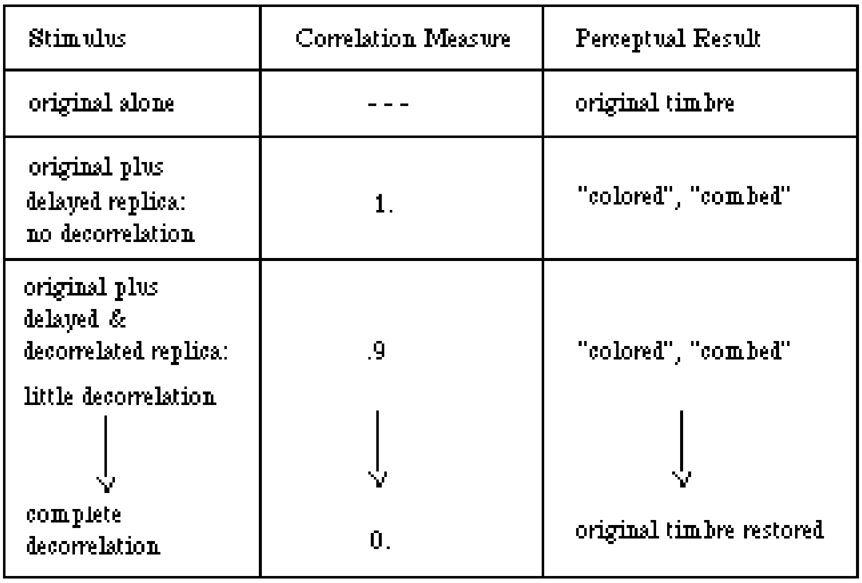
Figure 4.

We can continuing in this way decorrelating the same grain with different, increasing offsets, but under the condition of maintaining two constraints: (1) a stereo setting, and (2) an "unary shape" (no repetitions); if these constraints are not respected, we will switch from the time-space domain to the frequency domain (see next section). In terms of space, the 2 msec grain sounds completely "dry". However, when replicated and decorrelated by an interval of about 3 msec, we are introducing space perception, that is, creating a field where sounds are localized in more than one point in the listening space (Sound example 7). This constitutes an elementary example of what is called here inter-channel temporal decorrelation.

Now, if we increase the inter-channel decorrelation more substantially, another threshold arrives: not only we can hear "space", but we begin to perceive "directionality", that is, movement (through the Azimut, the horizontal plane) inside the sound image. This is a quite obvious effect, but nonetheless very important for our purpose here. The next examples will take again the full source sound (Sound example 4) and proceed to perform various degrees of decorrelation, up to a 48 msec offset (Sound example 8), where directionality can be evidently grasped (I recall that this is a straight example, without any interpolation, filter or analysis/synthesis devise controlling the trajectory between the channels).

The Space Domain

We must stress again and again the fact that the creation of space by decorrelating audio signals is effective when the decorrelated replica are placed in different channels, that is, when they refer to an inter-channel relationship. As we will see in a moment, a multi-channel setting, where many different signals are decorrelated at different rates, is the most effective. Of course, the examples shown here are stereo mix-down of multi-channel originals, but they still work as expected: as complex waveforms built up of several decorrelated layers. Indeed, stereo is here the minimum setting: in the case of a monophonic output, as I said, decorrelations will not cause space perception but a series of effects mostly related to the frequency domain: the signals will be "colored" or "combed" (as in flanging, etc.). Figure 5, taken from Kendall (1995), summarizes these perceptual results.



Stimulus	Correlation Measure	Perceptual Result
original alone	- - -	original timbre
original plus delayed replica: no decorrelation	1.	"colored", "combed"
original plus delayed & decorrelated replica: little decorrelation ↓ complete decorrelation	.9 ↓ 0.	"colored", "combed" ↓ original timbre restored

Figure 5.

Decorrelation of musical signals as a technique to create perceptual space is well known in audio industry: stereoizers and all kind of spatializators are working on the principle of a monophonic input signal which is replicated without changes and then decorrelated and routed to different channels. Gary Kendall, cited above, have apply the principle in very sophisticated ways (Kendall 1994, 1995). He used phase specification in order to perform a FFT analysis and an IFFT to get controlled decorrelations, creating libraries of FIR (Finite-Impulse-Response) filter coefficients to get precise measures (from .9 to 0.; see Figure 5). He also used IIR (Infinite-Impulse-Response) filters to create "dynamic decorrelation", by far more interesting, as Kendall says, because "dynamic variation produces a spatial effect akin to the sound of an environment with moving reflecting surfaces" (Kendall 1995).

We can recall here the work done at CNMAT's Sound Spatialization Theatre about Volumetric Modeling of Acoustic Fields (Kaup et al., 1999), where the authors incorporate, among other things, decorrelation techniques as control features of both magnitude and phase spectra, in order to make equal signals "differ in the speakers where the vector panning is operative", reducing in this way the "precedence effect" (Haas 1951) and improving the perception of volumetric sound projection for all people in the audience, no matter where they are seated. They acknowledge the fact that "decorrelation techniques give rise to considerable ambiguity as to the location of the source", concluding that "here appears to be a real trade-off between the enveloping nature of the spatial audio experience and the precision of the localization" (Kaup et al., op. cit.).

Phase Distortion

Inter-channel decorrelation is certainly not a product of a negative phase value. A situation where many decorrelated signals are summed up is not causing necessarily phase distortion phenomena. When a significant phase distortion comes out, this is a sign that at least one of the inter-channel decorrelated signals have lost phase coherence. This is why is important, when working in an Computer Music Studio, to have continuous access to information about global phase status by means of a phase correlation measuring display. Once the source of phase distortion is detected, we can slightly move the signal in the Azimut inter-channel plane looking for the point where the distortion occurs, in order to suppress or to attenuate its negative value; usually this takes a very small offset quantization, so small that the decorrelation effect does not suffer any significant alteration. Moreover, many moments where slight negative differences in phase relationship occurs can be kept without any post-correction, as they appear in quick time-varying situations, contributing in fact to enhance the dynamic effect of decorrelation (see Figure 1 for an example).

Panning, Delay, and Decorrelation

Decorrelation, as described here, is of course very different from simple panning. The last aims to positioning sounds in a stable field, and also to make them moving from one channel to another, but always inside this stable field. By "stable field" I mean a space where inter-channel settings are not subject to time-varying decorrelation. However, decorrelation can work together with panning, as in the CNMAT's Sound Theatre "vector panning" cited above (Kaup et al. 1999; but see also Pulkki 1997). In this particular case, the purpose of using decorrelation is to break the "precedence" effect: the "first wavefront" (Lindemann 1986) which makes the sound space image collapsing if the audience is too near to one particular loudspeaker. Kendall have demonstrated that decorrelation preserves the sound space image, no matter where the audience is seated. This is something that panning alone cannot do. In fact, panning only does source positioning, fixed or mobile, but without defining the space itself, that is, without altering and controlling the inter-channel offsets of the signals.

Concerning delay: as we saw already, delay alters often (specially when the offsets are small) the frequency content of the signal, unless inter-channel decorrelation is also involved. Delay becomes decorrelation from the moment that control is allowed over inter-channel offset times, which includes a constantly updated information about phase's status (an information that delay does not need).

Some algorithms of the GRM Tools package (Favreau et al., 1999; Teruggi 2001), for example, where access to time offsets between delay lines as well as to separate channel assignments are possible, can achieve space definition. Starting from here, the interesting thing would be to obtain time-varying control of offsets, along of time-varying routing of decorrelated signals through multi-channel arrays, as to get precise control over multi-layered (polyphonic) textures. The plug-in version of this package allows something of this, while depending of the flexibility of the host program (normally a multi-channel graphic sound file editor).

Multi-local Strategies

However there are other options that may correspond to diverse compositional strategies. As I have already said, decorrelation can be controlled dynamically by means of functions stored in an array of look-up tables, or in a phase specification library. These functions can be seen as "attractors", linear or non linear. They can be currently implemented using open packages as, for example, MSP, SuperCollider, or OpenMusic. I have used since years some interpolation algorithms based on polar coordinates, starting with some designed by Stephen MacAdams in 1982, which contained specifications about several spectral configurations. I managed to include in these specifications some data affecting waveform decorrelation to control the spectrum's space definition (McAdams and Wessel 1981, Vaggione 1984). My piece *Fractal C* (1984) was built algorithmically by defining inter-channel decorrelation as changing offset rates in a "multi-local" fashion, in order to give to each "unary shape" (belonging to any time scale: spectrum, sound object, texture) a particular, "scintillating" spatial quality (Vaggione 1989).

"Multi-local" strategies have been later developed in Fractal Theory under terms like "multi-resolution", "multi-fractal", etc., sometimes using wavelet-like representations (see for example Arneodo 1996). These strategies are evidently based on a multi-scale approach, but not necessarily showing the typical properties of rough self-similarity. In a somewhat parallel way, I have myself used the concept of "singularity", in the mathematical sense, to mean the power of single events (morphologies, "unary shapes") to structure a space in which many fractional dimensions are present (Vaggione 1997). Compositional decorrelations of waveforms can be based on these concepts, if controlled dynamically, that is, not concerning single global processes but "multi-local" ones.

Spatial Polyphonies

In any case, inter-channel temporal decorrelation of audio signals constitute an interesting way of working in the direction of a musical situation developing an approach to "poly-spatiality".

Let me introduce another musical example, taken from my electroacoustic work *Schall* (Vaggione 1995). This fragment is based on a diatonic glissando performed on an acoustic piano and sampled as to be processed by diverse means. What interest me to show here is the constant interplay of layers having different decorrelation offsets (Figure 6, sound example 9). Note that this interplay not only creates a spatial activity in the horizontal dimension (which diverse rates of speed and channel's splits positions resulting in crossed sound trajectories) but also in depth (in the perception of the far and the near). Hence layering decorrelated textures constitutes also a mean of controlling perceptual image distance.

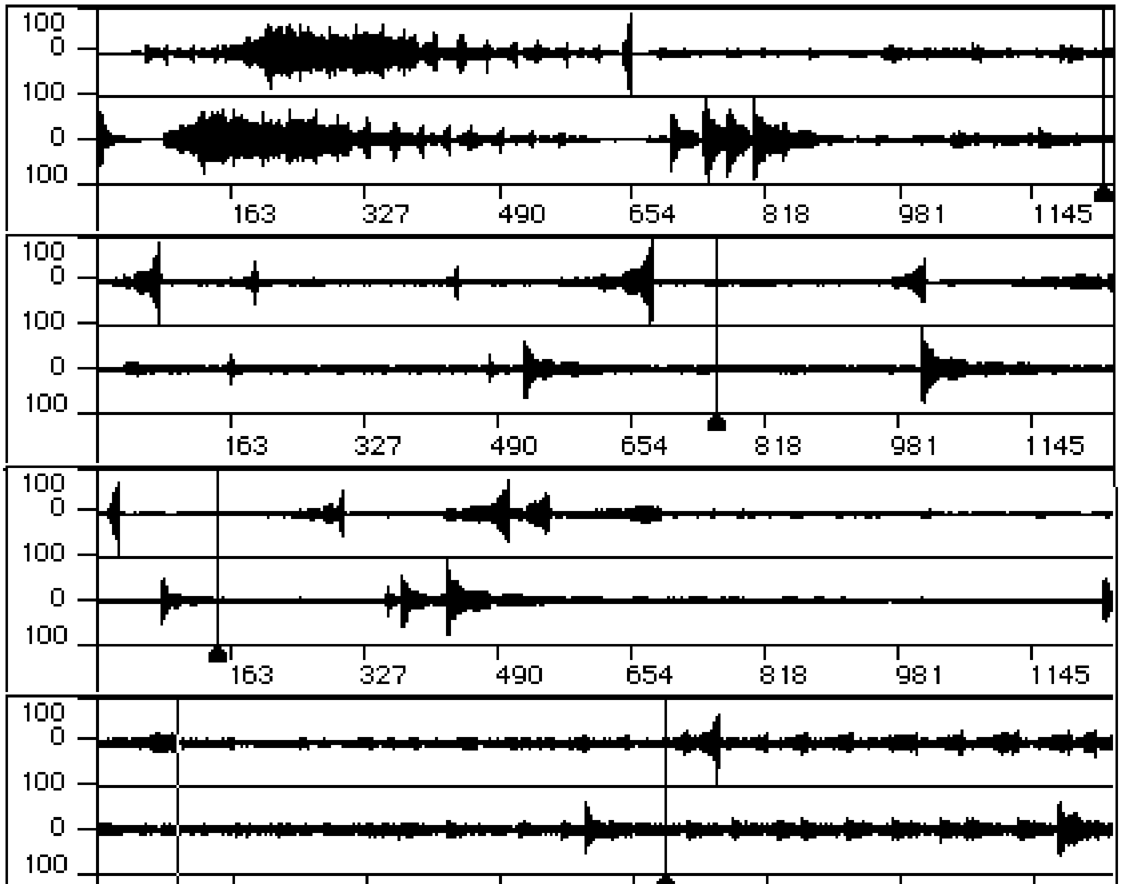


Figure 6 (example from *Schall*).

Conclusion

I have referred in this presentation to a simple and straightforward (but somewhat difficult to conceptualize) way of defining musical spaces. Temporal decorrelation of audio signals are imbedded in many spatialization systems. However, my point have been to show its direct use in electroacoustic music composition. Starting with the very fact of an impossible simultaneousness, and the assessment that music benefits of this fact to create "alive" sounds, I pointed out that our perception is sensitive to very slight temporal decorrelations. The difference of delaying with monophonic output (which leads to operations in the frequency domain) was stressed in relation to inter-channel decorrelation (which leads to operations in the space domain). Diverse ways of controlling inter-channel decorrelation were mentioned (straight waveform manipulation, interpolation, variable look-up table design, convolution, FIR and IIR filters, and direct phase control through FFT/IFFT). Finally, I have recalled that this approach is interesting for electroacoustic music composition when used in a multi-layered approach, where diverse channels with diverse time-varying inter-channel decorrelation values create not only a perceptual diffuse space, but also a spatial depth and a multi-directional dynamics. Combining decorrelated audio signals is a mean to create complex sound space images, internally articulated and controlled, aiming to approach the idea of composed poly-spatiality.

References

Arneodo, A. (1995) *Ondelettes, multifractales et turbulences.* Paris: Diderot.

Budon, O. (2000) "Composing with Objects, Networks, and Time Scales: An Interview with Horacio Vaggione". Computer Music Journal 24 (3): 9-22. Cambridge: MIT Press.

Chowning, J. (1971) "The Simulation of Moving Sound Sources". Journal of the Acoustical Society of America 19 (1): 2-6.

Chowning, J. (2000) "Digital Sound Synthesis, Acoustics, and Perception: A Rich Intersection". Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFx-00).

Favreau, E., Racot, G., and D. Teruggi (1999) "Evolution des outils, évolution des idées". *Interfaces homme-machine et création musicale*. Paris: Hermes Science.

Green, D. (1971) "Temporal Auditory Acuity". Psychological Review 78 56): 540-551.

Haas, H. (1951) "Über den Einfluss eines Einfachechos auf die Hörsamkeit Effect". Acustica 1: 49-58.

Kaup, A., Khoury, S., Freed, A. and Wessel, D. (1999). "Volumetric Modeling of Acoustic Fields in CNMAT's Sound Spatialization Theatre". Proceedings of the 1999 International Computer Music Conference, San Francisco: ICMA.

Kendall, G. (1994) "The Effects of Multi-Channel Signal Decorrelation in Audio Reproduction". Proceedings of the 1994 International Computer Music Conference. San Francisco: ICMA.

Kendall, G. (1995) "The Decorrelation of Audio Signals and Its Impact on Spatial Imagery". Computer Music Journal 19 (4): 72-87.

McAdams, S. and D. Wessel (1981) "A General Synthesis Package Based on Principles of Auditory Perception". Paper presented at the 1981 International Computer Music Conference, Denton, Texas.

McAdams, S (1984) *Spectral Fusion, Spectral Parsing and the Formation of Auditory Images.* Ph.d Thesis, Stanford University.

Lansky, P. (1990) "The Architecture and Musical Logic of Cmix". Proceedings of the 1990 International Computer Music Conference. San Francisco: ICMA.

Lindemann, W. (1986) "Extension of a Binaural Cross-Correlation Model by Contralateral Inhibition, II. The law of the first wavefront". Journal of the Acoustical Society of America, 74:1728-1733.

Pulkki, V. (1997) "Virtual Sound Source Positioning Using Vector Base Amplitude Panning". Journal of the Audio Engineering Society 45 (6): 456-66.

Rash, R. (1978) "The Perception of Simultaneous Notes Such as in Polyphonic Music". Acustica 40: 21-33.

Rash, R. (1979) "Synchronization in Performance Ensemble Music". Acustica 43: 121-131.

Risset, J.C. (1969) *An Introductory Catalog of Computer-synthesized Sounds.* Murray Hill: Bell Laboratories. Reissued in *The Historical CD of Digital Sound Synthesis.* Computer Music Currents 13, Mainz, WERGO.

Roads, C. (1996) *The Computer Music Tutorial.* Cambridge: MIT Press.

Roads, C. (2001) *Microsound.* Cambridge: MIT Press.

Strang, G. (1970) "The Problem of Imperfection in Computer Music". Beauchamp, J. and H. Von Foerster (Eds.): *Music with Computers*. New York: Mc Graw Hill.

Teruggi, D. and S. Martin (2001) "L'interpolateur, une interface de contrôle multi-paramétrique". Proceedings of the JIM (Journées d'Informatique Musicale). Bourges: IMEB/ADERIM.

Vaggione, H. (1984) "The Making of *Octuor*". Computer Music Journal 8 (2): 48-54. Reprinted in C. Roads (Ed.): *The Music Machine.* Cambridge: MIT Press (1989).

Vaggione, H. (1989) "Modelle der Unvollkommenheit in der Computer Musik". Symposium Chaos und Ordnung. Graz: Steirischer Akademie.

Vaggione, H. (1997) "L'espace composable. Sur quelques catégories opératoires dans la musique électroacoustique". In Chouvel, J.M., et M. Solomos (Eds.): *L'espace: musique/philosophie.* Paris: L'Harmattan.

Vaggione, H. (1998) "Son, temps, objet, syntaxe. Vers une approche multi-échelle dans la composition assistée par ordinateur". In *Musique, Rationalité, Langage*. Paris: L'Harmattan. Also available in Italian: "Suono, tempo, oggetto, sintassi". In Musica/Realta 60 (1999): 121-151. Milano: LIM.

Vaggione, H. (1995) *Schall,* Electroacoustic composition. CD Chrysopée Electronique, Bourges: Mnémosyne Music Media (http://www.gmeb.org). Distributed by CDeMusic (http://www.CDemusic.org/artists/vaggione.html).

Vaggione, H. (2000) *Agon,* Electroacoustic composition. Berlin ICMC-CD. San Francisco: ICMA. Distributed by CDeMusic (http://www.CDemusic.org/artists/vaggione.html).