



Rhetorical Strategies for Sound Design and Auditory Display: A Case Study

Pietro Polotti ^{1,2,*} and Guillaume Lemaître ^{2,3}

¹ Conservatory of Music "G. Tartini", Trieste, Italy

² IUAV, Venice, Italy

³ IRCAM, Paris, France

This article introduces the employment of rhetorical strategies as an innovative methodological tool for non-verbal sound design in Human-Computer Interaction (HCI) and Interaction Design (ID). In the first part, we discuss the role of sound in ID, rhetoric, examples of rhetoric employment in computer sciences, and existing methodologies for sound design in ID contextualized in a rhetorical perspective. On this basis, the second part of the article proposes new potential guidelines for sound design in ID. Finally, the third part consists of a case study. The study applies some of the proposed guidelines to the design of short melodic fragments for the sonification of common operating system functions. The evaluation of the sound design process is based on a learning paradigm. Overall, the results show that subjects can more effectively learn the association between sounds and functions when the sound design follows rhetorical principles. The case study focuses on a very specific instance in the extensive field of auditory display and sonic interaction design. However, the results of the validation experiment represent a positive indication that rhetorical strategies can provide effective guidelines for sound designers.

Keywords – Auditory Display, Interaction Design, Musical Rhetoric, Rhetoric, Sonic Interaction Design.

Relevance to Design Practice – Computers and interactive devices are supposed to rapidly evolve into communicating agents. This paper draws designers' attention to rhetoric as a rich source of strategies that can be employed in the design of the sonic aspects of interactive applications. A case study illustrates the methodological proposal in a factual way.

Citation: Polotti, P., & Lemaître, G. (2013). Rhetorical strategies for sound design and auditory display: A case study, *International Journal of Design*, 7(2), 67-82.

Introduction

Auditory Display (AD) and Sonic Interaction Design (SID) are fast growing research domains. The former investigates strategies and modalities for providing information by means of non-verbal sounds. The latter deals with the use of sound in ID. New energies have recently been injected into these fields by a number of European-supported research efforts such as the projects S2S² (Sound to Sense, Sense to Sound, see <http://smcnetwork.org/>), CLOSED (Closing the Loop Of Sound Evaluation and Design, see <http://closed.ircam.fr/>), and the European COST-Action Sonic Interaction Design (see <http://sid.soundobject.org/>). Nevertheless, it is commonly held within the International Community for Auditory Display (ICAD, see www.icad.org) that AD and SID lack consolidated methodological guidelines.

This article proposes rhetoric as a rich repository of principles and techniques and a solid methodological base on which to build new knowledge and expertise in the domains of AD and SID. Classical rhetoric deals with economy and effectiveness of communication through speech. The idea of transferring the tradition and experience of rhetoric from verbal language to another realm is centuries old. Since the 16th century, the transposition of classical rhetorical principles and techniques from the verbal domain to the musical has made an important contribution to the transformation of instrumental music into an independent art form. Instrumental music has since been able to develop its own structures and discourse independently of verbal expression, theatrical representation or dance. Inspired by this

example, this study examines the potential to employ musical rhetorical strategies for the design of non-speech audio in the realm of ID and HCI.

We argue that what we propose is congruent with a general trend in ID pointing to an increasingly human-centered and ergonomic technology, both physically and psychologically. In this sense, the re-invention and adaptation of communication techniques developed within pure humanistic frameworks to the context of ID seems to be a coherent and promising research strategy in favor of such a general goal.

Background

Sound in ID

Today, an increasing number of everyday objects and machines include sonic resources in both an active and a passive sense. This is coherent with the simple fact that our perception of the

Received April 22, 2012; Accepted April 14, 2013; Published August 31, 2013.

Copyright: © 2013 Polotti and Lemaître. Copyright for this article is retained by the authors, with first publication rights granted to the *International Journal of Design*. All journal content, except where otherwise noted, is licensed under a *Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License*. By virtue of their appearance in this open-access journal, articles are free to use, with proper attribution, in educational and other non-commercial settings.

*Corresponding Author: pietro.polotti@conts.it

world is inherently multimodal and that users rarely receive information through only one sensory modality. When feasible, ID applications seek to include several perceptual channels at once. Besides visual and haptic aspects, sound has to be systematically considered in order to pursue a more human-like technology (see, for example, Spence & Zampini, 2006, for the importance of sound in product design). In the near future, we expect to see a multitude of technological devices endowed with expressive and listening capabilities in terms of speech and non-speech audio communication. The envisioned acoustic scenario of the future will include thousands of new artificial sounds that will pervade our everyday life and consume the “environmental acoustic bandwidth” for the transmission of auditory information. In particular, we expect a huge proliferation of non-verbal sounds, whose advantage is to be pre-attentive, independent of any specific language and, if properly designed, shorter and more intuitive than verbal messages (Brewster, Wright, & Edwards, 1995). Such acoustical hypertrophy requires adequate action aimed at defining ways to best exploit the communication potentialities of non-speech audio, while avoiding the degeneration of the acoustic environment into an overcrowded soundscape. Strategies for designing artificial sounds in a concise and effective way tackle both these aspects at once by optimizing the communication process and reducing the sonic impact in terms of physiological and psychological fatigue. This represents the opposite of what, in the context of alarm design (see Patterson, 1990), an ambulance siren produces. These issues form the core of the debate surrounding disciplines such as AD and SID (see Hermann, Hunt, & Neuhoff, 2011, for a complete overview on the state of the art of AD and SID).

Rhetoric

Classical rhetoric is defined as the art and technique of persuasion through the use of oral and written language (see, for example, the modern edition of the treatises by Quintilianus, 1920/1996). In Ancient Greece, rhetoric was the queen of the *techné*: the

technology of acting upon minds through speech. Throughout the centuries, classical rhetoric developed rules, strategies, collections of cases, categorizations of subjects (the so-called rhetorical *loci*), and expressions (the so-called rhetorical figures) to provide the speaker with a wide repertoire of tools for supporting their arguments. Intentionally or not, people in general—and lawyers, politicians, philosophers, and writers in particular—have always adopted rhetorical procedures to convince their audience of a certain point of view.

From a technical point of view, rhetoric comprises five parts, which constitute an integrated system for communication: *inventio*, *dispositio*, *elocutio*, *actio*, and *memoria*. *Inventio* relies on the availability of categories of topics classified and organized according to their argumentative function (i.e., the *loci*) and consists in selecting a number of them, contextually relevant and appropriate to the target audience. *Dispositio* deals with the sequential distribution and structural organization of the arguments that will be employed. *Elocutio* concerns the form, the style, and, in general, the actual ways of expression, by means of which the arguments will be presented. *Actio* regards the speech utterance, namely the intonation of the voice, the pauses, the accents, and even the body posture and gestures, especially of the hands. Finally, *memoria* is the cement of everything and allows the orator to effectively dominate the speech in its entirety. The orator must avoid undesired hesitations or, worse still, interruptions. Memory provides the orator with a clear global image of the speech subjects and arguments as a whole and enables him, for example, at any convenient moment to anticipate other parts or recall references to what has already been said. Likewise, it is easier for the audience to memorize a well-structured and figuratively well-characterized speech. This point will be the basis for the validation of the design adopted in our case study.

In the perspective introduced in the middle of the 20th century by Perelman and Olbrechts-Tyteca (1958/1969), rhetoric was intended to mean the study of argumentation, i.e., the science of establishing contacts between minds by means of speech. The two Belgian philosophers direct attention away from the realm of rationalist thinking and logic to that of veridical statements and opinions. Their main idea is that to face the complex, multifaceted, and non-univocal nature of the human world, we need to set aside the universality of formal thinking and language, and rely instead on discourse and argumentation. In the human world, in fact, we cannot show or prove anything; we can only try to convince somebody of the advantages of a choice, or the veridicality of a point of view, by gaining the adherence of the mind of the interlocutor(s). Perelman and Olbrechts-Tyteca call the achievement of adherence between the speaker and the audience as the establishment of the contact of minds. This adherence is central and has to be attained by choosing proper arguments and styles that have to take into consideration not only the specific subject, goal, and context, but also the knowledge, culture, and feelings of the audience.

In such a theoretical framework, the form of the speech has as relevant a role as the coherence and robustness of the arguments. The lexical choices, their connections, the selection of examples, of unusual and novel points of view and expressions,

Pietro Polotti studied piano, composition, and electronic music in the Conservatories of Trieste, Milan and Venice, respectively. He holds a Master's degree in Physics from the University of Trieste. In 2002, he obtained a Ph.D. in Communication Systems from the EPFL of Lausanne, Switzerland. Presently, he teaches Electronic Music at the Conservatory “G. Tartini”, Trieste, Italy. Since 2004, he has been collaborating with the University of Verona on various European sound projects. Recently, his interests have moved towards sonic interaction design and interactive arts. He is part of the Gamelunch group (www.soundobject.org/BasicSID/Gamelunch). In 2008, he and Maurizio Goina started the EGGS project (Elementary Gestalts for Gesture Sonification, see www.visualsonic.eu).

Guillaume Lemaitre graduated from the Université du Maine in France (Ph.D. acoustics 2004, MSc acoustics 2000), and from the Institut Supérieur de l'Électronique et du Numérique (MSc electrical engineering 2000). He has been conducting several sound design projects with INRIA (the French national research institute for computer science) and Ircam in Paris. He has also conducted basic research studies in auditory cognition with the department of Psychology at Carnegie Mellon University (Pittsburgh, PA) and the University Iuav of Venice, Italy. His research activities include psychoacoustics, auditory cognition, perception and cognition of sound sources, sound perception and action, sound quality and sound design, sound interactions and vocal imitations of sounds. He is currently working for Genesis Acoustics, a French company based in Aix-en-Provence where he is in charge of perceptive studies and the design of product sounds.

are the factors that make the thesis of the speaker convincing and the argumentation successful. As strategies of appearance, form and style, emotional response and aesthetics become parameters to play with to establish the effective communication of specific semantic contents. Emotions, in particular, are closely linked to presence. As the treatises of ancient rhetoric emphasize, concreteness and exemplification are central elements for capturing the attention and participation of the audience. Presence, on the other hand, is an inherent property of sound as a physical phenomenon. Reciprocally, in our experience of the world, sound is one of the primary means for perceiving presence, together with movement. The former relates to our auditory perception, the latter to our visual and tactile experience. Moreover, the need to deal with emotions means that rhetoric cannot be reduced to structure or, worse still, to mere ornamentation: rhetoric involves a deep knowledge of the listener's cultural background, psychology and expectations. When using rhetoric in the design of non-verbal sound, we must take into consideration all of these aspects, as well as the fact that we do not aim at *convincing* somebody, but rather at *being convincing* about what we want to represent through sound. Indirectly, this could mean inducing a swifter and clearer understanding of the communicative contents of a sonic message in the listener, and/or prompting (convincing) them to react in a certain way according to a specific sonic event.

Rhetoric had a general revival during the second half of the 20th century. In the last few decades, many studies have aimed at using rhetoric for poetry hermeneutics (see, for instance, Lotman, 1993) or transposing rhetoric to the visual domain (see, for example, Groupe μ , 1999; Rastier, 2001). A revival of a rhetorical approach to music interpretation also involved contemporary music (Benzi, 2004). One of the main aspects emerging from this debate concerns the semantic valence of rhetorical structures. Rhetoric can be viewed as an instrument for the investigation of language and reality, operating at a semantic level, or as a set of ornamental functions with no semantic relevance. Once more, we specify that we refer to Perelman and Olbrechts-Tyteca's (1958/1969) position, which rejects the conception of rhetoric as a mere embellishment tool. Thus, by aesthetics, we mean something that is not superficial, but is related in a synergetic way to the semantic contents of what we aim to represent and communicate. In this sense, our aim is essentially pragmatic; we endeavor to show that rhetorical techniques can be a prominent support in sound design for AD in a factual and concrete way. The design procedure discussed in the case study takes into consideration a plausible semantic correspondence between the rhetorical level and the functional level of an operating system to facilitate the identification of the sound-function associations.

Use of Rhetorical Principles in AI and ID

In the past few decades, rhetoric gained attention within Artificial Intelligence (AI) and computational linguistic research through Mann and Thompson's (1988) formulation of the Rhetorical Structure Theory (RST). The theory met with considerable success (see, for example, Popescu, Caelen, & Burileanu, 2007). RST considers rhetoric as a method of analyzing and defining the structure of language. This is a fundamental aspect of speech

formulation somewhat similar to the second component of rhetoric, the so-called *dispositio*. However, in RST the essential persuasive character of rhetorical practices is missing. As discussed in the next section, this persuasion is instead related to the other two parts of rhetoric, *elocutio* and *actio*, which are applied in the design of the experimental materials for the case study. The fifth and last rhetorical component, *memoria*, will constitute the test bench for the experiment.

Rhetoric is intrinsically connected to dialogue construction. The issue of dialogue definition in human computer systems is an established field of research. In the introduction to a special issue of the *International Journal of Human Computer Studies* on collaboration, cooperation and conflict in dialogue systems, Jokinen, Sadek, and Traum (2000) agreed that "the research paradigm has changed from the view where the computer is seen as a tool, to one which regards the computer as a communicating agent" (p. 867). In the same special issue, an attempt to exploit the nature of rhetoric as the art of convincing can be found in an article by Grasso, Cawsey, and Jones (2000). Their work is specifically devoted to rhetoric and its use in a system providing advice and trying to convince people about nutrition and health issues. The authors, likewise, pointed out that "RST..., usefully and successfully applied in the generation of explanatory texts and tutoring dialogues..., says, however, little about how to generate persuasive arguments, which was our primary interest" (p. 1080). Classical rhetoric was also the source of inspiration of a work on evaluation of websites by De Marsico and Levialdi (2004).

There are many other cases where rhetoric comes into play, even if not so explicitly. One instance is the work on pattern analysis by Frauenberger and Stockman (2009) discussed in the next section. In general, the concepts of metaphor and analogy, which are fundamental tools of rhetoric, have been strongly present in the HCI community ever since the field came into being. An example is the paper submitted in 1984, although only published years later, by Carroll and Mack (1999). In that paper the cognitive power of the open-ended character of metaphors as a stimulus to the construction by analogy of new mental models, was fully expressed in a rhetorical framework. Within the domain of ID, Pirhonen (2005) provides an interesting work about the life cycle of analogies and metaphors. The author argues that the strength of a metaphor depends on the frequency of its employment; the more it is used the more it becomes an ordinary expression, losing its communicative power. In respect of the use of metaphors in the specific domain of AD, it is also worth mentioning the work of Walker and Kramer (2005), which underlines the ambiguity that a sound designer can encounter when trying to build more "intuitive" sonic metaphors.

The Necessity of Guidelines in the Domains of AD and SID

In the fields of AD and SID, defining general guidelines is a crucial issue. Barrass (2005a) discusses the need to go beyond the state of the art in terms of the foundational aspects of the discipline. In another paper, Barrass (2005b) addresses the issue from a perceptual point of view. Perceptual aspects are, of course, one of

the facets of the problem. However, we think that it is crucial to go beyond perception and consider the investigation of the structural and semantic features involved in sound design.

In a recent paper, Frauenberger and Stockman (2009) propose design pattern analysis as a point of reference for the discipline. Their article offers a concise, yet complete overview of the available methodologies, and points out the lack of a unitary and robust framework for the discipline. In particular, the authors highlight how research papers in the SID field usually do not reveal the rationale of their design decisions. As an alternative, they introduce a method based on pattern mining in context space. Being aware of the strengths and weaknesses of a pattern-based approach, they propose using context as an organizing substrate. In their opinion, this would provide designers with useful reference models of situated ID practices and allow rapid interpretation of preexisting examples in order to deal with new problems. The results sound promising and the method promotes the growth of the discipline by building upon previous knowledge. Somewhat along similar lines, Hermann, Williamson, Murray-Smith, Visell, and Brazil (2008) advance the idea of a *Sonic Interaction Atlas*, a big database hierarchically organized into categories providing designers with a powerful and easily accessible repertoire of specimens.

It is noteworthy that both studies reflect one of the keystones of rhetoric, the definition of the above-mentioned *loci*. The *loci* are the result of a classification of the argumentation types according to specific criteria. They provide the orator with a well-organized database of formulas, paradigms, examples and strategies that she or he can browse, select and employ to build a discourse and promote their theses. As already said, this corresponds to the first phase of rhetoric, the *inventio*, that is, the retrieval of argumentative materials that are then selected, organized and employed in the speech.

The aforementioned European project CLOSED has also attempted to define general principles here. The main goal was to define evaluation criteria for use in cyclic iterations of prototyping and validating steps in the framework of a mature sound design discipline (Lemaitre et al., 2009). Far from providing definitive results, the project has highlighted the need for further development of shared evaluation practices and methodologies for SID.

From a sound generation perspective, the work of Drioli et al. (2009), for example, raises the need for powerful and intuitive tools for sound sketching and production in ID. The perceptually-based sound design system introduced in their paper allows one to explore a synthetically-generated sonic space, having at its disposal a number of “sonic landmarks” whose timbres thoroughly permeate the sonic space and could be continuously interpolated. The main goal of the sound design system is to provide designers with a tool requiring no previous knowledge of sound synthesis and processing, but enabling a direct phenomenological approach to sound selection and refinement. The definition and implementation of an equivalent software tool, usable by designers for the production of rhetorically-based audio for interactive systems, is far beyond the scope of the present article. Such a task would involve a tremendous multidisciplinary effort ranging from AI to prosody and musical expressivity synthesis. Nevertheless, this would prove fertile ground for investigation for the future of the present research and a substantial development

in achieving an operational level at which a designer could easily access a palette of sonic-rhetorical tools for generating sounds to be adapted, tested and employed in ID applications.

These examples highlight the efforts made to address the lack of both solid methodological beacons and robust tools for the evaluation and the generation of sounds in the fields of AD and SID. Within this context, our work aims to contribute to the definition of general guidelines by introducing the use of rhetoric-based methodologies into sound design in a way that goes beyond an RST view of rhetoric and recovers the figurative and expressive strength of rhetorical formulas. In a sense, we regard rhetoric as the domain of strategies of appearance able to achieve a convincing and more immediate communication; an “intuitive argumentation” in the spirit of Merleau-Ponty’s (1964) phenomenological approach, when he says that “a movie is not thought; it is perceived (p. 58). Likewise, we look for sounds that are intuitively understandable and immediately *convincing* because of their *auditory appearance*.

Methodological Proposals

Rhetoric and Sound for ID

By way of analogy to the theoretical framework discussed in the section “Rhetoric”, we propose a number of guidelines for sound design in ID. As a primary goal, sound in AD and SID seeks to achieve effective communication. To do so, sound must be

1. appropriate in the context (referable to the rhetorical *inventio*),
2. coherently and effectively organized in time (referable to the rhetorical *dispositio*),
3. carrier of meaning and able to strike the imagination (referable to the rhetorical *elocutio*),
4. expressive and thus able to leverage an emotional response (referable to the rhetorical *actio*), and
5. easily memorizable (referable to the rhetorical *memoria*).

The first point is somewhat analogous to Frauenberger and Stockman’s (2009) already-quoted proposition of pattern mining in context space. The second is to be developed and considered in cases of design of complex AD or SID applications articulated in time, where different elements can be concatenated in different ways. This corresponds to what is denoted as “macro-form” in music. The third regards the choice of concrete musical or sonic elements and their parametric organization at a “micro-form” level. The fourth concerns the correct and effective utterance of the musical/sonic contents to emphasize the stratagems of the sound design and allow a more effective comprehension and communication at a perceptual, cognitive, and emotional level (see Lemaitre, Houix, Susini, Visell, & Franinovic, 2012, for an example of feelings elicited by sounds during the manipulation of a sonified tangible interface). In this context, the fifth is thought of as “passive” memory; while in a speech, the orator has to master the arguments through their memorization, within an AD or SID application, sonic outputs and feedbacks have to be easily learnable to be rapidly memorized and correctly interpreted during use.

Particular expedients that rhetoric provides the speaker with are the so-called rhetorical figures. Rhetorical figures are part of the *elocutio* and are defined as forms and expressions that differ from a regular use of language. A figure exists only when it is possible to distinguish the unfamiliar use of a structure from its normal use in a particular instance and context. On the other hand, the uncommon use that generates the figure is more effective if the distinction fades out through the discourse; the figure progressively integrates into the speech as a coherent and organic element of it, bringing, at the same time, the argumentative strength of the different world the figure refers to. This is the case of analogies and metaphors; the closer a metaphor comes to the represented subject, the stronger its effect. Perelman and Olbrechts-Tyteca (1958/1969) define a rhetorical figure as:

- A structure that is independent of the contents and that can be distinguished from it.
- A structure where its use is remote from the normal employment and hence is able to catch the attention of the listener and to spur their emotional reaction, reasoning, and adherence to the subject of the speech.

However, the balance between novel and ordinary is subtle. The more a metaphor is used, the less effective it becomes (Pirhonen, 2005). Examples of other rhetorical figures that are employed in this work in a musical version, are the *anadiplosis*, the *epanodos* and the *epanalepsis*. The *anadiplosis* figure takes place when the same word is used at the end of a sentence and at the beginning of the following one; e.g., “For this reason, we need the utmost *attention*. *Attention* is, in fact, fundamental in order to ...” The *epanodos* figure occurs when two words are repeated in reverse order—e.g., “One can always *laugh* and *weep* about the world. *Weep* for the misfortunes of life, *laugh*, however, about the ...” The *epanalepsis* figure employs the same expression at the beginning and end of a sentence; e.g., “*Go ahead* in your work, with determination, *go ahead!*”

As a first step in a more general investigation, the experimental focus of this work is on this specific aspect of rhetoric, that is, the rhetorical figures and their transposition and use in the domain of AD. We seek to show that going beyond an RST view of rhetoric and combining a skillful sonic *elocutio* with an effective sonic *actio* can produce a successful sound design practice for simple applications such as that of our case study.

Musical Rhetoric for AD and SID: A Possible Paradigm?

Technically, musical rhetoric provides sound designers with a large set of techniques to create sounds following rhetorical principles. The idea of transferring rhetorical means and strategies to the musical domain by proposing a plausible parallel between verbal and musical structures dates back to some centuries ago. Indeed, this is what happened in music in the past, especially during the 16th and 17th centuries when, for the first time, music was considered as an abstract language able to convey emotions by means of its rhetorically-built structures, in a way similar to the discourse of an orator. It took centuries for music to conquer its autonomous space, independent of both science and text.

During Classical Antiquity and the Middle Ages, music was either an abstract representation of the *Harmonia Mundi* or else the ancillary for poetry, religious texts, and dance. Only in the Baroque era did a novel practice, elicited by multiple social, economic, and cultural factors, free European instrumental music from its subsidiary role. From a technical point of view, the systematic transposition of rhetorical principles to music for the definition of a “language of emotions”, the so-called *Theory of the Affections* (see, e.g., Bartel, 1982/1997; Wilson, Bülow, & Hoyt, 2001), played a fundamental part in this process. A vast repertoire of techniques, practices, examples, and rules inherited from classical rhetoric provided music with the proper means of building its own self-standing forms and expressive content. This set of precepts and guidelines, developed from the 16th up to the 18th century (for an example see Wilson et al., 2001, the modern edition of the work by Gottsched, 1750; Mattheson, 1739/1981), formed a new discipline generally known as musical rhetoric.

As already discussed, rhetoric is the art of successfully achieving human reception by correctly addressing the rational, aesthetic and emotional dimensions simultaneously. These two last aspects are very important. Since the times of the ancient Greek orators, the skillful adoption of figurative rhetoric (the *elocutio*) and the correct and expressive utterance of the speech (the *actio*) have been considered fundamental ingredients of oration, on a par with the robustness of logical reasoning. In particular, the speech had to be properly delivered; prosody, volume, rhythm and pauses had to highlight the rhetorical construction to effectively communicate content and argument. This is closely analogous to the practice of musical rhetoric during the Baroque era, where a piece of music composed according to the above mentioned *Theory of the Affections* had to be performed with the proper expressiveness to make the musical discourse based on rhetorical structures evident and the transmission of “affections” successful. In other words, musical interpretation was a fundamental part of the game (see, for example, Bach, 1753/1949 or Timmers & Ashley, 2007).

In this work, we claim that the principles of musical rhetoric can be transferred to AD and SID. Sounds in AD and SID are non-verbal, uttered by machines and devices, and have to be properly perceived, interpreted and understood by humans for functional purposes. They have a precise semantic content, referable to something external to the sound itself. Therefore, the employment of rhetorical and musical rhetorical strategies leveraging on aesthetic and emotional aspects is particularly relevant. Furthermore, temporal organization of sound events is a key aspect in AD and SID. The analogy with the art of temporal organization of words, expressions and sentences, and, of course, with music is therefore compelling. To a certain extent and in a certain cultural/aesthetical context, at least that of the Western music of the Baroque era, one could think of musical composition as the twin discipline of rhetoric. We propose to extend this parallelism to non-verbal sound design for AD and SID. Therefore, music rhetoric holds both as a paradigm of transposition of verbal rhetorical practices to a non-verbal domain and as a direct source of principles transferable from music composition to sound design.

Earcons in AD and Their Evaluation Methods

We focused on a specific type of sounds used in computer interfaces: earcons. Earcons are short musical melodies that signal an action occurring during the manipulation of an interface (Blattner, Sumikawa, & Greenberg, 1989). Many studies concerning earcons have been reported in the last two decades. Brewster's (2008) survey about earcon design and evaluation, reiterates the design guidelines proposed by Blattner et al. (1989) concerning the composition of earcons on the basis of simple motifs and how they can be combined and hierarchically organized in families according to their intended symbolic meaning. For example, the earcons corresponding to different error messages could have the same rhythmic structure and differ in timbre or pitch range according to the specific error and could also be combined to represent multiple and related errors. Some other design guidelines are illustrated as an outcome of the work of Brewster, Wright, and Edwards (1994). The fundamental result is that the more complex earcons are the more effective in terms of learnability; significant differentiation in timbres, rhythmic patterns makes earcons more recognizable. Alternative, Brewster's group evaluation work advises that pitch should not be used as a discriminating parameter on its own. Melodic structures can be extremely refined, structured, and various, but they can be hardly distinguishable and memorizable. The case study tests the effectiveness of rhetoric guidelines right on this most critical parameter that is pitch.

In being musical structures, the relationship between the meaning conveyed by an earcon and its musical structure mostly relies on a symbolic convention (see Jekosch, 1999). As in the case of the design of other auditory interfaces (Edworthy, Hellier, & Hards, 1995; Rogers, Lamson, & Rousseau, 2000), the users of an interface endowed with earcons need a certain period of exposure to the earcons to successfully associate sounds and meanings. Warning signals represent the most common example of an auditory interface based on symbolic sound signals. In fact, in many contexts (aircraft, operating rooms, and so on), many different warning signals occur incessantly and concurrently. In such conditions, users may become incapable of deciding what a warning signal refers to, whether a warning is really urgent or not, and therefore whether to respond to it or not. To evaluate the effectiveness of the design of warning signals, several approaches have been developed.

The simplest technique consists in asking listeners to rate the perceived urgency of the signals (Edworthy, Loxley, & Dennis, 1991; Hellier, Edworthy, & Dennis, 1993). However, such a method cannot then be used to evaluate the ability of a signal to convey a specific meaning. To account for the possible significance conveyed by a signal, semantic scales have been elaborated (Edworthy et al., 1995) where the most common paradigm is to vary the signal-meaning relationships and to measure reaction times while users perform specific tasks requiring them to understand the intended meaning (Burt, Bartolome, Burdette, & Comstock, 1995; Haas & Casali, 1995; Haas & Edworthy, 1996;

Suied, Susini, & McAdams, 2008). Interestingly, such methods have shown that the degree of representation of a sound (see, for instance, the work by Jekosch, 1999, regarding symbolic to indexic relationships) has a great influence on the reaction times (Belz, Robinson, & Casali, 1999; Graham, 1999; Guillaume, Pellieux, Chastres, & Drake, 2003). However, for a computer interface, the reaction times are probably not proper indicators of the relevance of the auditory design. Rather, the strength of the association between a sound and its meaning is crucial, particularly in interfaces without visual displays (Fernström, Brazil, & Bannon, 2005).

Our case study introduced musical rhetorical principles for the design of earcons in a computer interface. Consistent with the fifth proposed guidelines (*memoria*), we used rhetorical principles to make earcons easier to memorize. Our goal was precisely to increase the strength of association between the earcons and their meaning. As in the works of Rath and Rocchesso (2005), Rath and Schleicher (2008), and Lemaitre et al. (2009), we adopted a learning paradigm as evaluation criterion. More precisely, we observed how swiftly listeners could memorize the association between each earcon and its corresponding meaning through the increasing number of times they were exposed to the earcons.

Early studies of the memorization of natural sounds (Bartlett, 1977; Bower & Holyoack, 1973) showed that sounds are memorized as a composite system comprising salient acoustical features and meaningful labels. This was inspired by Paivio's (1971) theory of dual coding for images, according to which images are coded in two independent codes (images and verbal labels). More recent research in the context of cognitive neuroscience has shown however that sound identification does not only rely on linguistic mediation (Dick, Bussiere, & Saygin, 2002; Schön, Ystad, Kronland-Martinet, & Besson, 2009). Research suggests that auditory stimuli elicit both a sensory traces and an amodal, semantic representation in memory that includes a large spread of associated concepts (Friedmann, 2003; Näätänen & Winkler, 1999). As regards musical melodies, Deutsch (1980) showed that listeners better memorize melodies with a clear structure. Özcan and Van Egmond (2007) studied memorization of associations between sounds and labels (text, images, pictograms) for product sounds, and Keller and Stevens (2004) for auditory icons. These latter authors found that associations between sounds and images/labels were better memorized when a direct semantic relationship existed between sounds and labels (e.g., sound and picture of a helicopter) in comparison to an ecological (e.g., sound of helicopter and picture of a machine) or a metaphorical relationship (e.g., sound of helicopter and picture of a mosquito). Finally, cognitive neuroscience studies closely links auditory perception to motor representations (Aglioti & Pazzaglia, 2010).

This suggests that listeners can use different types of codes to memorize earcons and to associate them to a computer function. Listeners can use salient acoustic features (timbre, pitch range, etc.), linguistic labels (name the sounds), musical structures, semantic associations (among which visual imagery), and motor reactions elicited by the sounds to associate a sound and a function. Designers can use these different elements to make

earcons easier to memorize. We chose to concentrate only on the melodic structure and leave the timbre of the earcons aside, as well as possible cross-modal aspects. Such musical melodies do not bear direct “natural” relationship to any operation function of the computer interface. However, as discussed in the case study, musical rhetoric provided us with a set of rhetorical figures that we used to create analogies between the musical structure and the computer function at a metaphorical level. The hypothesis was that since the rhetorical earcons introduced into the signal itself some representation by analogy of its meaning, their symbolism would be more striking and convincing. Thus, the adherence of the mind of the user—their memorization—would occur faster than in the case of non-rhetorical earcons.

Earcons for Computer Interfaces: A Case Study

We applied our thesis to the very specific case of sound design for an elementary computer interface. Specifically, we concentrated on the possibilities offered by the application of musical rhetoric to the design of a set of earcons, with a particular focus on the use of rhetorical figures. We designed an interface with eight operating functions and three sets of eight earcons assigned one-to-one to the eight operating functions. The design of the first set of sounds followed rhetorical principles. This set was used as the baseline for comparison. In the two other sets, we arbitrarily associated earcons and functions, without using any rhetorical principle.

The design centered on the rhetorical principle of *memoria*; make earcons easier to memorize in order to associate them to their respective function. The case does not deal with complex structures such as a speech or an entire musical composition, but with musical cells (the earcons) associated with very specific and discrete computer functions. However, the same principle holds good. A rhetorically well-designed earcon is more characterized by and convincing in terms of what it represents and, thus, more easily memorizable than when rhetorical strategies are weaker or absent. Therefore, we conducted a set of experiments aimed at validating the improvement of subjects’ memorization performance due to the introduction of rhetorical principles in the design process.

Interface Design

We adopted an elementary graphical interface implemented in Java. In the training phase, users performed ordinary filing and editing functions on simple geometrical figures. The users’ specific task was to perform eight sonified functions (*new*, *save*, *cut*, *paste*, *move*, *delete*, *undo*, and *redo*) a fixed number of times. The graphical interface was a simple metaphor of a computer application, representing a basic desktop/working-area and a clipboard. By typing shortcut key combinations, the subjects manipulated a small square, moving it through the desktop, deleting it, cutting it, and so on (see Figure 1). Typing these shortcut keys triggered the playback of the corresponding earcon.

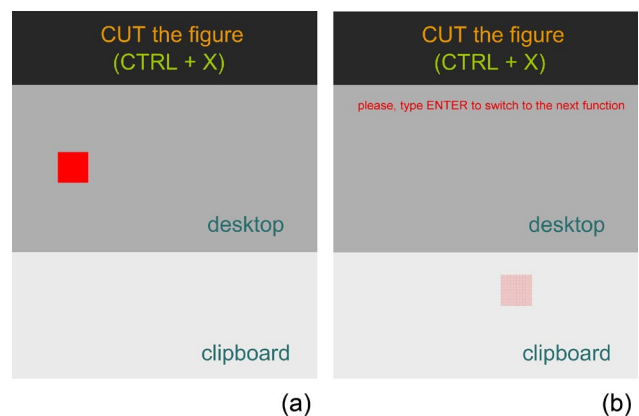


Figure 1. Training interface, an example.
Visual representation of the cut function (a) before performing the action and (b) after having performed it. The corresponding earcon is played back as soon as the user types the key combination Ctrl+X.

We composed two sets of eight earcons each. The former set (denoted R0) contained rhetorical principles and the latter either none at all, or, as discussed later, hardly any. To perform the experiment twice and consolidate the validation results, we composed two versions of this latter set (NR0 and NR1) denoted in the following as “non-rhetorical” sets. The rhetorical set differed from the non-rhetorical ones only in the additional and thoroughly intentional presence of rhetorical figures. Earcons are often monodic. Since it was possible to work on a single parameter, namely melody, we decided that this could be the simplest work bench for our research. All the earcons were monodic and identical with respect to underlying tonality, melodic range, number of notes, metre, tempo, and rhythmic/phrase structure (see Figures 2, 3, and 4). This identity was pursued to exclude all parameters other than the melodic as elements that might distinguish and thus identify the earcons. The rhetorical design affected the melodic aspect exclusively and the aim of the design being to isolate this aspect. On the basis of these premises, we studied the effect of the use of rhetorical stratagems in one of the earcon sets on the users’ learning performance with respect to the non-rhetorical cases. For this purpose, the maximum of variety in the melodies, particularly a one-to-one discriminability within each set, was pursued. A tolerance threshold in terms of coincident notes (same notes in the same metrical position) was set to five; no more than five identical notes out of nine were allowed. This contributed to the achievement of a common grade of identifiability on the basis of pure melodic criteria in each of the three earcon sets. The final sets of earcons were defined after iterated design cycles through numerous discussions with colleagues in the labs and informal experimental tests to obtain a uniform one-to-one discriminability. In summary, the question of the case study was: If a set of earcons, distinguishable only at a melodic level and identical in all of the other musical parameters, were designed according to melodic rhetoric principles, would they be more easily memorizable?

The design procedure was as follows. We visualized the function in an iconic-spatial and/or conceptual sense and transposed by analogy the image/concept into a musical

rhetorical figure. The rhetorical figures adopted in the set R0 (Figure 2) were selected on the basis of the principles of *elocutio* in musical rhetoric (Gottsched, 1750). The other two sets, NR0 and NR1 (Figures 3 and 4) were freely designed even though some rhetorical contents were present as regards, for example, the ascending and descending motions of the melodies, which are clear metaphors for appearing or disappearing. We achieved a *cantabile* character in all of the earcons in the R0 set. The earcons were played expressively on an electronic piano by a professional classical musician to underline the structure of the melodies. As discussed in the section “Musical Rhetoric for AD and SID”, this point is crucial and consistent with the integration of the *actio* in the rhetorical construction. Rhetorical figures and their expressive utterance are the two sides of a living organism and have to be considered as a whole. The performing style was that of Baroque music as played and taught in Western music schools. The interpretation of such short melodic cells involved dynamic expression (accents, crescendos, diminuendos) and micro-accelerations/decelerations, albeit within the constraints of a rigid beat.

The earcons were played and recorded at 60 BPM, but played back during the experiment at a very fast tempo (200 BPM). This made it difficult to analyze the rhetorical figures at a conscious level while listening to them. The goal was rather to emphasize the meaning/association in a “holistic”/intuitive fashion. The earcons of the NR0 set were played back on the same sampled electronic piano used for R0 by a computer-generated MIDI file, that is, without dynamic or temporal expressivity. This

Figure 2. The earcon set obtained by a rhetorical design and denoted as R0. The rhetorical figures are highlighted by means of the slurs.

Figure 3. The first set of earcons obtained by free design and denoted as NR0.

Figure 4. The second set of earcons obtained by free design and denoted as NR1.

choice was coherent with the fact that a non-rhetorical condition can be interpreted as absence of both *elocutio* and *actio*. By contrast, the NR1 set was played on the electronic piano by the professional classical musician. The purpose of considering two non-rhetorical sets was twofold: To corroborate the experimental results with a second case and to verify if the results obtained with the NR0 set could have been affected by the lack of expressiveness in the playback of the earcons. No emotional level was considered at all, not even in the humanly expressed sets. Each earcon was played *legato* and the dynamic variations (*crescendo/diminuendo*, accents, and so on) were only meant to highlight the musical structure at a cognitive level. The slurs in the score of Figure 2 were only there as markers of the rhetorical structures.

The choice of the rhetorical figures was as follows:

- *New*—The *anabasis* (ascending melody) was chosen to give the impression by iconic analogy of something that is emerging, appearing *ex novo*. The *anticipatio*, that is, the anticipation of each note on the upbeat position, was introduced to accentuate and underline the step-by-step ascent.
- *Save*—The *anadiplosis* figure takes place when the same word is used at the end of a sentence and at the beginning of the following one. Here, the conceptual analogy was that saving something means to fix a certain point in the work in progress and to restart from that point for further operations.
- *Delete*—By employing a *catabasis* (descending melody) jointly with the *anticipatio*, this earcon was designed as a symmetrical version of the earcon assigned to the *new* function to give the idea of disappearing.
- *Undo*—The *epanodos* figure occurs when two words are repeated in reverse order. The result was a retrograde motion of the melody, representing the cancellation of an action.
- *Redo*—The *epanalepsis* figure consists of employing the same expression at the beginning and end of the sentence. Such a figure seemed to be appropriate for depicting a previous action that is finally repeated after having been cancelled.
- *Cut*—For this function, we adopted a *hyperbole* to provide an iconic analogy of the “suspension” of the cut object in the clipboard. A *hyperbole* occurs when the melody exceeds its usual range towards the higher register. The clipboard is associated with an upper/suspended space.
- *Paste*—The *hypobole* employed for the sonification of this function realized a symmetrical structure with respect to the *cut* function. A *hypobole* occurs when the melody exceeds its usual range towards the lower register.
- *Move*—This function was rendered by means of an *anabasis* followed by a *catabasis* depicting by iconic analogy an object “lifted and dropped” in another position on the screen.

It is worth repeating that the presence or absence of rhetorical formulas is not dichotomic. Whenever there is a repetition or a similitude, we are somehow in a rhetorical framework. As an example, the earcons for the *new* function in Figures 3 and 4 show an ascending melody similar to the corresponding version in Figure 2, that is, they both realize an *anabasis*. In this sense, both the earcons have some rhetorical contents, but the one in Figure 2 has an increased degree of rhetoric provided by the *anticipatio*.

Experiment 1:

First Set of Non-Rhetorical Earcons

This experiment compared R0 and NR0. The two sets of earcons were designed according to the principles described in the previous section without any further specification.

Method

Participants

In total, 40 subjects (16 women and 24 men) volunteered as listeners and were paid for their participation. They ranged in age from 20 to 48 years old (median: 31 years old). All of them reported normal hearing. A subset of 20 participants was selected as musicians (professional musicians, musicologists, sound engineers, acousticians). They had reported from previous questionnaires considerable skill in music or sound analysis. The other 20 were considered non-musicians. They had not reported any specific education in music or sound.

Stimuli

The 16 earcons had a maximum level varying from 66.0 dB(A) to 71.1 dB(A). They had 16-bit resolution, with a sampling rate of 44.1 kHz.

Apparatus

The stimuli were amplified binaurally over a pair of Sennheiser HD250 linear II headphones. Participants were seated in a double-walled IAC sound-isolation booth.

Procedure

Each group of participants (musicians, non-musicians) was split into two subgroups so that in each group ten participants did the experiment with the non-rhetorical sounds, and ten with the rhetorical sounds. There were therefore four groups, resulting from the crossing of two factors (musical practice and design strategy). The participants were first introduced to the graphical interface of Figure 1 and trained to use it. They then performed eight trials, each consisting of a training and a test phase. During each training phase, the interface indicated them to perform each of the eight functions. This procedure was repeated three times in every training phase so that a participant performed 24 manipulations in each trial. Participants were instructed to memorize the sounds associated with the functions. During each test phase, the earcons were presented one by one. The participants had to choose among the eight possible functions. The visual interface used during the training was not presented during the test phase. Figure 5 represents the simple graphical interface of the test phase. The participants had to type the numbers corresponding to the functions associated with the earcons played back one by one. After one earcon was played back, the participant had as much time as they needed to answer, but they could not listen to the earcon again. After each test phase, the participants received a score indicating how many correct

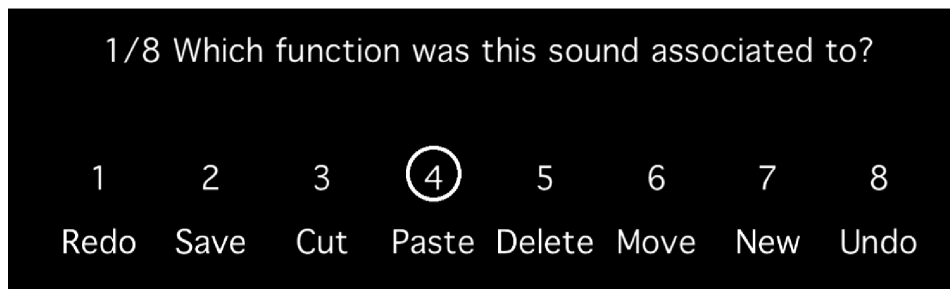


Figure 5. Graphical interface used during the test phase.

associations they had remembered, but without any indication of right and wrong answers. The order of presentation of the earcons was randomized for each subject and for every test phase.

We were specifically interested in the improvement of performance across the experiment. Thus, we measured the performance after each trial. The experiment therefore had a mixed design with musical training (musicians, non-musicians) and design strategy (rhetorical, non rhetorical) as a between-subject factor and the trials (1 to 8) as a within-subject factor.

Results

The experiment aimed to study how quickly users could learn the correspondences between functions and earcons, depending on the number of training trials. The dependent variable of interest was the amount of correct associations.

The experiment had the number of trials as a within-subject variable and the design strategy (rhetorical, non-rhetorical) of the earcons and the musical training (musicians, non-musicians) of the participant as between-subjects factors. The data were submitted to a mixed-design analysis of variance (ANOVA). Table 1 displays the results Figure 6 represents the percentage of

correct associations averaged over all of the participants in each group and all of the functions. In general, the scores increased significantly across trials, $F(7, 273) = 14.0, p < 0.01$. The effect of the design strategy was significant, $F(1, 39) = 16.7, p < 0.01$, and had the greatest effect on the scores ($\eta^2 = 0.32$). All told, the rhetorical sounds led to better memorization scores. The effect of the musical training was also significant, $F(1, 39) = 11.1, p < 0.01$. Globally, the musicians performed better than the non-musicians. The interaction between the design strategy and musical training was not significant, indicating that the increase of memorization due to the rhetorical earcons was the same for the musicians and for the non-musicians. Interestingly, the interaction between the trials and the design strategy was significant, $F(7, 273) = 2.3, p < 0.05$, showing that the increase of performance across the trials was not the same for the rhetorical and non-rhetorical earcons. The scores improved slightly faster for the rhetorical earcons. However, the effect was small ($\eta^2 = 0.03$). The interaction between the trials and the musical training was significant, $F(7, 273) = 3.4, p < 0.01$, denoting that the increase of performance across the trials was not the same for the musicians and the non-musicians. The scores improved slightly faster for the musicians, but the effect was also small ($\eta^2 = 0.09$).

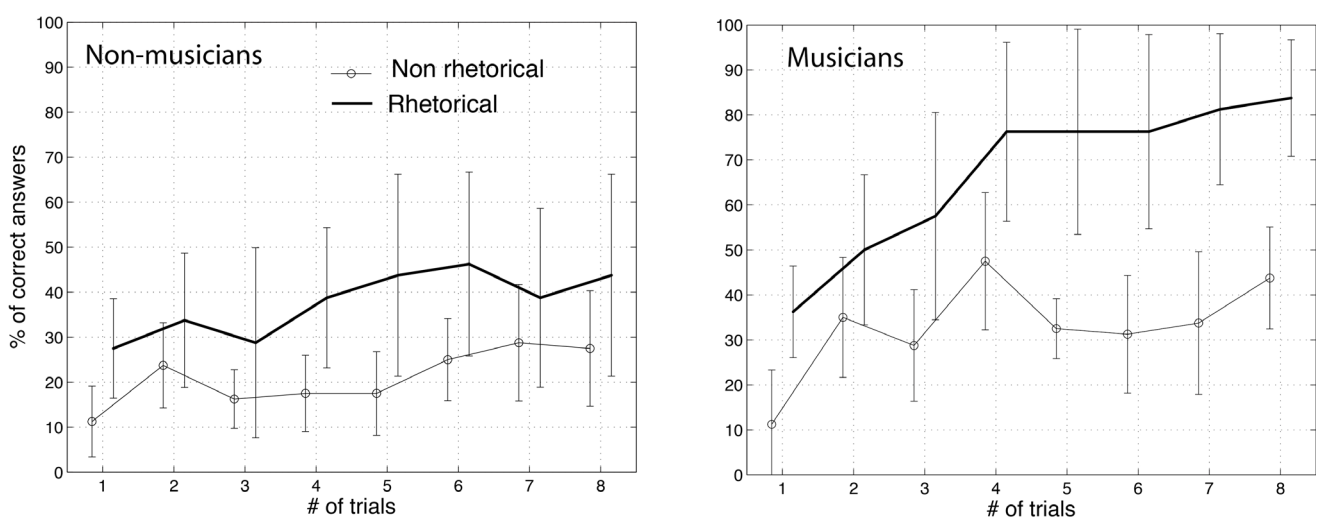


Figure 6. Experiment 1. The figure represents the percentages of correct associations (scores) between earcons and functions, averaged across all the functions and all the participants, in each of the four groups of participants. The left panel represents the results for the two groups of non-musicians, and the right panel the results for the two groups of musicians. The dark lines represent the scores for the set of rhetorical earcons, the light-gray lines represent the scores for the non rhetorical earcons. Vertical bars represent the 95% confidence intervals.

Table 1. Analysis of variance of the Experiment 1.

Source	df	F	η^2	p	GG
Between subjects					
DS	1	16.7	0.32	0.000**	
M	1	11.1	0.24	0.002**	
DS x M	1	2.0	0.05	0.1690	
Error	36	(3112)			
Within subjects					
T	7	14.0	0.14	0.000**	0.000
T x DS	7	2.3	0.03	0.029*	0.047
T x M	7	3.4	0.09	0.002**	0.006
T x DS x M	7	1.2	0.00	0.2970	0.305
Error	252	(239)			

Note: Values enclosed in parentheses represent mean square errors. DS = Design Strategy; M = Musical training; T = Trials; *df*: degree of freedom; *F*: Fischer's *F*; η^2 : proportion of variance attributable to the independent variable; *p*: probability associated with *F* under the null hypothesis. ** $p < 0.01$. * $p < 0.05$. GG = probability after Geisser-Greenhouse correction of the degree of freedom.

Discussion

The analysis of variance reported three major effects on the memorization performances. Firstly, the group of musicians performed better than the group of non-musicians. Previously, only a few studies reported differences between expert (including musicians) and non-expert listeners (see Lemaitre, Houix, Misdariis, & Susini, 2010, for a review). The difference observed here possibly indicates that the rhetorical strategies used to design the earcons correspond to musical figures that were already known by the musicians, probably due to their practice and training, and not by the non-musicians.

The effect of training was another major factor influencing the memorization performances. For both groups of participants, the performance increased with the repetition of trials. The learning rate of the set of rhetorical earcons was, however, greater for the musicians, who memorized almost perfectly the associations between earcons and functions after the eight trials. This indicates that once they recognized the musical figures, the musicians were rapidly able to associate them with the corresponding functions. One could argue that this association was purely symbolic: recognizing a musical figure, naming it, and relating it to the corresponding function. Nevertheless, against this hypothesis we note that non-musicians were also able to increase their performance across trials, without any conscious knowledge of the matter.

Finally, the analysis showed that the rhetorical earcons led to better memorization performances. This effect was more evident in the musicians than in the non-musicians. However, for both groups of participants the rhetorical figures made the sound-meaning association easier to remember. This clearly suggests that the representational character of the sound-meaning relationship introduced by the rhetorically guided design was a strong cue for the listeners. It was, however, possible that this result depended

only on this specific set of earcons. In particular, the fact that the non-rhetorical set was devoid of expressivity may have made these sounds more difficult to memorize. In order to investigate this possibility, we replicated the same procedure in a second experiment, using the NR1 set played by a professional pianist.

The non-significant interaction between design strategy and musical training indicated that the size of the effect of the design strategy did not depend on subjects' musical training. Therefore, we decided to use only one group of participants in the second experiment. In addition, we sought to magnify the potential effect of the rhetorical design strategies on the memorization performance. Thus, we used only musically trained participants, because the results of the first experiment showed that the performance increased more rapidly for these subjects.

Experiment 2:

Second Set of Non-Rhetorical Earcons

Experiment 2 replicated Experiment 1, using the non-rhetorical set NR1 instead of NR0.

Method

Participants

Fifteen participants took part in the experiment (six female and nine male). They were aged from 23 to 48 years old. All of them were musicians (13 were professional musicians, 2 amateur musicians). None of them had previously participated in Experiment 1. Seven used the rhetorical set R0 and eight used the non-rhetorical set NR1.

Stimuli, apparatus, and procedure

We used the same equipment and stimuli as in Experiment 1, except for the substitution of the new set of non-rhetorical earcons NR1. The procedure was the same as in Experiment 1, except that there was no musical training factor. The experiment therefore had one between-subject factor (design strategy: rhetorical and non-rhetorical), and one within-subject factor (number of trials).

Results

The data were submitted to the same analysis of variance as in Experiment 1. Table 2 and Figure 7 report the results of this analysis. The design strategy had the most important significant effect, $F(1, 13) = 11.3$, $p < 0.01$, $\eta^2 = 34.7\%$. The rhetorical earcons were better remembered than the non-rhetorical earcons. On average, the percentage of correct associations was 84.0% for the former set, and 46.9% for the latter. The trials also had a significant effect on the percentage of correct associations, $F(7, 91) = 7.53$, $p < 0.01$, $\eta^2 = 8.8\%$. The subjects better memorized the earcons as they acquired more experience of them. However, contrary to Experiment 1, the interaction between the design strategy and the trials was not significant, $F(7, 91) = 1.22$, $p = 0.313$. The performance increased at the same rate in the two groups.

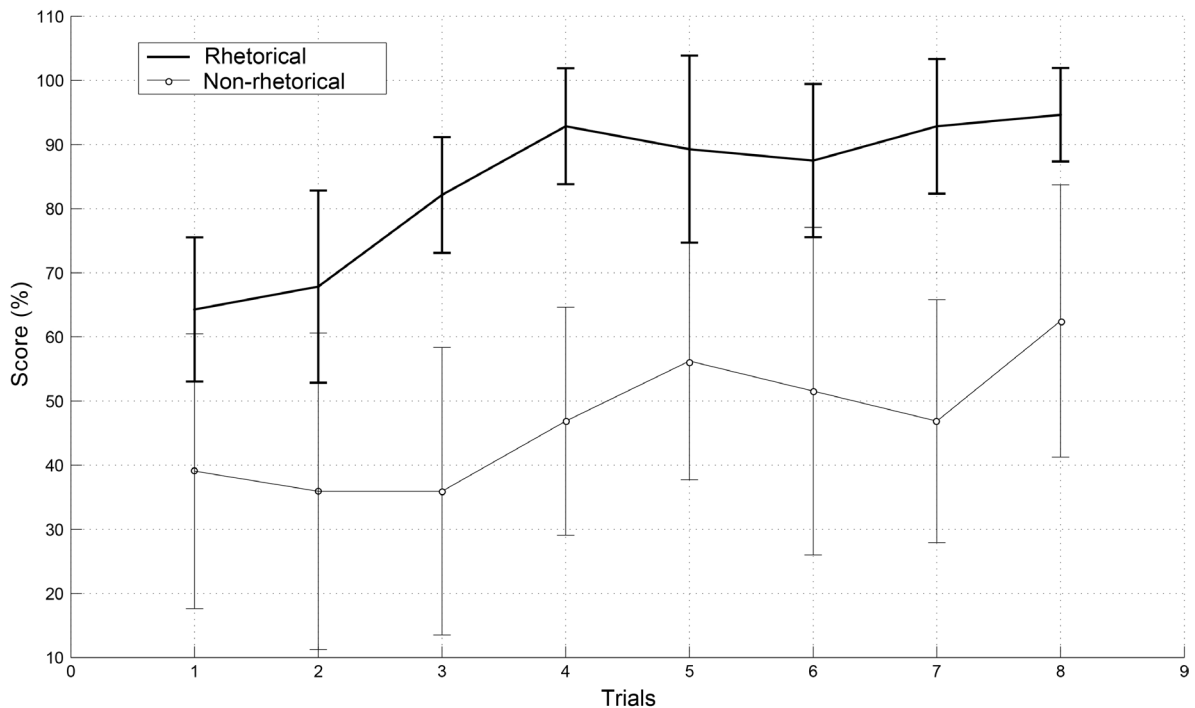


Figure 7. Experiment 2. The figure represents the percentage of correct associations between earcons and functions averaged across all the functions and participants. The dark line represents the scores for the rhetorical set, the light-gray line the scores for the second non-rhetorical set NR1. The vertical bars represent 95% confidence intervals.

Table 2. Analysis of variance of Experiment 2.

Source	df	F	η^2	p	GG
Between subjects					
DS	1	11.3	34.70	0.005**	
Error	13	(41176)			
Within subjects					
T	7	7.5	8.80	0.0000**	0.000
T x DS	7	1.2	1.42	0.3000	0.313
Error	91	(17946)			

Note: Values enclosed in parentheses represent mean square errors. S = Subjects; DS = Design Strategy; T = Trials; *df*: degree of freedom; *F*: Fischer's *F*; η^2 : proportion of variance attributable to the independent variable; *p*: probability associated with *F* under the null hypothesis. ** *p* < 0.01. * *p* < 0.05. GG = probability after Geisser-Greenhouse correction of the degree of freedom.

Discussion

The results of Experiment 2 are similar to those of Experiment 1. Rhetorical earcons were better memorized than the earcons of the second non-rhetorical set NR1. This confirms that the effect observed in Experiment 1 did not depend on the particular set of sounds.

Altogether, the results of these two experiments indicate that using rhetorical strategies to design earcons leads to better learning performances by users. There are different possible explanations for the better memorization performance of the rhetorically designed set of earcons. The rhetorical figures

introduced analogies between earcons and functions that have made associations easier to memorize. Rhetorical figures have certainly made the earcons overall more distinct one from the other and thus easier to remember. One could argue that any design strategy able to reinforce the structure of the earcons would provide listeners with cues that facilitate memorization. Some structure was also present in the non-rhetorical sets. Indeed, it is possible that other systematic methodologies would reach similar results, while focusing only on the melodic parameter. The advantage of our approach is that transposing verbal or musical rhetorical techniques to the design of auditory interfaces is rather straightforward and based on an already existing and consolidated communication theory. At the same time, we claim that the experimental results demonstrate, at least for this particular example, that rhetorical figures are convenient tools to effectively improve the design of earcons. The rhetorical figures used to build the rhetorical set were based on melody, that is, the combination of pitches. This is somewhat at odds with the recommendations of Brewster et al. (1994) and Brewster (2008). On the basis of evaluation tests, these authors recommended using timbre as the main cue to differentiate earcons associated to different functions. They also suggested that pitch is not an effective cue to create earcons because subtle melodic differences are not easily memorized. It is therefore significant that the pitch structures created in the current study by means of rhetoric figures increased the memorizability of the earcons. Our interpretation is that the rhetoric strategies led to melodic structures that were so powerful that they overcame the limitation of pitch memorization suggested by Brewster et al. (1994).

However, we must also stress that our experiments mainly used trained musicians as subjects. Brewster et al. (1994) showed that musicians and non-musicians recalled earcons equally well, but we cannot completely exclude that trained musicians memorize melodic structure more easily than lay listeners. This limits the generality of our conclusions to an extent and further work is required to study the effect of musical expertise.

Since the experiment also involved visual stimuli, some interference between the two sensory modalities could have occurred. For example, in the graphical interface, the cut and paste functions used a visual metaphor as well; items were moved from desktop to clipboard, generating a downward motion of the small square, and from clipboard to desktop, generating an upward motion of the square respectively. The corresponding earcons used ascending and descending melodies. Listeners' recovery of the function could have been mediated through visual imagery of these motions, although it should be noted that the graphical interface was not present during the test phase. Another possible cross-modal effect is the fact that the earcons were associated with the finger motions on the keyboard required to perform the actions. The rhetorical figures might have introduced some connection between the melodic structure and the pattern of keys employed in the training phase, which also mediated subjects' association of sounds and functions. It should be noticed, though, that during the test phase, the subjects did not have to type the key combinations used in the training. Any possible multimodal or cross-modal effect is beyond the scope of this study and we point out how both visual metaphors and finger positions were identical in the training phases of all the three earcon sets. Alternatively, we think that a rhetorically-informed design of visual or proprioceptive aspects would be consistent with a wider perspective involving the application of rhetorical principles, not only to sound, but to ID in general.

Far from being a general proof, the case study represents a particular example of what the application of rhetoric principles to sound in ID can be, providing an initial concrete test bench. Even if limited to the very specific case of earcon design, the adopted strategy has proved effective in terms of sound-meaning association and has also provided an example of non-conventional use and reinterpretation of rhetorical formulas and indications. In our view, the results of this particular case study allow us to envisage that the use of rhetoric as methodological guidance can enhance the design of sounds for AD and SID applications well beyond the case of earcon design.

Future Developments

Our plan is to go beyond music and explore the cases of electroacoustic and everyday sounds. The former category includes synthetic or processed sounds, where timbre is the main element and no source is recognizable according to the acousmatic notion introduced by Pierre Schaeffer (see Chion, 1990). The latter category involves the consideration of cultural and anthropological perspectives according to Murray Schafer's concept of soundscape (Schafer, 1994). This will

require a wide range of multidisciplinary contributions. Polotti and Benzi (2008) presented an exploratory work on these aspects in a recent edition of the International Conference on Auditory Display. The complexity of the parameters involved in the perception of both categories of sounds is not an easy task. In the experiment presented in this article, we considered a specific case of non-verbal sounds, namely music, and we were able to uncouple a single parameter, that is, melody. We think that it will be possible to apply a similar validation strategy to the case of synthetic and everyday sounds by isolating single psychoacoustic and cognitive parameters.

The definition and implementation of qualitative studies, including interviews with sound designers or a more systematic analysis of how previously-proposed design guidelines could be retroactively understood in the context of rhetoric, is another way to support our idea. In the "Background" section of the article, we provided some hints in this direction. As an example we briefly discuss a rhetorical reading of a recent work (Rocchesso, Polotti, & Delle Monache, 2009). In the article, continuous sonic interaction was investigated in the case of three kitchen scenarios: the action of screwing the two elements of a moka, a fast, and iterated slicing of a vegetable on a cutting board, and a sonified dining table. All three scenarios involved a rhetorical principle. In the moka case, the authors made use of the figure of *emphasis*. The term emphasis comes from the Greek, *en* = "in" and *phasis* = "to shine", which together signifies "to bring to light". In the sonic case, this could be translated as "to bring to a conscious hearing". This was defined elsewhere as the principle of "minimal yet veridical" (Rocchesso, Bresin, & Fernstrom, 2003):

- record the real sound or simulate it,
- modify it to emphasize its most perceptually and semantically relevant features,
- extend it temporally over the whole action of screwing in the two pieces of a moka,
- modulate it interactively to indicate moment-by-moment the state of the process in a semantically evident fashion.

In the cutting of vegetables, the temporal behavior of the sonic feedback was compelling in terms of its *actio*. The audio feedback rhythm was adaptively more or less synchronous with the cutting rhythm to induce the user to recover the regularity of their movement in case of time deviations. The reference to the musical rhetorical figures of *accelerando* and *rallentando* is manifest. In the dining table scenario, the principle of contradiction, that is, the figures of *antithesis*, and *synaesthesia*, was a crucial argumentative element for the achievement of an engaging sound design. The former figure was employed to contradict the expected sound produced by some action, for example, a gurgling liquid sound accompanied the action of stirring solid food such as a salad. The latter was used to define a sonic identity/characterization of liquid ingredients. The *antithesis* had the explicit purpose of generating surprise, or even hilarity, to let people experience the importance of the sound feedback of their actions. The *synaesthesia* aimed to confirm the identity of the ingredient, monitoring its quantity, thus reinforcing the enactive experience and anticipating its taste.

We plan a new set of experiments, inviting designers other than the authors to approach a task, first adopting their usual methodologies and then trying to apply some of the guidelines we formulated as a future test bench of this methodological proposal. Another essential research track connected to the integration of musical rhetoric should be that of automatic expressive performance. Whereas we stated that *actio* is part of the game, we cannot expect to rely on human performance of the sonic items of an AD application; proper algorithms for the expressive execution of rhetorical structures should be developed. This would be a parallel research effort that represents a field of research on its own (see, for example, Dahlstedt, 2008).

Acknowledgments

This research was partially supported by the project CLOSED (<http://closed.ircam.fr/>) and by the Action COST- IC0601 on Sonic Interaction Design (<http://sid.soundobject.org/>). We wish to thank Carlo Benzi, Giorgio Klauer, Nicolas Misdariis, and Patrick Susini for many insights.

References

1. Aglioti, S. M., & Pazzaglia, M. (2010). Representing actions through their sound. *Experimental Brain Research*, 206(2), 141-151.
2. Bach, C. P. E. (1753). *On the true art of playing keyboard instruments* (W. J. Mitchell, Trans.). London, UK: Cassell (First edition published 1949).
3. Barrass, S. (2005a). A comprehensive framework for auditory display: Comments on Barrass, ICAD 1994. *ACM Transactions on Applied Perception*, 2(4), 403-406.
4. Barrass, S. (2005b). A perceptual framework for the auditory display of scientific data. *ACM Transactions on Applied Perception* 2(4), 389-402.
5. Bartel, D. (1997). *Musica poetica: Musical-rhetorical figures in German baroque music*. Lincoln, NE: University of Nebraska Press (Originally published in German in 1982).
6. Bartlett, J. C. (1977). Remembering environmental sounds: The role of verbalization at input. *Memory and Cognition*, 5(4), 404-414.
7. Belz, S. M., Robinson, G. S., & Casali, J. G. (1999). A new class of auditory warning signals for complex systems: Auditory icons. *Human Factors*, 41(4), 608-618.
8. Benzi, C., (2004). *Stratégies rhétoriques et communication dans la musique contemporaine européenne (1960-1980)* [Rhetoric strategies and communication in the European contemporary music]. Lille, France: Editions A.N.R.T.
9. Blattner, M. M., Sumikawa, D. A., & Greenberg, R. M. (1989). Earcons and icons: Their structure and common design principles. *Human Computer Interaction*, 4(1), 11-44.
10. Bower, G. H., & Holyoak, K. (1973). Encoding and recognition memory for naturalistic sounds. *Journal of Experimental Psychology*, 101(2), 360-366.
11. Brewster, S. A., Wright, P. C., & Edwards, A. D. N. (1994). A detailed investigation into the effectiveness of earcons. In G. Kramer (Ed.), *Proceedings the 3rd International Conference on Auditory Display* (pp. 471-498). Reading, MA: Addison-Wesley.
12. Brewster, S. A., Wright, P. C., & Edwards, A. D. N., (1995). Parallel earcons: Reducing the length of audio messages. *International Journal of Human-Computer Studies*, 43(22), 153-175.
13. Brewster, S., (2008). Non-speech auditory output. In J. A. Jacko & A. Sears (Eds.), *The human computer interaction handbook* (2nd ed., pp. 247-264). Hillsdale, NJ: Lawrence Erlbaum Associates.
14. Burt, J. L., Bartolome, D. S., Burdette, D. W., & Comstock, J. R. Jr. (1995). A psychophysiological evaluation of the perceived urgency of auditory warning signals. *Ergonomics*, 38(11), 2327-2340.
15. Carroll, J. M., & Mack, R. L., (1999). Metaphors, computing systems, and active learning. *International Journal of Human-Computer Studies*, 51(2), 385-403.
16. Chion, M. (1990). *L'audio-vision. Son et image au cinéma* [Audio-vision. Sound and image in cinema]. Paris, France: Editions Nathan.
17. Dahlstedt, P. (2008). Dynamic mapping strategies for expressive synthesis performance and improvisation in computer music modeling and retrieval. In S. Ystad, R. Kronland-Martinet, & K. Jensen, (Eds.) *Computer music modeling and retrieval: Genesis of meaning in sound and music* (pp. 227-242). Heidelberg, Germany: Springer.
18. De Marsico, M., & Levialdi, S. (2004). Evaluating web sites: Exploiting user's expectations. *International Journal of Human-Computer Studies*, 60(3), 381-416.
19. Deutsch, D. (1980). The processing of structured and unstructured tonal sequences. *Perception & Psychophysics*, 28(5), 381-389
20. Dick, F., Bussiere, J., & Saygin, A. P., (2002). The effects of linguistic mediation on the identification of environmental sounds. *Center for Research in Language Newsletter*, 14(3), 3-9.
21. Drioli, C., Polotti, P., Rocchesso, D., Delle Monache, S., Adiloglu, K., Annies, R., & Obermayer, K., (2009). Auditory representations as landmarks in the sound design space. In *Proceedings the 6th Conference on Sound and Music Computing* (pp. 315-320). Porto, Portugal: Casa da Música.
22. Edworthy, J., Hellier, E., & Hards, R. (1995). The semantic associations of acoustic parameters commonly used in the design of auditory information and warning signals. *Ergonomics*, 38(11), 2341-2361.
23. Edworthy, J., Loxley, S., & Dennis, I. (1991). Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *Human Factors*, 33(2), 205-231.
24. Fernström, M., Brazil, E., & Bannon, L. (2005). HCI design and interactive sonification for fingers and ears. *IEEE Multimedia*, 12(2), 36-44.

25. Frauenberger, C., & Stockman, T. (2009). Auditory display design: An investigation of a design pattern approach. *International Journal of Human-Computer Studies*, 67(11), 907-922.
26. Friedman, D., (2003). Cognition and aging: A highly selective overview of event-related potential (ERP) data. *Journal of Clinical and Experimental Neuropsychology*, 25, 702-720.
27. Gottsched, J. C. (1750). *Ausfuerliche Redekunst nach Anleitung der Griechen und Roemer wie auch der neuern Auslaender* [The art of speech according to Greeks and Romans as well as the new foreigners]. Leipzig, Germany: Leipzig, Breitkopf.
28. Graham, R. (1999). Use of auditory icons as emergency warnings: Evaluation within a vehicle collision avoidance application. *Ergonomics*, 42(9), 1233-1248.
29. Grasso, F., Cawsey, A., & Jones, R. (2000). Dialectical argumentation to solve conflicts in advice giving: A case study in the promotion of healthy nutrition. *International Journal of Human-Computer Studies*, 53(6), 1077-1115.
30. Groupe μ . (1999). Magritte, pièges de l'iconisme ou du paradoxe en peinture [Magritte, traps of icons or about the paradox in painting]. In N. Everaert-Desmedt (Ed.), *Magritte au risque de la sémiotique* [Magritte and the semiotic risk] (pp. 87-109). Brussels, Belgium: Publication des Facultés universitaires S. Louis.
31. Guillaume, A., Pellieux, L., Chastres, V., & Drake, C. (2003). Judging the urgency of nonvocal auditory warning signals: Perceptual and cognitive processes. *Journal of Experimental Psychology: Applied*, 9(3), 196-212.
32. Haas, E. C., & Casali, J. C. (1995). Perceived urgency of and response time to multi-tone and frequency-modulated warning signals in broadband noise. *Ergonomics*, 38(11), 2313-2326.
33. Haas, E. C., & Edworthy, J. (1996). Designing urgency into auditory warnings using pitch, speed, and loudness. *Computing and Control Engineering Journal*, 7(4), 193-198.
34. Hellier, E. J., Edworthy, J., & Dennis, I. (1993). Improving auditory warning design: Quantifying and predicting the effects of different warning parameters on perceived urgency. *Human Factors*, 35(4), 693-706.
35. Hermann T., Hunt A., & Neuhoff, J. G. (Eds.) (2011). *The sonification handbook*. Berlin, Germany: Logos Verlag.
36. Hermann, T., Williamson, J., Murray-Smith, R., Visell, Y., & Brazil, E., (2008). Sonification for sonic interaction design. In D. Rocchesso (Ed.), *Proceedings of the CHI Workshop on Sonic Interaction* (pp. 35-40). New York, NY: ACM.
37. Jekosch, U. (1999). Meaning in the context of sound quality assessment. *Acustica united with Acta Acustica*, 85(5), 681-684.
38. Jokinen, K., Sadek, D., & Traum, D. (2000). Introduction to special issue on collaboration, cooperation and conflict in dialogue systems. *International Journal of Human-Computer Studies*, 53(6), 867-870.
39. Keller, P., & Stevens, C. (2004). Meaning from environmental sounds: Types of signal-referent relations and their effect on recognizing auditory icons. *Journal of Experimental Psychology: Applied*, 10(1), 3-12.
40. Lemaitre, G., Houix, O., Misdariis, N., & Susini, P. (2010). Listener expertise and sound identification influence the categorization of environmental sounds. *Journal of Experimental Psychology: Applied*, 16(1), 16-32.
41. Lemaitre, G., Houix, O., Visell, Y., Franinovi'c, K., Misdariis, N., & Susini, P. (2009). Toward the design and evaluation of continuous sound in tangible interfaces: The Spinotron. *International Journal of Human-Computer Studies*, 67(11), 976-993.
42. Lemaitre, G., Houix, O., Susini, P., Visell, Y., & Franinovic, K. (2012). Feelings elicited by auditory feedback from a computationally augmented artifact: The Flops. *IEEE Transactions on Affective Computing*, 3(3), 335-348.
43. Lotman, Y. M., (1993). Painting and the language of theater: Notes on the problem of iconic rhetoric. In A. Efimova & L. Manovich (Eds.), *Tekstura: Russian essays on visual culture* (pp. 45-55). Chicago, IL: Chicago UP.
44. Mann, W., & Thompson, S. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3), 243-281.
45. Mattheson, J., (1981). *Der vollkommene Capellmeister: A revised translation with critical commentary* (E. C. Harriss, Trans.). Michigan, MI: UMI Research Press (Original work published 1739).
46. Merleau-Ponty, M., (1964). The film and the new psychology (H. L. Dreyfus & P. Allen Dreyfus, Trans.). In M. Merleau-Ponty (Ed.), *Sense and non-sense* (pp. 48-59). Evanston, IL: Northwestern University Press. (Original work published 1948)
47. Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, 125(6), 826-859.
48. Özcan, E., & Van Egmond, R. (2007). Memory for product sounds: The effect of sound and label type. *Acta Psychologica*, 126(3), 196-215.
49. Paivio, A., (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, and Winston.
50. Patterson, R., (1990). Auditory warning sounds in the work environment. *Philosophical Transactions of the Royal Society B*, 327(1241), 485-492.
51. Perelman, C., & Olbrechts-Tyteca, L. (1969). *The new rhetoric, a treatise on argumentation* (J. Wilkinson & P. Weaver, Trans.). Notre Dame, IN: University of Notre Dame (Original work published 1958).
52. Pirhonen, A. (2005). To simulate or to stimulate? In search of the power of metaphor in design. In *Future interaction design* (pp. 105-123). London, UK: Springer.
53. Polotti, P., & Benzi, C. (2008). Rhetorical schemes for audio communication. In P. Susini & O. Warusfel (Eds.), *Proceedings of the 14th International Conference on Auditory Display*. Paris, France: Institut de Recherche et Coordination Acoustique.

54. Popescu, V., Caelen, J., & Burileanu, C. (2007). Logic-based rhetorical structuring for natural language generation in human-computer dialogue. In *Lecture notes in computer science* (Vol. 4629, pp. 309-317). Berlin, Germany: Springer.
55. Quintilianus, M. F. (1996). *Institutio Oratoria* [Institutes of Oratory] (H. E. Butler, Trans.). Cambridge, MA: Harvard University press. (Original work published 1920)
56. Rastier, F. (2001). L'hypallage et Borges [The hypallage and Borges]. *Variaciones Borges*, 11, 3-33.
57. Rath, M., & Rocchesso, D. (2005). Continuous sonic feedback from a rolling ball. *IEEE MultiMedia*, 12(2), 60-69.
58. Rath, M., & Schleicher, R. (2008). On the relevance of auditory feedback for quality of control in a balancing task. *Acta Acustica united with Acustica*, 94(1), 12-20.
59. Rocchesso, D., Bresin, R., & Fernstrom, M. (2003). Sounding objects. *IEEE Multimedia*, 10(2), 42-52.
60. Rocchesso, D., Polotti, P., & Delle Monache, S. (2009). Designing continuous sonic interaction. *International Journal of Design*, 3(3), 13-25.
61. Rogers, W. A., Lamson, N., & Rousseau, G. K. (2000). Warning research: An integrative perspective. *Human Factors*, 42(1), 102-139.
62. Schafer, R. M. (1994). *Soundscape: Our sonic environment and the tuning of the world*. Rochester, VT: Destiny Books.
63. Schön, D., Ystad, S., Kronland-Martinet, R., & Besson, M. (2009). The evocative power of sounds: Conceptual priming between words and nonverbal sounds. *Journal of Cognitive Neuroscience*, 22(5), 1026-1035.
64. Spence, C., & Zampini, M. (2006). Auditory contributions to multisensory product perception. *Acta Acustica united with Acustica*, 92(6), 1009-1025.
65. Suied, C., Susini, P., & McAdams, S. (2008). Evaluating warning sound urgency with reaction times. *Journal of Experimental Psychology: Applied*, 14(3), 201-212.
66. Timmers, R., & Ashley, R. (2007). Emotional ornamentation in performances of a Händel sonata. *Music Perception*, 25(2), 117-134.
67. Walker, B. N., & Kramer, G. (2005). Mappings and metaphors in auditory displays: An experimental assessment. *ACM Transactions on Applied Perception*, 2(4), 407-412.
68. Wilson, B., Bülow, G. J., & Hoyt, P. A. (2001). Rhetoric and music. In S. Sadie & J. Tyrrell (Eds.), *New grove dictionary of music and musicians* (2nd ed., pp. 260-275). New York, NY: Grove.