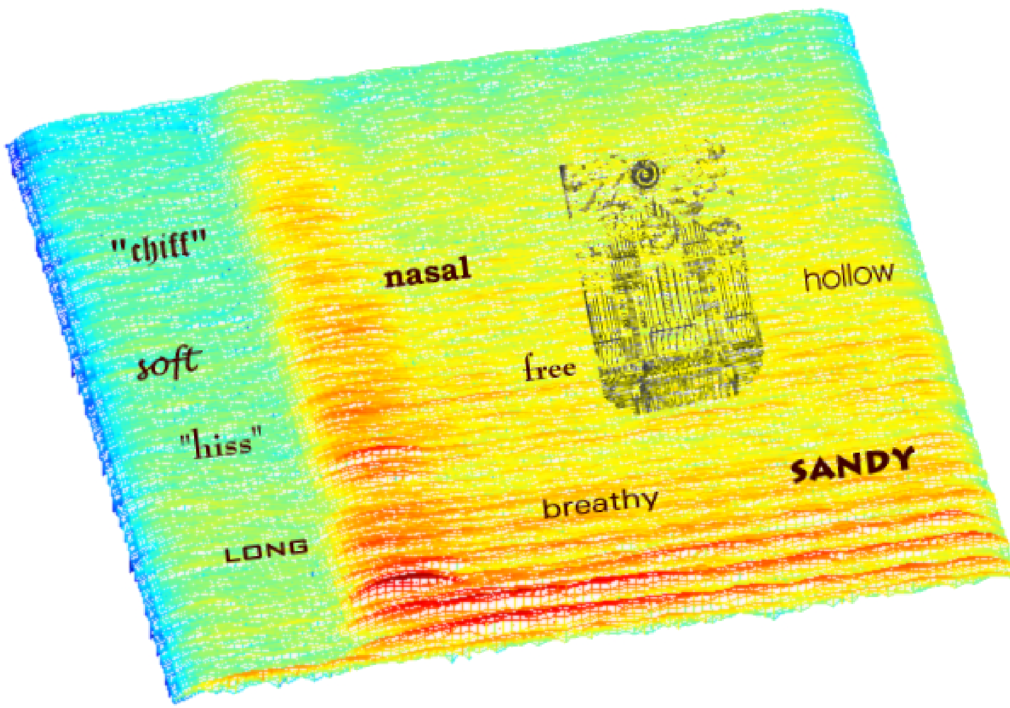
Sound Quality of Flue Organ Pipes
An Interdisciplinary Study on the Art of Voicing



Vincent RIOUX

Department of Applied Acoustics
Chalmers University of Technology
Göteborg, Sweden 2001

Report F 01-02

THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Sound Quality of Flue Organ Pipes:
An Interdisciplinary Study on the Art of Voicing

VINCENT RIOUX

Department of Applied Acoustics
Chalmers University of Technology
Göteborg, Sweden 2001

Sound Quality of Flue Organ Pipes
An Interdisciplinary Study on the Art of Voicing
Vincent RIOUX

Göteborg 2001
ISBN 91-7197-995-6

Doktorsavhandlingar vid Chalmers Tekniska Högskola
Ny serie nr 1680
ISSN 0346-718X

This thesis was prepared within the frame of the research project 'Changing Processes in North European Organ Art 1600-1970,' conducted at the Göteborg Organ Art Center, Göteborg University, and at Chalmers University of Technology, Göteborg, with financial support from the Bank of Sweden Tercentenary Foundation, Göteborg University and Chalmers University of Technology.

Doctoral Thesis in Applied Acoustics

Abstract

Rioux, V. (2001). *Sound Quality of Flue Organ Pipes. An Interdisciplinary Study on the Art of Voicing*. Department of Applied Acoustics, Chalmers University of Technology, Sweden

This thesis presents a research study carried out in collaboration with a department of applied acoustics, a department of musical acoustics and an organ workshop. The description of the sound quality of flue organ pipes has received fairly little attention either in organ-building or scientific literature, despite its importance in the overall quality of the instrument, probably due to the difficulties inherent in doing so. This thesis addresses this issue while focusing on the process of voicing a flue organ pipe. The treatment of such a topic requires an interdisciplinary approach in the use of methods and results that have originated in acoustics, signal processing, experimental psychology and linguistics.

The voicing process performed on a flue organ pipe can change several characteristics of the pipe sound. A review of practical and theoretical works shows that there is no well-established way of describing this voicing process, despite the fact that one can recognize different styles of voicings generally emerging from the aesthetics of a particular historical period or from the remarkable creativity of an individual organ-builder. Gathering information from a particular research organ workshop engaged principally in the reconstruction of baroque organs, an attempt is thus made to compare the descriptions obtained from an expert voicer with physical analyses made on experimental pipes or psychological analyses made on a pool of test participants.

Papers I and IV address specific methodological problems related to the administration and type of listening tests which were used in this work. The methodology presented in Paper I makes active use of computers for the administration, storage and analysis of listening tests. A specific programming environment was built for this purpose. Paper IV focuses specifically on a test module which combines the advantages of classical methods such as free verbalization, categorization, multiple comparison and semantic differentials.

Papers II and III show that it is possible to produce a structured list of verbal descriptors suitable for the description of flue organ pipe sounds through a combination of qualitative and quantitative analyses of test participants' responses. Three main steps were involved. First, an original list was constructed from the description of an expert voicer. This list was then extended by means of a qualitative analysis of test subjects' verbal comments. Finally, a factor analysis of a semantic differential test provided a structured normative lexicon.

Keywords: musical acoustics, psychoacoustics, sound quality, timbre, flue organ pipes, verbal descriptors, listening tests, time-frequency analysis, deterministic/stochastic separation.

Report F 01-02

Vincent Rioux, Chalmers Room Acoustics Group, Department of Applied Acoustics, Chalmers University of Technology, Sven Hultins gata 8A, SE-41296 Göteborg, Sweden, Telephone: +46 (0) 31-772 22 00, Fax: 46 (0) 31-772 22 12, Email: vr@ta.chalmers.se

Résumé

Cette thèse présente une recherche effectuée en collaboration entre un laboratoire d'acoustique appliquée, un laboratoire d'acoustique musicale et un atelier de facture d'orgue. Ce travail se propose d'explorer les différentes voies donnant accès à une description de la qualité sonore des tuyaux d'orgue à embouchure de flûte. Face à une telle entreprise, une approche interdisciplinaire s'impose, impliquant l'utilisation de méthodes et résultats issus de l'acoustique, du traitement du signal, de la psychologie expérimentale et de la linguistique. Sont tout d'abord abordés divers travaux ayant pour objet la facture de ces tuyaux d'orgue et plus particulièrement leur harmonisation. C'est à la lumière de ce premier chapitre que l'on remarquera les questions les plus saillantes. Comment rendre compte d'un travail artisanal voir artistique à l'aide d'outils scientifiques? Quelle est notre perception des sons travaillés par un organier? Est-il possible d'exprimer verbalement les modifications de la qualité sonore d'un tuyau durant son harmonisation? Quels types d'analyse physique peut-on mettre en regard de tels paramètres subjectifs?

En réponse à ces questions une série de tests d'écoute, administrés à l'aide d'une interface homme-machine dédiée, ont été effectués sur des sujets spécialistes du monde de l'orgue. Ceci a principalement aboutie à la création d'une liste de descripteurs verbaux susceptibles de favoriser la communication des caractéristiques de ces sons (principalement entre spécialistes). De plus, cette approche fondée sur une exploration des relations entre champs cognitifs et linguistiques se double d'une volonté de mise en correspondance avec le monde physique. Tous les tests d'écoute sont ainsi basés sur des enregistrements de « vrais » tuyaux ou sur un tuyau expérimental répondant à la fois aux contraintes de facture et d'investigations scientifiques.

Les articles I et IV traitent des problèmes méthodologiques spécifiques liés à l'administration et aux types de tests d'écoute utilisés dans ce travail. La méthodologie présentée dans l'article I s'appuie sur une utilisation poussée de l'ordinateur à des fins de présentation, stockage et analyse des tests. Un environnement de programmation a été créé tout spécialement. L'article IV se propose de détailler un module de test combinant les avantages des méthodes classiques telles que les comparaisons multiples, la catégorisation et verbalisation libres.

Les articles II et III démontrent qu'il est possible de produire une liste structurée de descripteurs verbaux adaptés à la description du son des tuyaux d'orgue à bouche grâce à l'emploi conjoint d'analyses qualitatives et quantitatives des réponses des sujets testés. Trois grandes étapes sont employées. Tout d'abord, une liste originale est construite d'après les descriptions fournies par l'organier. Cette liste est alors étendue par les résultats d'une analyse qualitative effectués sur les commentaires des sujets. Enfin, une analyse factorielle permet de proposer une liste normative et structurée.

Report F 01-02

Vincent Rioux, Chalmers Room Acoustics Group, Department of Applied Acoustics, Chalmers University of Technology, Sven Hultins gata 8A, SE-41296 Göteborg, Sweden, Telephone: +46 (0) 31-772 22 00, Fax: 46 (0) 31-772 22 12, Email: vr@ta.chalmers.se

Acknowledgements

This thesis would not have been written without the help and support of numerous people.
I first want to thank my supervisor Mendel Kleiner for his kind invitation and numerous advice. I then wish to address my sincere thanks to Daniel Västfjäll for sharing ideas, work and friendship.
To all members of the GOArt group, thank you very much for your warm acceptance and remarkable interdisciplinary work. No need to say that Munetaka Yokota, in particular, deserves all my gratitude for his constant support and help.
I also wish to thank all students and researchers at the “Laboratoire d’Acoustique Musicale” of Université Paris 6 particularly Benoît Fabre, Michèle Castellengo, Jean-Dominique Polack and Claire Ségoufin.

Ingrid Bersch and Åsa Abelin helped me with linguistics and speech therapy issues. Knut Guettler and Yinong Ding took some of their time to perform spectral analyses. Scott Peterson kindly provided us with an electromagnetic valve. Anne Cleaves carefully read and corrected my drafts. The friendship and hard work of Matthias Scholz and Malte Kob accompanied this work several times.

To all of you thanks a lot!

I finally want to warmly thank all the organ research workshop employees and all the Teknisk Akustik team for the nice work atmosphere they contributed to create, not to mention Börje Wijk and Gunilla Skog for invaluable help with practical matters.

This thesis is dedicated to my family and friends.

The financial support of the Bank of Sweden Tercentenary Foundation, Göteborg University and Chalmers University of Technology is gratefully acknowledged.

TABLE OF CONTENTS

Note:

- Words that appear in *italics* in the text are defined in the glossary.

- An accompanying hypertext document that has been made to illustrate the text of this thesis (audio, films, test data, routines, documentation, etc.) can be found at: http://www.ta.chalmers.se/homepages/vincent/index.htm.

This complementary information is also on CD-ROM may be obtained from the Department of Applied Acoustics, Chalmers, SE-41296, Gothenburg, Sweden.

This document is referred to in the text as "htmlDoc."

Preface

This thesis is the result of a project conducted in cooperation with the Gothenburg Organ Art Center (GOArt), the Chalmers Room Acoustics Group (Department of Applied Acoustics, Göteborg), and the Laboratoire d'Acoustique Musicale (Université Pierre et Marie Curie, Paris 6). The major part of the work reported here was done within the framework of the North German Baroque Organ Project, which led to the reconstruction of a baroque organ in Örgryte Nya Kyrka (Davidsson, 2000).

The interdisciplinary nature of this project (bringing together the work of musicologists, musicians, organ-builders and scientists) is reflected in this thesis, which is itself addressed to readers who might not share the same backgrounds. The structure of this thesis is thus somewhat hybrid, consisting of both a set of chapters and a number of articles and appendices. The primary goal of the present work is to identify and explore methods and experiments that can serve to construct a normative description of the *sound quality* of flue organ pipes.

In Chapter I, we delimit the area of flue pipe voicing as experienced by organ-builders and seen by scientists, seeking moreover to emphasize the descriptive approach used in both fields. This is followed in Chapter II by a more comprehensive treatment of the problems raised by the perception and description of sound. Chapter III is devoted to the presentation of the basic physics of flue instruments by means of measurements made on an experimental pipe. Further, this chapter mentions some engineering methods that are useful for the digital processing of sound files.

Chapter IV details a set of experiments, the results of which have previously been presented at acoustics or organ conferences.

Chapters I, II and III can be read separately and are also intended as contextual introductions to Chapter IV, and to Papers II, III and IV.

PAPER I presents an overview of methods used in order to try to map physical and perceptual *attributes* or features of some flue organ pipe sounds and subsequent results obtained. This paper documents our first attempt to check the generalization of some verbal descriptors used by an organ-builder.

PAPER II is devoted to an analysis of free verbal descriptions made by participants, resulting in the proposal of an extended list of verbal descriptors. The structure of this list is arbitrary. (Daniel Västfjäll did 50 percent of the work.)

PAPER III is a companion to Paper II and achieves a structured analysis of the list of verbal descriptors by means of factor analysis of a dedicated listening test. (Daniel Västfjäll did 20 percent of the work.)

PAPER IV presents a new listening test method based on a computer graphical interface.

We tried as much as possible to use hypertext technology to ease both on-screen reading, audio consultation and overall illustration. This is reflected in the creation of an extra piece of information available as an html document on a CD-ROM or a web page.

I - Voicing of a Flue Organ Pipe

> A **voice** is also a sound that resembles or reminds you of a human voice. (Collins Dictionary)

In his famous treatise, Dom Bedos de Celles (1766-78) describes the art of organ building from an encyclopedic point of view. All aspects of the technical work are reviewed in detail, from the construction of tools and architectural considerations to pipe scaling, but the final adjustments of the pipes, the voicing, are somehow eluded. This is probably due to the inherent difficulty of this task.

In this chapter, we will try to describe the basics of pipe making and indicate what the art of organ voicing is concerned with. This very introductory text should provide some initial indications of what can be expected of a scientific enquiry and define the parameters of the primary elements of our study.

First, a summary of interviews and recordings made between February and April 1997 in the Organ Research Workshop[1] will introduce the basic knowledge involved in pipe making and pipe voicing. It is important to note at this point that voicing techniques are intrinsically linked to esthetic values and thus have followed historical trends. We do not presume here to give a general overview of voicing techniques over the whole historical development of the organ, which spans more than two thousands years (see Williams (1988) for a complete study of the history of the organ). In this thesis, detailed aspects of voicing will concern a case study of an attempt to reconstruct a North German baroque organ. For the sake of generalization, we will thus be interested in generating a set of methods, experiments and results that can be applied and expanded to different kind of voicing techniques and aesthetics, and even to other types of sounds or musical instruments.

Pipe making and scaling are topics that have received much more attention than voicing in both technical (Bédos de Celles (Dom), 1766-78; Cavaillé-Coll, 1979; preface of Monette, 1992) and scientific literature (Miklós *et al.*, 1998; Sundberg, 1966). Nonetheless, some sources are available on this topic in various publications. We will then proceed with a review of previous works that have been done by organ-builders and scientists. Finally, we propose a way to follow the voicing of a single pipe, and discuss the concept of voicing steps.

[1] Organ-builder Munetaka Yokota leads the Organ Research Workshop at Varbergsgatan 2, Göteborg, Sweden. He will be referred to as the Voicer or the Organ-builder in the rest of this document.

I.1 - Summary of notes taken in the Organ Research Workshop[1]

These notes were taken during a period of nearly three months and further refined through the constant cooperation carried on between the author of this thesis and the Organ-builder. Interviews were made in the workshop and complemented by notes, recordings and sketches. This work was mainly motivated by a specific project concerning the construction of six flue-organ pipes copied from two historical pipes (see Table 1).

I.1.a - Pipe making and scaling

During the planning of the construction of an organ, the *stop* (or registers) and mixture composition have to be specified (the design process). This will fix the ratio between length and diameter of the pipes, which will in turn determines the radiation impedance (Miklós *et al.*, 1998) of the passive ending and thus in a first approximation, the harmonic balance of the resulting tone. A "micro-scaling" dealing with the mouth height and width is then decided, following empirical rules. These choices have been rather well classified (see Figure 1 for an illustration of a sample of stops names and scaling).

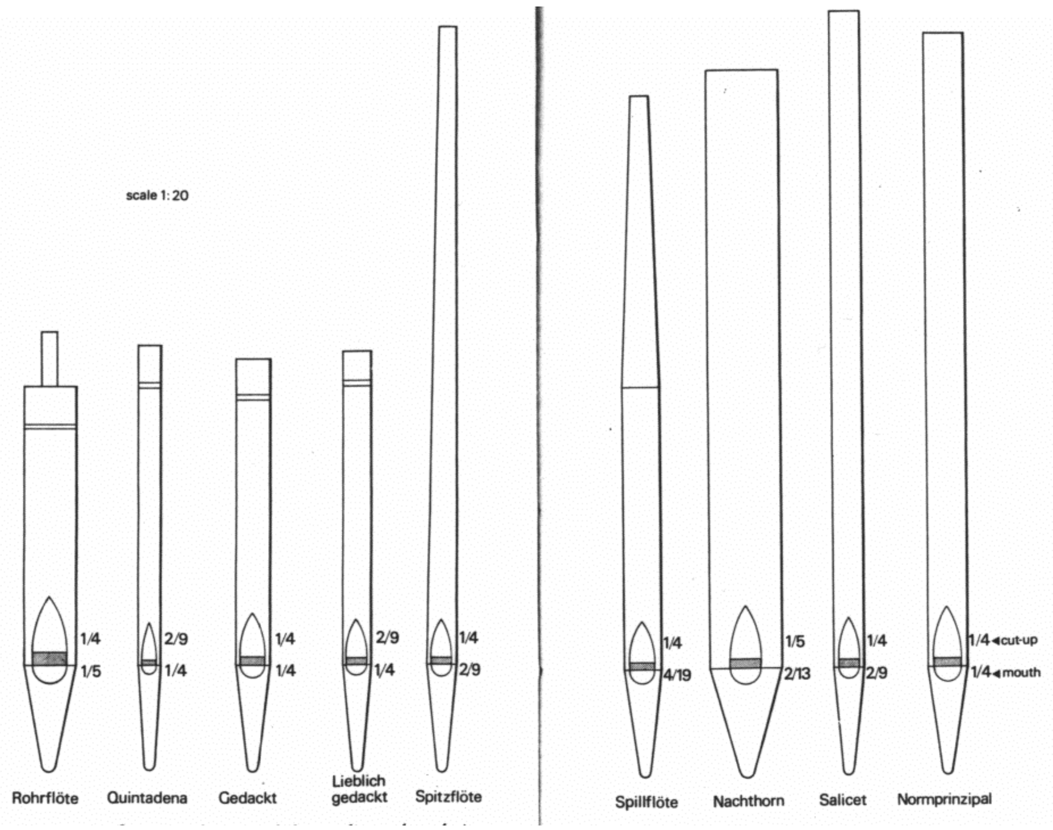


Figure 1. Most common stops of flue organ pipes with the mouth scale specification ("micro-scaling"). The first ratio indicates cut-up (mouth height) to mouth width (H/W) and the second one mouth-width to pipe circumference. from (P. Williams, & Owen, B., 1988)

It is only after these two steps that the pipe-making process really begins. At this point, the alloy and the thickness of the walls, the languid, mouth shapes and the toe size are decided (see Figure 2 and Figure 3). All these choices considerably influence the amount and type of voicing to be done and consequently deserve special attention.

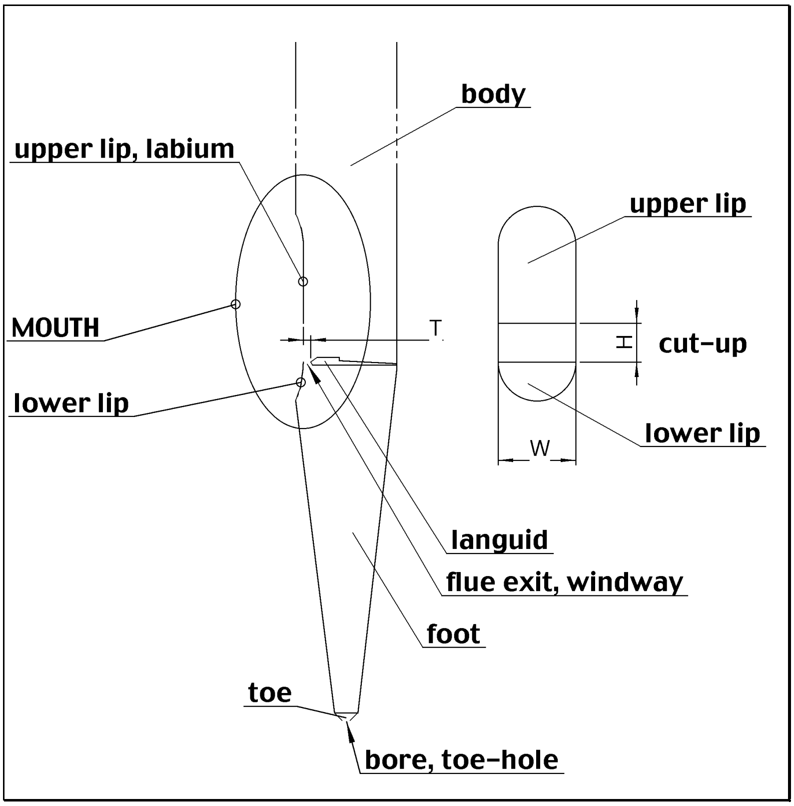


Figure 2. The different parts of a flue organ pipe.

The main steps of an historical pipe-making process, as implemented in the GOArt research workshop, are presented in chronological order with some added remarks:

1. Casting:
Flat sheets of alloys are cast on a special sand bench.
2. Scraping:
The scraping is one of the several methods—to which hammering also belongs—aimed at ”cleaning” the rough metal sheets just cast.
3. Specifying and **cutting** metal pieces for the body, the foot and the mouth:
The body, the foot and the mouth shapes (see Figure 2) are cut with special attention to the edges at which they will join.
4. Forming toe:
The surface of the toe (St) is to be compared with the area of the flue exit (T·W) (see Figure 2).
5. Forming the languid:
Cast and plane a thick plate of metal according to the shape described in Figure 3 below.

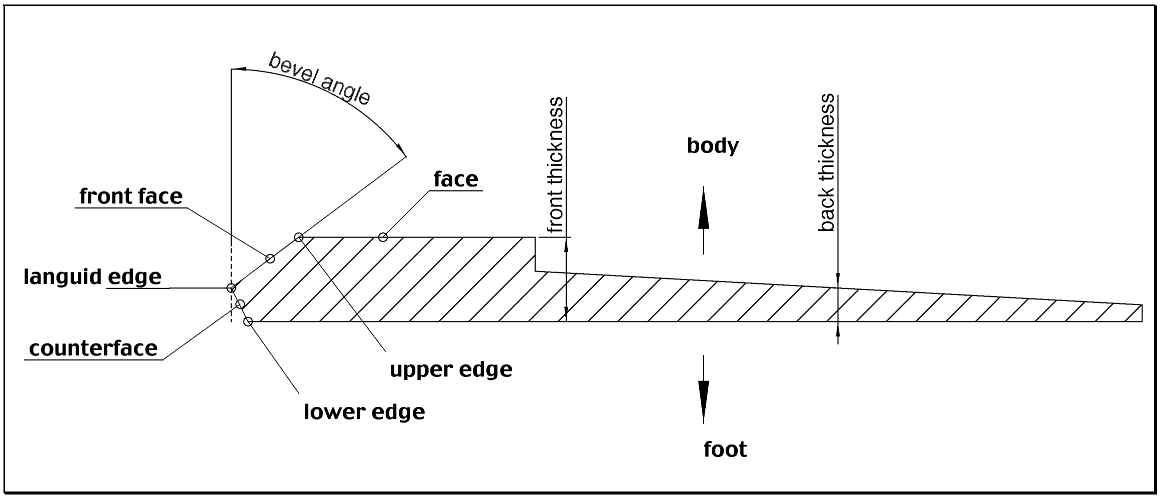


Figure 3. Sketch of a languid in the Schiörlin[2] style, illustrating the complexity of its geometry.

6. Forming the cylinder shape of the body and the conic shape of the foot.
7. Mounting the languid.
8. Soldering body to foot.

At this point, the pipe may “speak” or not: it might not sound at all when blown. The art of voicing begins to be employed at this stage and continues until the final step, when the pipe is placed in the organ chest.

I.1.b - Pipe voicing / task analysis

Voicing techniques refer to the geometrical changes made by the organ-builder on each pipe of each *stop* (see Figure 1). These techniques have been developed to allow extremely fine changes in the *sound quality* of organ pipes. The voicing techniques are based on empirical or trial-and-error knowledge which has been accumulated, refined (and sometimes forgotten) over centuries. As already stated, they have gone through changes according to the aesthetics of different times. The ultimate goal for a voicer is to harmonize all the pipes of an organ so that criteria (both conventional and personal) for the following **topics** are satisfied:

- temperament (tuning)
- room acoustics influence
- pipe coupling influence
- pressure fluctuations influence (Carlsson, 1999)
- control of the timbre over a whole stop
- blending of different *rank* for mixtures
- blending of different stops.

[2] Pehr Schiörlin, Swedish organ-builder who built the Jonsered organ in 1783.

The voicer changes the geometry of the pipes according to the following desired transformations of the **sound**:

- modification of the timbre
- adjustment of the stationary part (relative levels of harmonics, *noise* content)
- adjustment of the *transient*/speech (speed, noise character, harmonic content)
- *loudness* adjustments
- tuning compensations.

We must emphasize the fact that the sound production of flue organ pipes is dependent on many interconnected factors. These interdependencies make it difficult to delimit precisely where the voicing begins and ends in the pipe-making process. One could, for example, integrate in the voicing the choices of body, foot geometry, wind pressure (static pressure in the wind chest) or even the type of metal used for the pipe walls, as these choices will influence further refinements. The static and dynamic pressures in the foot are also of utmost importance for the sound produced by the pipe. The pressure rise start-up in the foot is directly correlated with the initial velocity of the jet and thus to the transient. This foot pressure should then be regarded as an important parameter of the voicing.

In what follows, however, the field to which the term “voicing” is applied is precisely delimited. In fact, if we make a distinction between fixed (e.g. scaling, alloy, wind pressure) and variable (mouth geometry) parameters of voicing, we choose to restrict our definition of voicing to all adjustments concerned with variable parameters. Consequently we will later be making a distinction between *scaling parameters* (fixed parameters) and *voicing parameters* (variable, adjustable parameters of the mouth area).

Voicers have designed a set of tools to modify the geometry of the mouth. Among many variables the following **geometrical parameters** were chosen, as they are of first order importance:

- mouth opening dimension (H/W ratio)
- cut-up height to jet thickness ratio (H/T)
- relative position of the upper-lip to the middle of the flue exit (y0)
- Proportion between the areas of the foot hole and the flue exit.

All these considerations are extracted from notes taken during a set of experiments. Our first experiment was based on the reconstruction of two historical pipes. These two pipes, a principal 4’ G# and a principal 4’ c3 were taken from an organ—made in 1783 by Pehr Schiörlin—in the Jonsered church (in the vicinity of Gothenburg, Sweden). Because the purpose of our research includes the study of the effect of the *nicking*, non-nicked pipes were selected, although they were not necessarily representative for the late 18th century but are closer to the mid-17th century style which corresponds to our special period of interest. Thus, the Voicer made three copies of each pipe. The pipes were of varying wall thickness, had varying compositions of the tin-lead alloy and also varied in regard to the presence of *nicking*. The main goal from the voicer’s point of view was to try to make each copy sound as close to the original as possible, regardless of the difference in material or the construction of the pipes, and thus to use the many resources of the voicing techniques.

The following nomenclature was set:

Table 1. Pipes nomenclature.

pipe number	**type / specifications**	**alloy composition**	**reference (htmlDoc)**
#1	original principal 4’ G#	(tin-rich pipe, 100% Sn speculated)	s0
#2	original octave 4’ c3	(lead-rich pipe, 17% Sn speculated)	
#3	copy of 1	(100% Sn with impurities, material is not strictly simulated)	s6
#4	copy of 2	(17% Sn)	
#5	thick wall copy of 1	(100% Sn)	s8
#6	high-tin wall copy of 2	(100% Sn)	
#7	like #3	with nicking added	s9
#8	like #4	with nicking added	

It has been possible to follow each step of the voicing as applied to the particular pipe, in this case the one corresponding to #3 (see Table 1). As for the pipe-making process, it is very interesting for both the organ-builder and the acoustician to try to define a chronological frame for the voicing procedure. In general the empirical techniques only provide a sort of “fuzzy” convergence towards the expected sound. The sequence of steps implemented during the voicing of the copy #3 of the G sharp 4’ principal pipe is presented on the next page with full respect to the chronology chosen by the organ-builder.

Step 0. ”How does the crude copy sound?”
The cut-up is very low (H small), and this is the occasion to record the influence of the modification of this parameter (H) by itself (htmlDoc:step0)
Step 1. “make the pipe sound”
The lower-lip is slightly hammered to reduce the width of the windway (htmlDoc:step1).
Step 2. ”cleaning”
The inside edge of the lower-lip and the edge of the languid (see Figure 4) may have some burrs. These are removed using a special knife in order to reduce a certain type of undesired *noise* (i.e. “**chiff**” or the noise of the stationary part).
Step 3. *pitch* tuning
Make a roll (tuning-slot) (htmlDoc: s1).
Step 4. ”fastening speech”
The languid is knocked down in order to allow the jet to go deeper into the pipe (htmlDoc: s2).
Step 5. comparison between the copy and the original pipe:
The copy has less fundamental than the original pipe. As regards the transient part, the original pipe displays a somewhat high *pitch* (first overtone at the transient) and its transient part is faster than the copy, whereas the copy does not have this prominent octave overtone.
Step 6. ”balance the harmonic content of the steady part”
To increase the fundamental, the cut-up is heightened (H is raised) (htmlDoc: s3).
Step 7. loudness adjustment
In order to compensate the low loudness of the sound, the windway is opened up (increase T). This technique is closely linked to the orientation of the languid. When the windway is opened up to give a ”**broader**” or ”**louder**” sound, it also affects the direction of the jet (directed outward) so that it has to be compensated by a lowering of the languid. Thereby, the jet goes further toward the inside (htmlDoc: s4).
Step 8 second correction and compensation with languid (htmlDoc: s5)
Step 9 ”last fine adjustments”
The cleaning of the sound is related to the windway opening so that the windway has to be closed a little, which in turn may involve a lifting of the languid (htmlDoc: s6).
Step 10 final judgment
The languid had to be set a little lower than the original one. This results in a somewhat ”**darker**” sound (the first harmonic is weak compared to the fundamental) (htmlDoc: s7).

All the steps presented above are the result of a complex interaction between the physical system and the perception of the voicer.

Table 2. The set of recorded steps.

Nomenclature	**Pipe**		**Desired effect**	**Comments / process**
s0	#1	original pipe		
s1	#3	voicing steps of a geometrically reconstructed pipe	pitch tuning	lentgh was cut
s2	#3		fastening the speech	the languid is knocked down
s3	#3		equilibrate the harmonic content of the steady part	cut-up raised
s4	#3		increase **loudness**	opening-up the windway
s5	#3		correction / compensation	windway down + languid lifted
s6	#3		final adjustment / cleaning	idem
s7	#3		idem	last recording
s8	#5	thick wall copy		
s9	#7	copy with **nicking**s on languid		

The study of such interaction must take into account the influence of the "interviewer" (here an acoustician) as depicted in Figure 4.

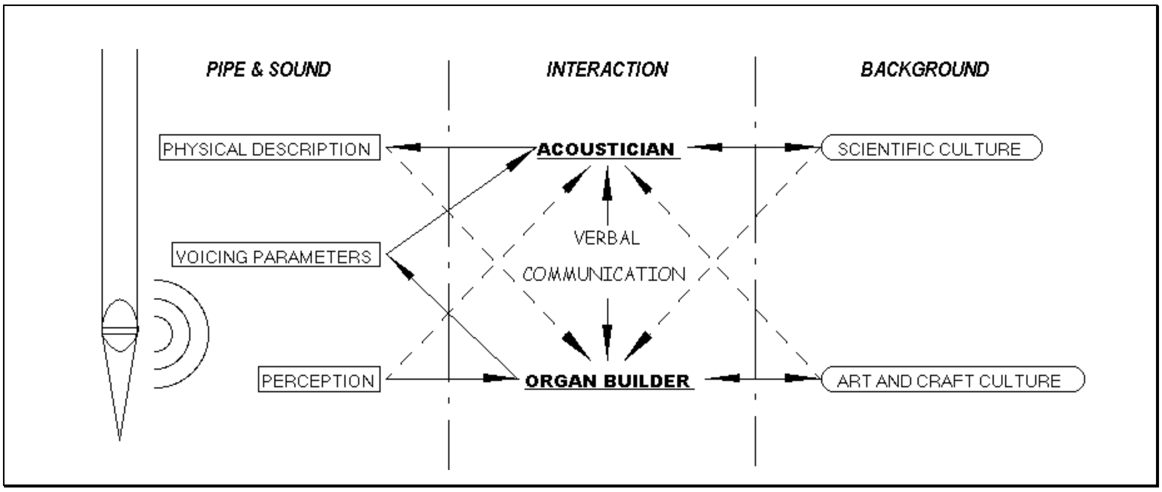


Figure 4. Graphical representation of the communication process used during interviews made in the organ workshop.

It may require several geometrical adjustments to perform a modification of a single *attribute* of the pipe sound. Each "perceptive" step is motivated by a desired change in the sound as perceived by the Voicer. The only possible way for an external observer to detect these perceptive steps is to use "verbal communication" (see Figure 4). Before exploring in details the possibility of defining a verbal protocol (see chapter II.3 - p. 32), we can already acknowledge the fact that these changes were experienced by the interviewer and emphasized by the Voicer through the use of verbal descriptors (see steps 2, 7, 10). We hence collected the descriptors used by the Voicer (see Table 3).

Table 3. List of descriptors used for the description of the voicing of flue organ pipes from the Voicer (private communication).

Part of the sound	**Attributes**
speech or **transient**	"Chiff"
	"Cough"
	"Hiss"
	"Kiss"
	Soft Vs strong
	Slow Vs fast
	Short Vs long
	Pitch (e.g. octave)
steady state	Amount of fundamental
	Amount of quint
	Amount of octave
	Stringy Vs fluty
	Round Vs sharp
	Full Vs thin
	Light Vs heavy
	Nasal
	Breathy
	Dirt Vs clean
	Free Vs tight
	Floating/spatial Vs oppressive
general impression	Intense
	Sandy
	Sweet
	Hollow
proportion between	Fundamental and overtones
	Speech and steady state
	Noise and musical tone

We will come back in Section II.3 - on the signification of the different types of descriptors found in this table.

I.2 - Review of practical writings (by organ-builders)

When mentioning organ-building writings, one might immediately think of the historical treatises of Dom Bédos de Celles (1766-78) or the theoretical works of Cavaillé-Coll (1895). Dom Bédos, despite his fundamental encyclopedic work, devoted only two chapters to the art of voicing, and as Monette (1992) puts it, “His material is accurate but sketchy, scarcely more than an expanded definition.”

Cavaillé-Coll, in his “Theoretical Works,” presented a set of experiments made on flue pipes motivated mainly by his desire to relate the physical functioning of an air-jet with that of a free-reed. According to him, in order to obtain a “good sound quality” out of a wooden pipe (geometry similar to a recorder, see Figure 5), a voicer should respect at least these four points :

1. The front surface of the block (inferior wall of the canal) must be plane and in the extension of the inferior part wall of the upper-lip.
2. The superior wall of the canal, either presenting a plane or curved surface, should be such that all the lines perpendicularly drawn to the direction of the current should be parallel to the inferior wall of the canal.
3. The edges of the walls must be sharp.
4. The bound of the lower-lip should be proportioned to approximately one-tenth of the cut-up height.

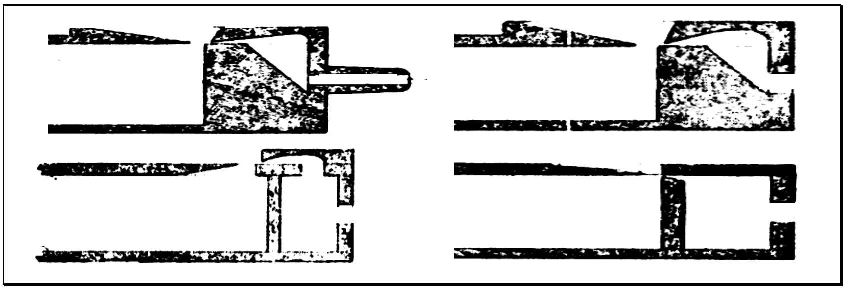


Figure 5. Different configurations of the mouth geometry of wooden pipes, from Cavaillé-Coll (1895)

Nevertheless, the sound quality which he often mentions is not further explained in any sense but by its opposition to “purity” (p.13) and “intonation stability”; this supposes then that there should exist an obvious “good sound quality” for a pipe. It is may be of interest to point out that Cavaillé-Coll uses a few verbal descriptors only to characterize a “bad” sound quality (“sourd” dull, “âpreté” roughness or harshness, and “frisement” rustling). He was certainly more interested in describing the physical properties of flue pipes than in describing their sound quality, probably because, as he wrote it, that when voicing, “On tâtonne” (“One gropes”) (p. 33). This led him to build an experimental pipe with a moveable upper-lip and conduct an extensive set of experiments. He remarked on the presence of a very high-pitched sound, which frequency could be related to the cut-up height (H), most probably the edge-tone (p. 17). He also affirmed that *nicking* should be avoided when possible but that it is often necessary on medium pipes.

Despite the difficulty of the task experienced by their predecessors, some contemporary organ-builders have tried to report on their experience of voicing. Probably the most complete technical work was published by Monette (1992) in his “Art of Organ Voicing.” Following the pipe making and scaling, Monette distinguishes three phases of voicing: the preliminary voicing, the regulation and the *tonal* finishing and fine tuning. All these phases are described in a very practical manner so that they may be transferred to actual practice immediately. His definition of Timbre is uncoupled from speech and considered to be the same as Tone Color (the harmonic spectrum, see Section III.2 -). We will later discuss an alternative choice. Monette discusses the importance of the *noise* components of the transient part. Although an expert voicer has to adjust their presence in a “natural” way (not too little, not too much), they are said to be essential in musical speech (p. 74). Verbal descriptors are used throughout the book, and some are explained in a glossary (bright, chiff, dull, slow, *white noise*), but no systematic way of describing sounds of flue pipes is given.

Monette also mentions issues affecting the sounds of pipes like the chest coupling (p. 70) and the influence of the room.

Table 4. List of verbal descriptors collected in “The Art of Organ Voicing” (Monette, 1992). Monette gives a definition of words in *italics.*

bright
chirp
cooler, **lighter**
dull
gritty
husky
lighter
nasal (basic vowel tone not being altered)
pungent, **sharp, caustic**
slow
neutral sung "**ah**"
“chiff”
"**ee**" with distinctive nasal quality
"**i**" as in (sit)
”**schwa”** (an unstressed mid-central vowel)
"**uh**" or "**aw**"
"**uh**" or "**oo**"

We will return to this table Chapter II, focusing particularly on verbal description.

In a similar vein, Pelto (1995) provides us with a technical and “perceptual” description of four voicing techniques: namely, the low cut-up, medium cut-up, open toe voicing and high cut-up . He made a comparative study of twelve diapason pipes. Four sets of such pipes were voiced according to the different voicing techniques. Merging properties of a combination of pipes were also under study; consequently he chose for each set, pipes constituting a major triad (f1-a1-c2).

Table 5. Description of four voicing techniques from Pelto (1995) Adapted with permission.

	low cut-up	medium cut-up	open toe voicing	high cut-up
ratio H/W	1/5	1/4	1/4	1/3
comments			large foot hole and small flue exit height (T)	
general note	W is one-fourth of the circumference of the pipe			
timbre	**thin** and **quiet**, like a gamba, **bright**, becomes **sharp** if voiced stronger; **cool** sound.	more **mellow** than a low cut-up, the second harmonic is powerful; a somewhat special timbre.	like the vowel u, **firm** and **bright** not **profound**, beautiful and active, well-balanced; rather a **warm** sound.	A **big** and singing sound, a well balanced, not particularly interesting timbre; an exceptionally **warm** and **broad** sound.
fundamental	feeble	medium	medium	powerful
attack/onset	long, detached from the stationary sound	pitch near to the pitch of the stationary sound	pitch near to the pitch of the stationary sound	fast, not striking, nearly unnoticeable
remarks	A distant and reserved sound; a wider scale could be used for a more rich sound	A little problematic to voice; some small nicks were needed	The sound comes very close to the listener	The sound surrounded the listeners. The scale could have been a little narrower

Note that here, four parameters are used for a voicing, namely the height of the cut-up (H), the width of the flue exit (T, the jet Thickness at the base of the flue), the air velocity and the fundamental frequency. These parameters are all involved to calculate what Harmuth Ising (1971) first called the Voicing Number (see Equation 1). According to Liljencrants (1999), the intonation number I of Ising should be more than 2 but less than 3: “With I=2 you get maximum efficiency. With higher values you get stronger harmonics and when you go past about 3 the pipe will overblow.”

$$I = \frac{U_0.\sqrt{T}}{f_0.H.\sqrt{H}}$$

where,

U_0, is the mean jet velocity at the flue exit.

$U_0 \approx \sqrt{\frac{2(P_f - P_0)}{\rho_0}}$ (Bernoulli stationary eq.)

T, the Thickness of the jet at the flue exit

f_0, the frequency of the fundamental.

H, the cut up Height.

P_f, static pressure in the foot.

P_0, atmospheric pressure.

ρ_0, air density

Equation 1. The "Voicing number" from Ising (1971).

The height of the cut-up is conclusively said to be related to the volume of the sound (increases with T) and to the proportion of high harmonics (increases when T decreases). Pelto, certainly because he is deeply involved in organ making, chose to deal with real organ pipes as objects of study. Though it is clearly closer to practical interest, the physical understanding may be somewhat overshadowed by the intricate complexity of such objects.

However, there is an alternative to this method, which will be further discussed (see Chapter III - Appendix I : Construction of the experimental pipe). Instead of dealing with real, complex structures, one can try to build or simulate a simplified prototype under the assumption that the main behavior is conserved. This allows one to take fewer parameters into account.

We believe that the two approaches are complementary and thus that they should be equally appreciated and also combined.

I.3 - Review of theoretical writings (by scientists)

In parallel to the work of practicians, physicists have studied in detail the physics of organ pipes (see section III.1 - 46). Only a few of them have tackled voicing as a specific subject of study.

Mercer, in two articles (D. M.A. Mercer, 1951; D.M.A. Mercer, 1954) gives qualitative details on the voicing of different kinds of flue organ pipes, principally on diapason pipes. The physical functioning is also investigated (though understanding has progressed somewhat by now). The experimental part is conducted on a "real" open diapason pipe, two feet long, speaking middle C (261 Hz). Several voicing factors (under the control of the organ-builder) are then given. Each adjustment is considered in isolation from the others. Mercer points out nevertheless that "this separate treatment of each adjustment is somewhat artificial, since an experienced voicer will frequently make several adjustments simultaneously." Even if no light is really shed on the language used by the organ-builder, some sound descriptors like round, dull, keen, bright are used.

Nolle (1979) built a prototype of a flue organ pipe on which the following characteristics could be adjusted or changed (see Figure 6):

- cut-up height (H or Y in his paper)
- relative position of the upper-lip (labium) to the middle of the flue exit (Y or X in his paper)
- cross-section shape of the resonator (circle or square).

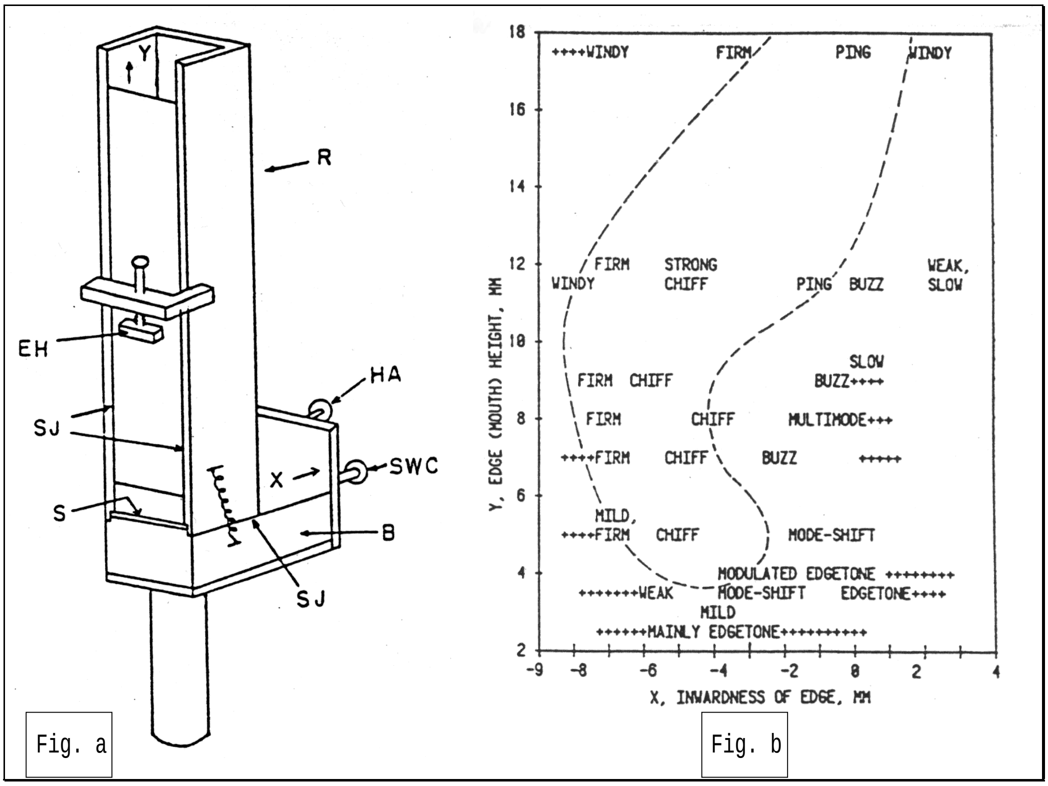


Figure 6. Adjustable flue pipe (a), and qualitative description of pipe speech (b) from Nolle (1979).

In Figure 6b, some verbal descriptors concerning the transient of the pipes are plotted against *voicing parameters*. The descriptors used by Nolle to describe the sound are explained in the following table:

Table 6. Short glossary of verbal descriptors used by Nolle (1979)

Buzz	similar origin as modulated edgetone (at the speaking frequency). The waveform with buzz present has high frequency parasitic oscillations which switch on at consistent points during the fundamental period, but which may not repeat in detail. This is apparently a burst of noisy edgetone which ceases whenever the fundamental oscillation moves the jet into an unfavorable position.
Chiff	used here only for the kind of transient suggested by the sound of the word, a brief burst of relatively broadband noise. Although the pitch of this is more ambiguous than that of the ping, the spectrum of the chiff appears to peak somewhere near a frequency at which pinging could occur.
Edgetone	refers to the well-known signal produced by a jet at a sharp edge.
Firm	indicates particularly stable oscillation which builds up quickly (as distinguished from slow build-up) but without ping or chiff.
Ping	refers to an almost periodic oscillation, typically near the third or fifth harmonic of the steady-state fundamental in the case of a stopped pipe, which gives a xylophone-like character to the attack.
Windy	indicates broadband noise accompanying relatively weak oscillation, usually with conspicuous amplitude instability.

The works of Mercer and Nolle are good illustrations of two basic ways of studying voicing techniques. The work of Nolle symbolizes here the simplifying/modeling approach. Instead of using a real pipe, he chose to work on a simplified prototype. This approach is somewhat closer to the theoretical works of Coltman (1976), Fletcher (1976), Fabre et al. (1996) and Verge et al. (1997). Still, it is important for our purpose to point out that Nolle is the only author quoted here who has put so much emphasis on the use of verbal descriptors, including *onomatopoeia*.

By contrast, the papers of Mercer illustrate studies concerning real pipes. In the same vein, McNeil (1983) carried out a set of experiments on different pipes for the purpose of isolating important factors in the art of voicing. Steady-state and transient parts are studied and their characteristics are related to change of various voicing parameters. At the end of his study, he makes an important remark: “My assumption is that *the artistic control of change* is the standard by which sound should be judged.”

Angster et al. (1997) made a precise and systematic study of the relationship between the voicing steps and the evolution of harmonics for two diapason pipes (2’ c and 4’ F). The goal of this study was to verify the match between measured and expected effects of different voicing adjustments. Voicing steps are classified (see Table 7) and we take the opportunity here to propose an alternative nomenclature. Angster et al. describe both stationary and transient parts of the sound using time-frequency analysis, though restricted to the evolution of harmonics (noise and mouth-tones are not considered).

Table 7. Nomenclature of geometrical steps (operations on the mouth geometry). Revised version of the nomenclature of voicing steps used in (Angster, 1997).

Description	**According to Angster**	**Condensed**	
Cut-Up	A	CU+ (/ CU-)	+ = increased
Languid	B	LA+ / LA-	+ = raised
Lower-Lip	B'	LL+ / LL-	
Flue Exit	C	FE+ / FE-	
Upper-Lip	D	UL+ / UL-	+ = bends up
Toe Hole	E	TH + / TH-	
Nicking	F	N	
Ears, rollers	G		
Tuning device	H		
Adjust Languid Edge		LAe	
Adjust Lower-Lip Edge		Lle	

Castellengo (1969) initiated a detailed description of the influence of scaling and voicing parameters on the sound of flue organ pipes. She discussed scaling parameters and their influence on the harmonic spectrum as well as on the pipe modes. Voicing parameters are also treated with special attention to their influence on the attack transient, the edge-tone and the so-called mouth-tones. Castellengo extended these experimental observations to the family of flute-like instruments (Castellengo, 1976) and made a detailed study of the attack transients (Castellengo, 1999).

Both Angster and Castellengo emphasize the fact that the study of isolated voicing parameters is nearly impossible in practice. A step of voicing is actually a modification of a set of parameters intended to change a single perceptual *attribute*. This is why, in order to get a valuable description of the voicing art, we need to be able to characterize these "perceptual" steps.

I.4 - Summary

From this review of various writings concerning the voicing of flue organ pipes, one might conclude that this is a complex subject. A voicer has to deal with a great many interconnected parameters (physical and perceptual as well as esthetic). A voicer has to take into account complex acoustical issues like characteristics of transients, harmonics balance, *noise* components, influence of pipe-wall vibrations (Kob, 2000; Svensson *et al.*, 1999), acoustic interference problems like mixture or coupling (Johansson *et al.*, submit. 2000) and room acoustics. With regard to the influence of room acoustics, it is interesting to note that in the literature we did not find specific mention of voicer assistants. As a matter of fact, a voicer should be able to listen to the sound of a pipe both at the source (where he has access to the mouth parameters) and at a distance, in the church, where he can perceive the sound as experienced by the audience. This is an important note for us, because in order to do that, a voicer will have to hire an assistant and ask him to report on the sound he perceived. There is here the trace of a communication process. This could effectively help an observer to disentangle the **project** from the **action**. The voicing task involves both artistic and technical skills. And as Monette (1992) says, we need to separate the taste from the technique. Taste is something that we cannot predict; it might even evolve within the course of a single individual career. In other words, we shall not try to tell or ask if one sound is better than another. Technique is supposed to be much more transparent. Our point here is that unfortunately it appears that techniques involved in the art of voicing are not easy to describe. If we can easily list the different types of modifications and their associated tools (the **action**), we do not have direct access to the goal of these modifications in terms of the sound perception (the **project**): There is something else besides taste that is not really considered as a technique but should be*: the evaluation of the sound quality*. The only judgments we will consider in our study concern the **relative** qualities of sounds. It is striking that these judgments are present in all the writings presented above. Voicers and scientists report on these kinds of judgments using verbal descriptors, but they do it in an informal way. In fact, we believe that there is a need for a formalization of the perceptual steps involved in voicing techniques and that this formalization can be approached by the elaboration of a structured and specialized lexicon.

II - Listening to and Describing Sounds

In the previous chapter, we very briefly introduced a framework for a description of the sensory and communication channels involved in the voicing process as perceived by an observer (in this case a scientist, Figure 4). We also presented a number of verbal descriptors concerning the *sound quality* of flue organ pipes. These descriptors were collected both from organ-builders' treatises and scientific articles (often quoting actual organ-builders' terminology). We want to stress here the **provenance** of such verbal descriptors because it provides us with some information about the type or degree of expertise of its users and hence a **context** of listening habits and culture. We do not of course expect that such verbal descriptors can fully grasp the perceived quality of the sound of a pipe, but their existence testifies, at the very least, to a need for communication.

In this chapter, we will try to sort out and address the rather difficult questions that may arise from such remarks as:

- Is there a need for communicating our perception of sounds?
  What are the possible communication channels?
- How do we listen to a "musical" sound?
  What are the physiological and cognitive processes involved in the evaluation of "musical" sound quality?
- What is timbre?
  What are the dimensions and features *of timbre?*
- Can we verbally describe sounds?
  How can we build a dedicated verbal lexicon?
- How can we have access to these descriptions?
  What are the adapted psychological methods?

This will lead us to introduce several concepts that psychoacoustics borrows from various fields of science such as physiology, linguistics and psychology.

Of course, all these questions are complex ones and cannot be answered fully at this time. However, they constitute the backbone of the context in which our listening tests were made (presented in Chapter IV and in the four articles collected in the Appendix). This chapter thus represents the core of our hypotheses and consequently the main paradigms of this research study.

Before proceeding, we kindly ask the reader to keep in mind that this thesis is concerned for the most part with the description of **small**, **relative, timbral** changes that occur to the sound of a flue organ pipe while it is being voiced. This is of crucial importance, because as we will see in what follows, a corpus of sounds associated with a certain context delimits different approaches to the study of auditory perception.

II.1 - Is there a need to communicate the sound quality of flue organ pipes?

It is likely that the majority of instrument-makers cannot or do not want to use words explicitly to describe the *sound quality* of their products. This certainly has to do with the very high quality of musical instruments that hardly suffer from sketchy and highly subjective description. It turns out in fact that our auditory lexicon of names is as a whole rather poor (see section II.3 - p. 32). Nonetheless, under certain circumstances, instrument-makers might have to communicate their auditory impressions of sounds. They have learned to perform a very detailed analysis of sounds (see Section IV.3 -). This analysis could be compared to a “virtual” digital analyzer (see Section III.2.c.iii - , p. 61) which would have integrated the physics of the instrument, the physiology of the human ear and the psychology of musical listening as well. In fact, instrument-makers also are and should be sensitive to the “sound quality” requirements of musicians and composers of their time or of remote times (e.g. baroque. In any case, how this kind of communication happens is not established. We have no definite language available to express musical sound quality. The sound of an instrument is deeply rooted in its design and can hardly be transmitted except by a somewhat fragile master-to-disciple movement. This might explain why, for example, we have kept some sounds from the past (coming from surviving original instruments, e.g. Stradivarius violins or old baroque organs) but we might have lost their design.

But this does not mean that it is impossible to describe sounds verbally, at least in a defined context. This is the contextual issue that we want particularly to address in this section. The table 8 is intended to clarify such possible contexts of verbalizations. It is most likely that the various contexts of listening and the different types of potential interlocutors will influence the type or refinement of the terminology used to describe sounds of flue organ pipes.

As an example, we reproduced below an extract of a formal exam given by the Organ-builder to his apprentices.

Listen to the music examples and write which type of stop is being used amongst the following stops: Trumpet, Vox Humana, Flute, or Principal:
a.
b.
c.
d.

How would you describe the sound of a flue organ pipe having a larger diameter in comparison with a flue pipe having a smaller diameter, yet the other factors are the same?

a. More mellow, fundamental sound.
b. Sharper, brighter sound.
c. Same timbre, but louder.

Though basic, these two questions give a very good example of two types of verbalizations. Question 1 is concerned with the description of a full stop. The terms “Trumpet,” “Vox Humana,” “Flute” and “Principal” refer to standardized dimensions and scaling of flue organ pipes. Except for “Principal,” all the terms refer to a previous sound experience from a different musical instrument. This type of exercise is an important skill in organ-building design: being able to map a perceived sound to a particular stop. In Question 2, the context of listening changes. Now, apprentices are asked to use their “internal communication channel.” From their past experiences only (their long-term memory), they should be able to match a given variation of a physical parameter to a perceptual parameter to be chosen among a reduced set. Compared with Question 1, the type of verbal descriptors also changed. “Sharper,” “brighter,” “more mellow” concern only a part of the sound: its timbre. Timbre should be distinguished from “loudness” (as suggested); we will come back to this point later on (see Section 2.c). The use of the comparative form of these adjectives emphasizes the need for a relative judgment. Finally, we must point out that in the case of the task delimited by Question 2, the matching between words and relative changes in sound is strongly dependent on the previous experience that an apprentice might have of a flue pipe. He must know not only the exact meaning of the words “sharp” or “bright” applied to the case of a flue organ pipe sound, but also the physical system itself (on which he can modify parameters and listen to their influence step by step).

Table 8. Listening and communication contexts. The listening abilities of an *organ-builder* (or voicer) develop themselves under different kind of situations. This table lists different listening situations that might occur under *working conditions*. For each “sound object” presented in a certain context, there might exist potential interlocutors. We indicated only the most likely ones. Situations reported and used in this thesis are outlined (marked in gray).

Context – situation	**sound object** ***listening type***	**type of interlocutor** **Communication channel**	**comments**
Voicing in workshop	sound of a single pipe	colleague (organ-builder)	
		observer (“neutral”)	
		student (apprentice)	
Voicing in church	sound of a single pipe	assistant	In this case, the sounds of pipes are always heard in relation to the room-effects.
	relative balance of the sound of all pipes of a register		
	sound of a single register		
	relative balance of the sound of all registers of the organ		
Organ demonstration	sound quality of registers	user (musician)	
	sound quality of each pipe		
Concert	interpretation	-	
	sound quality of registers	-	
	sound quality of each pipe	-	
Recording	all possible listening objects	all possible	
	quality of recording	recording engineer	
Conference	recordings and/or just talks	colleagues, users	
Listening tests	relative sound quality of recordings	test leader	
		internal	
Imagination	all previous listening activities	internal	our *long-term memory* is activated

In the preceding example, there is not much possibility for confusion regarding the correspondence between verbalizations and actual sounds. However, cases where no such clear correspondences can easily be established are abundant. For example, the Voicer often told us about the difficulty he had several times in understanding what musicians would naturally refer to as a “sweet” sound: *What is* *a sweet sound?* Such difficulties might arise when no proper listening context and/or communication protocol is established.

A former voicing assistant Allvar (2000) employed in the North German Baroque Organ project, provided us with some insights concerning the learning procedure of such a communication protocol. In this case, it is absolutely needed. A voicing assistant, placed at a precise spot in the church, must provide the voicer (situated in the organ chest) with a verbal description of the quality of sounds modified by the room acoustics. The process of adjusting each other's perception is long and difficult, and suffers from the low intelligibility of speech communication in a large room. However, as reported by the Voicer, a satisfactory degree of agreement can generally be obtained.

Communication of sound-quality impressions might be ambiguous, but it is necessary in a number of situations. Before trying to improve it, one must first understand the basis on which it is built. Above all, it involves listening. This is not in itself a simple activity, and the following section is intended to explain some of its aspects.

II.2 - How do we listen to musical sounds and to their quality?

II.2.a - Functions and modes of listening

We all know that a musical piece can provide us with an infinite number of possible ways to listen to it. For example, we can focus on the quality of particular instruments or of the orchestral blending, the style of the music, the interpretation, the evoked emotions, etc. And so, while listening carefully (or intently) we can focus on completely different aspects than sound itself. This can be a rather annoying problem in a communication framework. Consequently, these functions and modes of listening must be defined and set, before a proper code of communication can be established.

Pierre Schaeffer (1966 , pg. 116), the father of “musique concrète”, was probably the first to propose a taxonomy of musical objects. His approach of sound as an object (a *sonic objec*) in itself (voluntarily separated from its source), extracted thanks to what he referred to as “une écoute réduite” (a focused listening), certainly contributed to significant development of electroacoustic music (Bel, 1992). Throughout all his “traité des objets sonores,” he constantly refers to four sectors of listening, which are outlined in the table below.

Table 9. The functions of listening—translated from Pierre Schaeffer (1966).

4. **to comprehend** “*comprendre*” *for me*: signs *in front of me*: values (meaning-language) Emergence of a content of sound. *Reference* and *comparison* with extra-sonic notions	1. **to listen** “*écouter*” *for me*: indices, clues *in front of me*: external events (agent-instrument) Sound *emission*	1 and 4: objective (causes)
3. **to understand** “*entendre*” *for me*: qualified perceptions *in front of me*: sonic objects *Selection* of certain aspects of sound	2. **to hear** *“ouïr”* *for me*: raw perception, sketch of the object *in front of me*: raw sound object Sound *reception*	2 and 3: subjective (effects)
3 and 4: abstract	1 and 2: concrete	

In Table 9, Schaeffer organizes four listening functions, each of which can be distinguished in terms of its input (*in front of me)* and output (*for me*) and can also be classified according to its level of psycho/physical treatment (concrete vs. abstract) and type of socio/individual coding (objective vs. subjective).

This thesis deals with the perception of sound as a component of a musical instrument and thus is definitely oriented toward a detailed analysis of sound qualities at the level of a single pipe. In the terms Table 9 sets forth, we are thus mainly concerned with sectors 1 and 3.

All these functions are presented at a "neutral position"; nothing, until now, has been said about their dynamics and possible triggering. Most of the time the trigger is of reflex nature—auditory perception is primarily an alarm system constantly "on" (you cannot naturally shut your ears). Before proceeding with the mechanisms of listening, we wish to introduce concepts of modes of listening. In Figure 7, we sketch a possible model of the process of different listening modes. This model is built upon the hypothesis that our auditory perception is influenced by the status of a source-identification process (see next section, II.2.b - p. 26). If listeners are certain of the source, they can focus on a higher-level processing of the sensory qualities of the sound or simply not process the sound because they have already decided that it is an "irrelevant sound" not worthy of further attention.

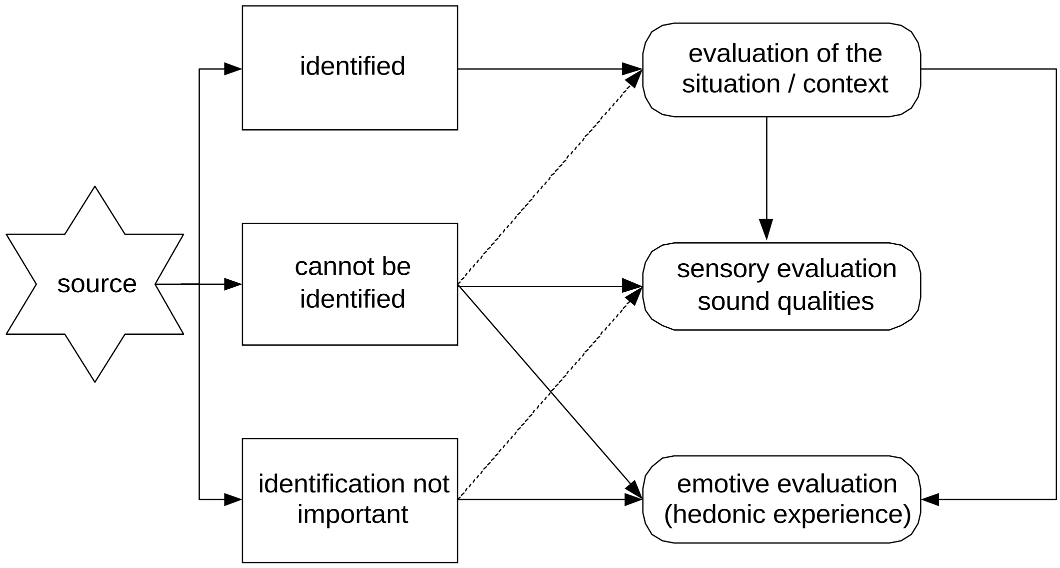


Figure 7. Modes of listening. Different attentional attitudes of the listener (reflex or conscious) toward the identification of the source of sounds lead to different kinds of evaluation. Adapted from Västfjäll (2000).

According to whether the source is (a) identified, (b) not identified or (c) not important, evaluations will vary in the following ways:

a) If the source is identified, people can evaluate the situation easily, which then enables a *high, detailed* level of sensory evaluation and possibly a hedonic evaluation. This is the listening mode, in which an expert (such as a voicer) will always work, and it thus constitutes the main listening mode of our study.

b) When the source is not identified, people cannot engage in high-level processing, since they are concerned with a very broad category of sounds. A negative hedonic evaluation (unpleasant) follows upon a *low* level of sensory evaluation.

c) This is the case of probably the most common mode of *musical listening* (Gaver, 1993). In this mode, the listener is not interested by the source (for some reason) and will probably enjoy the music (positive hedonic evaluation) more than evaluate sensory qualities.

Note that we make a clear distinction between hedonic (driven by emotional aspects (Västfjäll 2001a; 1999)) and sensory qualities. But what exactly do we mean by sensory qualities? Can these a priori subjective qualities be measured? Can we link them to measurable objective physical variables of sounds? These are questions that have been addressed by psychoacousticians.

II.2.b - Psychology and acoustics

In order to establish and promote a proper paradigm of research aimed at matching verbal descriptors and physical properties of sounds, one must make hypotheses on how information (carried by a sound) is processed by the auditory system (the ear) and further in by the brain. This in turn involves three “mother” disciplines, namely physics, physiology and psychology. Later, we will approach the acoustics (physics) of a flue organ pipe combined with a description of signal processing (see Section III.2 -). The present section, however, is devoted to the psychological side of such a project. We thus skip considerations concerning physiological issues and assume the reader to have a basic level of knowledge (see e.g. (Hartmann, 1997; Helmholtz, 1885; Leipp, 1989; Pickles, 1988; Roederer, 1995)).

II.2.b.i - Psychophysics vs. cognitive psychology

> Cognitive psychology is concerned with *information processing*, and includes a variety of processes such as attention, perception, learning, and memory. It is also concerned with the structures and representations involved in cognition. The greatest difference between the approach adopted by cognitive psychologists and by the Behaviorists is that cognitive psychologists are interested in *identifying in detail what happens between stimulus and response*.
>
> from Keple (1997)

> Psychophysics is commonly defined as the quantitative branch of the study of perception, examining the relations between observed stimuli and responses and the reasons for those relations. This is, however, a very narrow view of the influence it has had on much of psychology. Since its inception, psychophysics has been based on the assumption that the human perceptual system is a measuring instrument yielding results (experiences, judgments and responses) that may be systematically analyzed. Because of its long history (over 100 years), its experimental methods, data analyses, and models of underlying perceptual and cognitive processes have reached a high level of refinement. For this reason, many techniques originally developed in psychophysics have been used to unravel problems in learning, memory, attitude measurement, and social psychology. In addition, scaling and measurement theories have adapted these methods and models to analyze decision making in contexts entirely divorced from perception.
>
> from Baird (1978, p. 1)

As one can see from the definition of psychophysics given by Baird (1978), the research fields of psychophysics and cognitive psychology might overlap, but various researchers have made a distinction between the adepts of cognitive psychology and the defenders of psychophysics with regard to audition (see e.g.Maffiolo, 1999, pg. 51; Sarris, 1999 , pg. 41). While both fields regard human hearing as their object of study, one is mainly concerned with the responses of the auditory system to basic stimuli (psychophysics) while the other one integrates higher information processing in order to reveal the mechanisms involved in the perception of complex (everyday) stimuli.

Psychoacoustics is generally considered to be a sub-field of psychophysics. Psychoacousticians study the perceptual *attributes* of audio-signals. Studies made in this field mainly use simple and controlled stimuli (e.g. sinusoids or band-limited *stochastic* signals). One of the main hypotheses of this kind of study says that we can infer the perceptual reaction of an individual from the knowledge of the signal characteristics of a sound and the knowledge of the auditory system (studied in details by physiological acoustics). This paradigm can be classified as a bottom-up approach (see Figure 8), where “bottom” refers to basic auditory processes (e.g. cochlear mechanisms) and “up” is concerned with high-level cognitive processes (such as identification, recognition and lexicalization). According to this hypothesis, the study of very well controlled stimuli (e.g. synthesized) should provide a description of the essential mechanisms of audition. Well-known studies have been made under this scheme, concerning the perceptual attributes *loudness* (Zwicker *et al.*, 1999) and *sharpness* (Von Aures, 1985; Von Bismarck, 1974a).

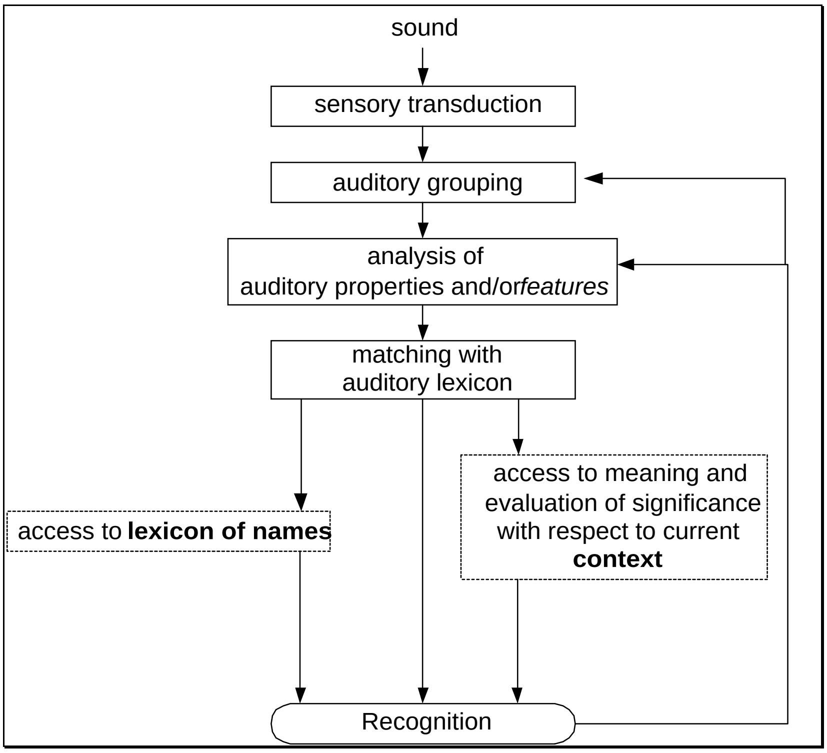


Figure 8. Schematic diagram of auditory perception modeled by a *bottom-up* organization, from McAdams (1993b).

In our case, input stimuli are “natural” sounds (not synthesized). And we expect outputs (responses of participants in listening tests) to provide us with a description of the *timbral* differences of stimuli. We will come back later to the specific qualities of musical sounds and *timbre* in particular. In order to close this discussion of the various streams of auditory perception, we have to state that the study of natural sounds has increased attention to a specific field of psychology, namely, ecological psychology, in line with cognitive psychology.

II.2.b.ii - Ecological psychology

Ecological psychology is strongly linked to the concepts of adaptation and categorization (Rosch, 1978). Originally developed for the study of visual perception (Gibson, 1986), it has further been extended to other perceptual modes—auditory (Ballas, 1993; Guyot, 1996; Maffiolo, 1999) or olfactory (David, 1997). In this context, the auditory perception is dependent on our representation of the world to which we must adapt. Moreover, this representation is based on a hierarchical organization of categories more than on dimensional or featural *attributes*. In this framework, the world is made of objects that share or do not share different kinds of properties. Objects are organized in hierarchical categories. These objects may produce sounds, and so categories of sounds are always emanating from their sources. These categories or groups are based on the concept of invariants. Ecological psychologists distinguish between *structural* invariants and *transformational* invariants (McAdams, 1993). *Structural* invariants specify the nature of the source (its structure, for instance, is used to distinguish a violin from a piano), while *transformational* invariants discriminate between the different types of transformation of the source (for example, plucked or bowed string). The main point of this alternative approach is that an input stimulus does not interact with the auditory system without a prior expectation of its properties: perception is always dependent on *a priori* knowledge. This point is outlined in Figure 9, where the concepts of **context** and **memory** now occupy the starting point of a flow chart otherwise fairly similar to the one presented earlier (see Figure 8).

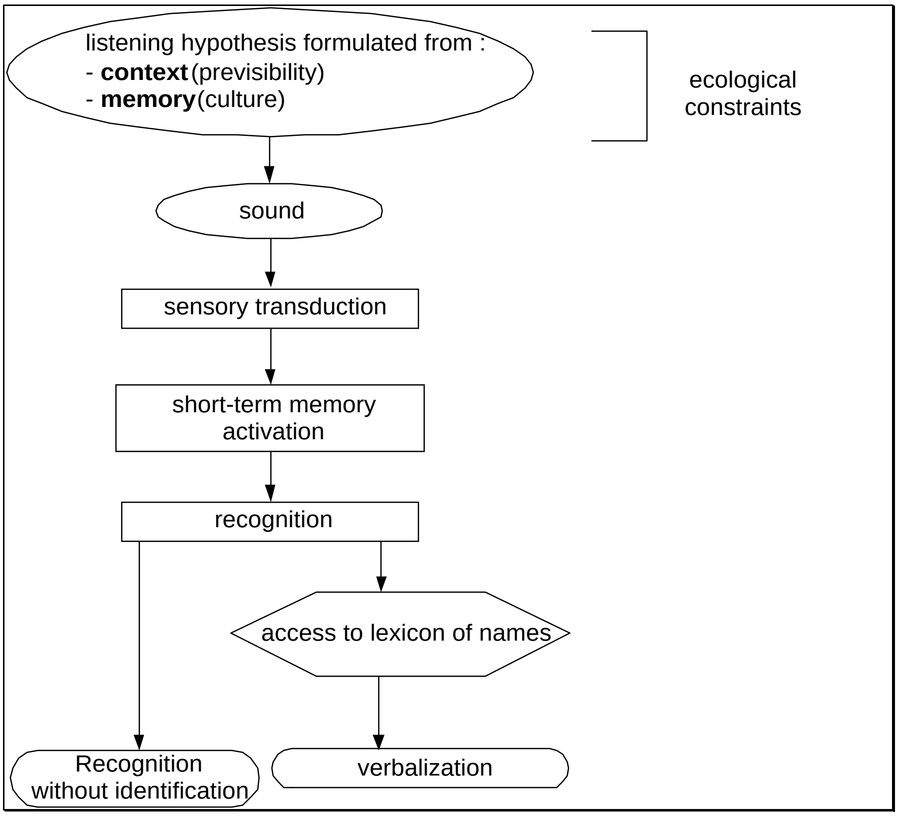


Figure 9. Schematic representation of auditory perception with an “ecological” point of view. *Top-down* organization. from Gaillard (2000)

Recognition means that what is currently being heard corresponds in some way to something that has already been heard in the past (McAdams, 1993). According to McAdams there is basically just a difference of naming between recognition and identification (a narrower recognition process using a possibly hierarchically organized lexicon).

The previous classification of invariants specifies a certain scale factor at which sources are generally examined, reflecting a typical listener in a "normal" situation, which might be affected by the quality of the acoustical surroundings (soundscape). Whether such a paradigm is applicable for the musical mode of listening could be discussed. In our case, we have already stated that the identification of the source is not relevant. The timbre of all our sound objects is produced by the same source, which is known. If a process of categorization applies to this kind of sounds, it will then deal with more basic components of sounds, which make up timbre. This is why we developed a method (see Paper IV) where both categorizations and comparisons could be done when assessing relative timbral modifications.

II.2.c - Timbre and musical sound quality

Up to now, we have used the term timbre with an implicit reference to its "intuitive definition." Unfortunately, there seems to be nothing like an explicit definition of timbre. The concept of timbre, since the pioneering work of von Helmholtz (1885), is generally explained by a subtractive approach: timbre is what is left of a sound when other characteristics such as pitch, *loudness*, duration are "extracted." Bregman (1990, pg. 92) gives a remarkable critique of this last assertion and a rather extensive explanation of the problem associated with the definition of timbre. Again, it is essential not to neglect the importance of the modes and functions of listening. From a cognitive-ecological point of view, timbre is what helps us to distinguish two different sources (e.g. musical instruments or voices). Castellengo (1994) as quoted by (Gaillard, 2000, pg. 44) recognizes thus two main components of timbre:

The sound emitted by a musical instrument carries the identity of its source through its production. This sound can then be represented by a number of **structural characteristics** related to *pitch* , intensity, etc.

A sound can be characterized by its similarities to or differences from another sound, but also by some characteristics which make it unique. These principles of comparison do not lie in the sound itself but in the way sounds are heard.

We can in fact distinguish two functions of timbre. Timbre can either represent a **signature** of a particular source or a set of **characteristics** (similar, different, or unique) which allow a comparison between two sounds (**sonic objec**ts) "disconnected" from their source. In our case, the physical source of sound under study is known and fixed, and hence only the second component of timbre comes into play here. We are interested in finding ways of describing how timbre is affected by voicing modifications of the same pipe. Each of these modifications produces a new sonic object, which can be **understood** (see sector 3 of Table 9) thanks to an "écoute réduite" (close listening) introduced by Schaeffer (1966), also reported by Handel (1991, pg. 172). This close listening will focus on certain characteristics of sound. But which characteristics are these? The principles of comparisons of sonic objects rely on such characteristics, but they too are subjective. Can we then find some objective correlates? This is, of course, one of the major issues of auditory research. While some success has been obtained regarding pitch and loudness, the acoustic correlates of timbre still resist analysis. Nevertheless, several researchers have established the **multidimensional** aspect of timbre. In his study, Grey (1977) found three main dimensions responsible for the discrimination of synthesized sounds imitating classical musical instruments (including some hybrid sounds obtained by morphing). These dimensions were found by multidimensional scaling of dissimilarity judgments. McAdams (1999a, 1999b), following the early work of Grey, recently proposed using acoustic correlates such as the log attack time, the spectral *centroid* (related to *brightness*, see Chapter III.2.c - , p.60), spectral smoothness and spectral flux. McAdams also distinguishes two types of timbral characteristics, namely feature**s** and **dimensions**. Dimensions can vary across a continuum scale, while features cannot; they might delimit categories or act as a sort of distinctive anchor. Pollard (1988) showed that not only line spectral components (old definition of timbre), but also starting *transient* (onset), balance between *stochastic* and periodic parts and inharmonic tonal components should be taken into account (for a more detailed explanation, see Section III.2 -).

We must remark that the approaches of Grey and Pollard can be distinguished from each other in that the first tries to exhibit the fundamental dimensions of timbre, while the other lists all its possible components. This is probably reflected in the type of listening activities to which they implicitly refer (normal for the former, detailed for the latter). In any event, in both cases we have to acknowledge the fact that any listener is able to focus only on a limited amount of characteristics. Listening is limited by attention and conscious knowledge of timbre characteristics (learned by experience in the case of a voicer, for example).

We also have to note that there is some evidence that pitch and loudness are also responsible for a certain sound quality and hence influence timbre (H. Fletcher, 1934, pg. 68). Moreover, both pitch and loudness are subjective quantities that cannot be accurately predicted. Nevertheless, we can consider that for very similar *sonic objects*, extremely small variations in pitch and loudness do not interfere significantly with timbre. This has always been the case for the stimuli presented in our listening tests, but we felt the need to implement a loudness *equalization* interface that could be included in a test session (see htmlDoc: LISE and Appendix V).

Finally, in the case of our study of flue pipe sounds, we will postulate that timbre is both a featural and multidimensional quantity reflected in temporal, spectral and spectro-temporal *micro*properties of sonic objects. Hence timbre is the quantity which enables us to distinguish:

- the same note played on different stops,
- different steps of voicing of the same pipe when loudness and pitch are kept constant.

We will focus primarily on the second point.

II.2.d - Summary

We believe that even though both theories (psychophsyical/cognitive) can be defended, they represent the two sides of the same coin. On the one hand, auditory perception has a “deterministic” part that is directly related to the evolution of our auditory “sensory” system (outer to inner ear, i.e. tympani, cochlea and nerve cells). On the other hand, our auditory “perceptual” system has followed the evolution of the human species. This system had to adapt to the requirements of the complex life environment on earth. If we remember that our auditory system is primarily an alert system, we can be sure that it has evolved according to the constraints of an ecological system and the demands of certain cognitive tasks.

In subsequent chapters we will make the following hypotheses concerning our auditory system:

- Auditory evaluation is contextual (it might change depending on a specific context or task).
- Our auditory evaluation of stimuli is relative. The auditory system is a “comparative” system.
- Internal representation of auditory stimuli (memory) is based on hierarchical categories that might overlap.
- A prototype or *attribute* determines each category. Attributes are either *dimensional* (continuous) or *featural* (present or absent).

Of course we should not forget that in the present study, only musical sounds are of interest, which means not only that sounds are complex but that they also should convey an artistic value. This artistic dimension is at the same time essential and transparent (difficult to grasp). Moreover, the definition that consists in saying that a musical sound comes from a musical instrument is no longer valid. New technologies and especially recording technologies have opened new frontiers to musicians. Musical expression through sounds is less and less bounded and can at the limit work without physical sound sources (we might think for example of the multiplicity of sound types used in contemporary compositions). In order to explain what a musical sound is, one has to move toward philosophical concepts. That is what one of the most influential researchers and composers dealing with electro-acoustic music did. Pierre Schaeffer, the father of the “musique concrète” in his “traité des objets sonores” (1966) discussed in detail the conceptual issues concerning different ways of listening to music. He was probably one of the first to introduce the concept of “sonic objects.” This concept presupposes the existence of a special way of listening, some kind of “focused hearing” (“écoute réduite”) which is interested only in sound in itself. The source of sound, so crucial in an ecological perspective, is consciously reduced to its lowest possible influence, when identifiable.

Finally, we try to propose flue organ pipe sounds and ways of listening to these sounds so that the attention of the listener does **not** focus on:

- esthetic judgment (beautiful-ugly)
- hedonic (pleasant-unpleasant) judgment
- emotive or affective judgment (e.g. anger, anxiety, elation (see Västfjäll *et al.*, 1999))
- musical adequacy
- adequacy with the full instrument: we are interested in the analysis of *micro*-temporal properties of the sounds of single flue organ pipes (not even seen as a component of a *stop* or an organ)

but on:

- relative timbral changes between a set of sounds (minimum a pair)
- the temporal parts of sound (transient, stationary)
- the spectral parts of sound (stochastic, periodic)
- information on the source (in the case of a voicer, through voicing parameters).

In order to get a description of this type of listening we need to have a language adequate for these small timbral differences. This wish was already expressed back in 1934 by Harvey Fletcher (1934), and it still resonates today. In the following sections we will discuss the use of verbal descriptors of sound and introduce existing and applicable methods of listening tests.

II.3 - Verbal description of sound quality

In the previous chapter on voicing, we introduced a number of verbal descriptors that were used both by organ-builders and scientists. We wish in this section to examine the lexical nature of such descriptors and their relevance for timbre perception. We will first review similar attempts made in acoustics and other fields of perception and then move to the specificities of musical sounds.

II.3.a - Use of verbal descriptors in acoustics

Several domains of research in acoustics are using verbal descriptors of sounds. We can roughly divide research studies into those using already established verbal lexicons produced by specialists and those aiming at building a lexicon from the bottom up. These two approaches can be criticized on the grounds that the interaction between the layers of the sensory evaluation of sounds and the semantics of associated verbal descriptors is rather unknown and might lead in many cases to a weak coupling between the use of a certain descriptor and a relevant perceptual *attribute* of a sound (and even more, to acoustics correlates). Nevertheless, it is obvious that in the majority of cases, research on verbal descriptors is often a remarkable point of departure for an exploratory study on the perceptual dimensions and *features* of sound.

Several studies, for example, have been concerned with sound reproducing systems (Gabrielsson *et al.*, 1985), sound quality of cars (Hempel *et al.*, Muckel *et al.*, 1999), sound quality of vacuum cleaners (Guyot, 1996), urban soundscapes (Maffiolo, 1999) or everyday (or natural) sounds (Björk, 1985). These studies are mainly focused on the *sound quality* of industrial products or environmental noise and generally do not get input from experts.

By contrast, Solomon (1958) did pioneering work on sonar noises and asked different types of participants from novice to experts (submarines operators) to verbally qualify these types of sounds. This work is particularly interesting as, just as in our case, Solomon tries to establish a lexicon based on a set of verbal descriptors collected from various sources (Solomon, 1958, pg. 422), namely:

- college students (free verbalizations)
- a list of scales used by Osgood and Suci (Osgood *et al.*, 1955)
- rational analysis of the sensory inputs of the human organism
- a list used in ship-classification teaching methods (experts).

He then performed a listening test based on Semantic Differential (see section 4) of 50 adjectives in order to detect attributes emerging from a group of descriptors. He found seven main relevant dimensions, which are reproduced in Table 10, below.

Table 10. Descriptors of the first seven factors found in the study by Solomon(1958), followed by unidentified descriptors.

magnitude	**aesthetic-evaluative**	**clarity**	**security**	**relaxation**	**familiarity**	**mood**
heavy-light large-small rumbling-whining wide-narrow low-high	beautiful-ugly pleasant-unpleasant pleasing-annoying smooth-rough	clear-hazy definite-uncertain even-uneven concentrate-diffuse obvious-subtle	mild-intense gentle-violent calming-exciting safe-dangerous simple-complex	relaxed-tense loose-tight soft-hard soft-loud gentle-violent mild-intense	definite-uncertain familiar-strange wet-dry active-passive steady-fluttering	colorful-colorless rich-thin happy-sad deliberate-careless full-empty

unidentified	dull-sharp, gliding-scraping, good-bad, green-red, hot-cold, labored-easy, masculine-feminine, powerful-weak, pushing-pulling, repeated-varied, resting-busy, rounded-angular, rugged-delicate, slow-fast, solid-hollow

The nature of these seven attributes is rather complex; they can be related to emotional, social or esthetic values. Only the attributes “magnitude,” “clarity” and ”relaxation” are possibly close to perceptual attributes like loudness, *sharpness*, noisiness or roughness. Nevertheless, some descriptors such as smooth-rough, simple-complex, mild-intense, steady-fluttering, rich-thin appear under emotional or esthetic attributes, whereas they can just as well serve for direct timbral description (see paper 0). As Solomon (1958 pg. 424) puts it, the interpretation of these “dimensions” is never completely free from the subjectivity of the factor analyst, and it is important not to forget the **frame of reference** of the listener (Solomon, 1958, pg. 422). It is also worth noting that the seven attributes found are obviously not directly correlated with any physical variables of sound. In any case, while each attribute tries to grasp a perceptual dimension of a set of sounds within a certain frame of reference, it is clear that the underlying descriptors very much reinforce the understanding of the meaning of the attributes. In other words, it seems unlikely to be able to find meaningful enough attributes which can summarize a list of descriptors: *a lexicon should be constituted by a set of attributes complemented by a list of descriptors.* Before drawing any further conclusions, then, we would like to point out that for such kinds of work the use of verbal descriptors should also be accompanied by acoustic analyses and, whenever possible, audio demonstrations.

II.3.b - Verbal descriptions of other perceptual modalities and Synaesthesia

A great number of previously mentioned descriptors for auditory stimuli are clearly borrowed from other modalities of perception like vision, touch or taste. Moreover, several studies have pointed out the existence of interaction between auditory perception and other senses like vision (Larsson *et al.*, 2000; McGurk *et al.*, 1976). Hence, some researchers in sound quality have been interested in transferring some knowledge (especially methods) from other areas, such as the food industry. Bech (1999), for example, quotes the successful work of Noble (1987), who constructed a remarkable wine aroma standardized lexicon, while Bech has addressed the problem of spatial sound characteristics. Conversely, researchers coming from the food industry have taken part in sound quality conferences (see e.g.Civille, 1998). It generally turns out that the vocabulary concerned with taste is more proficient and accurate than verbal descriptors of auditory perception. Thus Noble (1987) managed to present about one hundred non-ambiguous aroma descriptors for wine tasting. This lexicon was hierarchically organized from basic components (close to essential aroma, e.g. sherry, mushroom, oak, tea, fig...) to broad attributes of taste (floral, woody, fruity...).

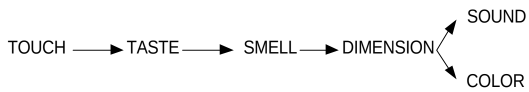
In a study concerning the link between sensory representations and personality characteristics , David(1997) explored the contrast between two senses, namely olfaction and audition. This study is strongly grounded in linguistic and cognitive analyses of free verbalizations made by subjects. Olfaction seems to be less lexicalized (in French) than hearing, but in both case she found that a specific vocabulary is rather poorly employed. At first glance, this finding contradicts the results of Noble concerning taste. However, we have to point out the importance of the frame of reference (or context), which differs greatly between these two works. In the case of Noble, the frame of reference is narrow (restricted to wine tasting), whereas in the case of David, the frame of reference is open. This supports our belief that *only a focused frame of reference/context can provide us with specific and precise descriptions of auditory perception*. This point is supported by research in other domains of perception like visual perception. A study of lexicons applied to colors (Dubois *et al.*, 1999) shows, for example, that the denomination of colors is linked to a specific task (such as art or decoration). This finding supports the hypothesis that a subject cannot be reduced to an information process unit, in that he is organizing his own cognitive (e.g. categorization) and linguistic landscape for a specific context.

Having considered these works on senses other than audition, it is appropriate to discuss the special case of words that can describe several sense modalities. This phenomenon is one instance of what is known as synaesthesia, which has been studied by linguists (the other instance of synaesthesia, concerning the confusion of different modalities at a perceptual level, is assessed by psychologists). Synesthetic words are very much present in the naming of auditory impressions and thus require particular treatment in this thesis.

Abelin (1988) distinguishes two types of “linguistic” synaesthesia:

- real synaesthesia:
  - from one sense modality to another (e.g. sight→hearing)
- partial synaesthesia, which concerns the extension of meaning
  - from one sense modality to something else (e.g. sight→mental qualities)
  - the opposite (e.g. dimensions & shapes→hearing).

Note that while Abelin’s work concerned the Swedish language, all terms and classifications were translated into English in the article, supporting the hypothesis that this work can be extended to English. We might argue here that terms expressing “dimensions and shapes” (often used in verbal description of sounds) could in fact be interpreted as coming from a sort of “meta-sense” consisting of a mixture of several senses plus a mental activity (in this case visual + tactile + geometry). A diachronic study (J. Williams, 1976) revealed that synesthetic words from sensory fields develop according to the following scheme:



While stating that it is difficult to know whether taste or smell comes first in time, Williams claims that this scheme is stable and that it might even apply to non-Indo-European languages. Considering that there is clearly some cognitive process involved in the development of such a rule, this constitutes an interesting point in support of the hypothesis that words of different languages could describe the same characteristics of sensations. This is a rather important point for this thesis, as we have worked with a voicer whose mother tongue is Japanese and with test participants coming from different countries (France, Germany, the Netherlands, Sweden, and the U.S.). Though everyone was familiar with American English, we still have to make the hypothesis that they shared some common cognitive representations of sounds and that they could translate them in an appropriate manner. In this line, Carterette (1982pg. 768), comparing cross-cultural studies of music perception, notes that “meaning and mood depend more deeply on social and theoretical systems than in the case with timbre.”

Table 11 gives us an example of common translations of synesthetic terms. Nevertheless, several Swedish terms might correspond to a single English word; for example, “sharp” can be translated as both “vass” and “skarp.” This is probably evidence of the well-known fact that adjectives develop according to the needs of a people in a particular context. An inspection of Table 11 makes it clear that verbal description of the sense of hearing is highly synesthetic. Moreover, Abelin took these adjectives from a list of 280, of which only fourteen are exclusively used for the hearing modality, thus supporting the fact that the denomination of auditory perception with specific terms is very limited compared with the other modalities (at least in the case of the Swedish language).

Table 11. Real and partial synesthetic terms applicable to the hearing sensory mode, adapted from Abelin(1988, pg. 23-30)

TOUCH→HEARING	mörk (dark) ljus (light) briljant (brilliant) dunkel (dusky)
TOUCH (surface structure)→HEARING	**vass (sharp)** **skarp (sharp)** skärande (cutting) **skrovlig (rough)** **sträv (rough)** len (smooth) torr (dry) jämn (even)
TOUCH (temperature)→HEARING	varm (warm) kall (cold) sval (cool) het (hot)
TOUCH (consistency)→HEARING	mjuk (soft) hård (hard) massiv (massive) grötig (muddled)
DIMENSION→HEARING	hög (high) låg (low) djup (deep) liten (small)
SHAPE→HEARING	tjock (thick) tunn (thin) grov (stout) platt (flat) fyllig (full) sprucken (cracked)
BODILY (physical) CHARACTER→HEARING	stark (strong) svag (weak) lätt (light) tung (heavy)

Abelin also notes that several adjectives can be regarded as onomatopoetic or sound symbolic. We will reserve a discussion of these very special descriptors for the following section, which is concerned with musical sounds.

II.3.c - Verbal description of musical sounds

II.3.c.i - A few approaches

"...It is commonly known that the timbre of tones coming from two different violins may be greatly different and we have no adequate language to express this difference. Such a language is greatly needed." Despite this early proposition of Fletcher (1934, p.67), few researchers have addressed the problem of verbal description of musical timbre and more generally of musical sounds. On a larger scale, "there are few systematic studies of descriptions or verbal attributes of music (Carterette *et al.*, 1982)."

Probably the first significant and systematic attempt of deciphering the perception of musical timbre through verbal description was made by von Bismarck (1974b). He used the steady parts of 35 sounds chosen to represent "the most prominent characteristics of voiced and unvoiced speech sounds as well as some musical sounds" and explicitly asked participants not to concentrate on the **identification** of the source nor on the **physical** components of the sound. Thirty scale labels were collected from previous studies, including the one by Solomon (1958), and thus **elicited** or **free** verbalizations were not allowed. These thirty scales were reduced to four main factors by a principal component analysis of the correlation matrix. Sixteen out of thirty labels were found to be "factorially pure" and are reproduced in the table below.

Table 12. Main attributes and their corresponding descriptors found by von Bismarck (1974). F1 denotes Factor 1.

F1 (44% of variance)	F2 (26% of variance)	F3 (9% of variance)	F4 (2% of variance)
dull-sharp	**compact-scattered**	**full-empty**	**colorful-colorless**
sharp, hard, loud, angular, tense, obtrusive, unpleasant, bright, high	compact, boring, narrow, closed, dead	full	colorless

Compared with the study by Solomon, the description of timbre is much more detailed. Moreover, von Bismarck provided signal analyses of the set of sounds. Nevertheless, we still feel that audio documentation is needed to fully grasp this type of work. These results led von Bismarck to further develop a *sharpness* psychophysical scale (Von Bismarck, 1974a).

Following a rather different approach, Samoylenko et al. (1996) developed a method of verbal protocol analysis suited to the analysis of **free verbal description** of musical timbre. Noting that for the techniques related to the use of verbal attribute scaling "the relations between classes of verbal attributes and acoustics properties are often quite weak," they focused mainly on a systematic analysis of the meanings and linguistic nature of the elicited labels with a categorization approach (Rosch, 1978). Their corpus of stimuli was constituted by a rather broad selection of synthesized sounds of various classical musical instruments, in line with the choices of Grey(1977). They found that subjects were using different approaches according to whether they were to rate similarity or differences between sounds. According to them (Samoylenko *et al.*, 1996, pg.

273), judgments of similarity mainly involve the comparison of specific feature***s*** of timbre as opposed to **holistic** entities (which concern the sound as a whole).

An alternative approach has recently been proposed by Faure et al. (2000). Motivated by the growing need for verbal descriptors for the creation of musical sound databases, they studied free verbalizations of twelve sounds equalized in pitch, loudness and duration. It was thus possible to exhibit a **verbal portrait** for each one of these sounds. Their *tour de force* relies on the fact that they manage to check that test participants were able to recognize auditorally a sound described only by its verbal portrait. This was probably the first time researchers were able to establish the existence of such a cognitive loop. This actually might also suggest that verbalizations could influence our perception of sounds (or at least our conscious perception of sounds). This was pointed out by Samoylenko et al*.*(1996), quoting Bower et al.(1973), who demonstrated the important role of verbal description in the process of memory and recognition. They found that subjects recognized sounds by reconstructing related **verbal portraits** (often of a visual image of their source). More information on cross-modal source monitoring can also be found in (Henkel *et al.*, 2000). An interesting illustration of this phenomenon could be found in the study of the fairly recent development of digital synthesizers. It is quite puzzling to notice that this technology did not bring so very many new sounds despite its early promises of profusion (all type of sounds are potentially synthesized by a computer). Instead, most sounds available on synthesizers are aiming to imitate existing instruments. This could be the result of our difficulties in imagining new sounds, remembering new sounds (which are not attached to a physical object any more) and creating new “interesting” sounds (synthesis methods may have to include “perceptual” or at least “physical” parameters (Cadoz *et al.*, 1981)). We explicitly refer to digital synthesizers, but it is certainly valuable to note that these machines are often built on the same principles as organs. The organ can in fact be regarded as the first system to perform complex *additive synthesis* (Comerford, 1993; Risset *et al.*, 1969; 1982). In the musical teaching domain, one can quote the influence that a teacher can have on his students concerning the mastering of their instrument. When the production and quality of sound is of crucial importance (e.g. strings or flutes), the student has to learn how to match physical sensations with his perception of the sound quality. Talking about violin teaching, Guettler (1999) affirms for example that “it is only when the term *acceleration* has been connected to something perceivable and to a particular *sound quality* that it will serve as useful information for a less-technically-inclined string player.”

All these approaches which use verbal description at their starting point have been criticized (see Samoylenko *et al.* 1996 pp. 256-257) for a good review of pros and cons]. Nevertheless, there are many hints tending to prove that this might still be the only possible way to decipher timbre perception. We judged that making this type of digression was essential in order to understand the listening tests presented in paper II (p. 97) & III (p. 99) and somehow depict their background. Before returning to the main subject of this thesis, we will examine a field of research that shares some direct relations with the voicing of organ pipes.

II.3.c.ii - An interesting parallel with speech pathology

We will not, unfortunately, develop the parallel between the physics of flue organ pipes and the physics of the voice (and especially the singing voice (see e.g. Miller, 2000)) to any great degree, though some obvious links exist (just considering the fact that the organ generally accompanies singers). Closer to the interests of voicing, we find in speech pathology a discipline relating to both the physical and perceptual dimensions of the voice. Our primary interest in deliberately choosing to talk about this discipline is that there is a definite need for the construction of a verbal lexicon appropriate for the description of voice disorders (for a simple review see Ramig, 1994). Speech pathologists are working with an interdisciplinary field, gathering input from speech clinicians, otolaryngologists (specialists of the oto-larynx), psychologists, neurologists, oncologists (cancer specialists), pediatricians, vocal coaches and singing teachers. Their main goal is concerned with etiology, in other words, to relate the effects (voice perceptual quality) with the causes (acoustic and physiological measures of the voice characteristics). These relations need to be described somehow and can be accessed through a lexicon. An example of a collection of descriptors used in studies by Ramig(1994) and Hammarberg et al.(1980) has been compiled and listed in the table below.

Table 13. Examples of verbal descriptors used in voice disorders evaluation, from Ramig(1994) and Hammarberg et al.(1980).

bitonality, breaks, breathy, chest register, coarse, coughing, creaky, vocal fry, diplophonia, flutter, grating, gravelly, guttural, throaty, hard glottal attacks, hard onset, harsh, head (falsetto) register, hissing / wheezing, hoarse, husky, hyperfunctional, lack of timbre, middle register, moments of aphonia / aphonia, monotonous, pitch, pressed, quality changes without voice breaks, raucous, relaxed, repressed / restrained, rough, strained, strangled phrase endings, throat clearing, tight, tremorous (flutter vibrato), unstable pitch, unstable quality, voice breaks / tendency towards voice break

Not all the descriptors listed in the previous table have defined acoustical correlates. Not so surprisingly, because of their common use, most words which come from everyday language suffer from this problem. Even if we all know what a breathy, coarse, harsh, hissing, rough, tight, etc. voice refers to, it is rather clear that these descriptors may reflect a complex span of multiple physiological parameters. Nevertheless, some of these physiological parameters can be identified, and Titze (1995), for example, proposes an alphabetic glossary of definitions of terms related to vocal fold physiology, some extracts being listed in the following table.

Table 14. A few examples of definitions of terms used in vocal fold physiology, adapted from Titze (1995)

bleat = flutter	phonation with amplitude or frequency modulations (or both) in the 8-12Hz range; also called bleat as the bleating of a lamb
breathy voice	containing the sound of breathing (expiration) during phonation; acoustically, breathy voice, like falsetto, has most of its energy in the fundamental, but a significant component of *noise* is present because of turbulence in or near the glottis...
creaky voice	a voice that sounds like a creaky door; acoustically, a complex pattern of subharmonics and modulations that reflect a complex pattern of vibration of the vocal fold.
pressed voice	phonation in which the vocal processes of the arytenoid cartilages are pressed together, resulting in a constricted glottis with relatively low airflow; the fundamental is weakened relative to overtones.
rough voice	an uneven, bumpy quality that appears to be unsteady in the short term, but stationary in the long term; acoustically, the wave-form is chaotic, with the modes of vibration lacking synchrony.
strained (tense) voice	a voice that appears effortful; visually, hyperfunction of the neck muscles is apparent; the entire larynx seems compressed.
twangy voice	a sharp, bright quality, as produced by a plucked string. Twang is often attributed to nasality, but it is probably more laryngeally based. It is often part of dialect or singing style.
whistle register	a register in which the sound appears as a whistle, high in pitch and flute-like in quality...

It is very interesting to note that some descriptors present in both tables are also used for the description of the timbre of flue organ pipes (see Papers II and III). In fact, we all have an idea of what voicing techniques are dealing with when we try to control small changes in our own voice.

II.3.c.iii - Different types of descriptors

From the first pages of this document, we have tried to gather a large number of verbal descriptors of sound coming from various sources, trying on the way to delimit the context of such verbalizations and their purpose. We have thus gathered some proofs that description of sounds by means of verbalizations is possible and at least needed in many cases. Nevertheless, we have until now implicitly assumed that some kind of isomorphism exists between verbal descriptors and some perceptual characteristics of the sound they describe. We wish here to explore these links in more detail. In the case of flue organ pipe sounds, we can have three types of characteristics to which verbalizations can apply:

- The “objective” source or physical characteristics (mouth geometry, size, scaling)
- The “objective” signal characteristics (temporal, spectral and spatial properties)
- The perceptual characteristics (as such or of the last two characteristics).

Objective refers here to the measurable values of the physical world. When listening to sounds only, verbalizations can of course apply only to the third point; we can in fact only guess (or try to identify) the modified characteristics of the source or the signal characteristics. This guessing will get better and better during the learning phase of an expert. This is why we generally get more verbalizations of the perceptual characteristics of source and signal from experts than from novices.

Verbal descriptors are words, most often adjectives but also possibly sentences. They can be of the following linguistic forms:

- specific, extended or synesthetic
- unipolar vs. bipolar
- sound symbolic, possibly onomatopoetic
- sonic analogies (metaphorical).

We have already covered aspects of **synaesthesia** (see Section II.3.b - , p. 34) and here we simply mention the fact that **specific** descriptors, for example, those coming from speech quality evaluation, can be **extended** to other domains of sound quality. An inspection of Table 15 shows that these three types of descriptors are used by the expert. When a descriptor does not have any antonym, it is said to be unipolar; by contrast, it is said to be bipolar when one or several antonyms can be displayed. Again, the expert uses these two forms of descriptors. This might seem to be a relatively unimportant distinction, but it actually has a rather large impact on the choices of scales in listening test experiments (covered in the next section).

The case of **sound symbolism** is rather special and thus requires particular treatment. A word is said to be sound symbolic whenever its sound structure is dependent on its meaning. We do not want to enter here into the debate concerning the importance of sound symbolism in languages and for this refer the reader to Abelin (1999) and Hinton et al. (1994) for examples. We will try, rather, to understand and study the presence of such type of words in our lexicon. If we adopt a recent model of a universal grammar depicted in Figure 10, the nature of a sound symbol is described by a coupling between the acoustic and logical forms. A sound symbol is actually also called a **phonosymbol** or a **phonestheme** (Abelin, 1999, pg.3). A limited case, known as *onomatopoeia*, is encountered when the logical form is totally absorbed by the acoustic form. An onomatopoetic word refers to its own sound.

We can recognize four words in Table 15 that are actually used by the expert as phonosymbols, namely “chiff,” “cough,” “hiss” and “kiss.” These phonesthemes turn out to be of an unusual form. The onomatopoetic part of these descriptors is embedded in the **first part of the word** (the two first *phonemes* that might need to be considered as *allophones*). Hence a pipe exhibiting a marked “chiff” will have a transient which sounds like “chi-”, /t i:/ in international alphabet notation (see and hear Wells *et al.*, 1995). Whenever it exists, the global form and meaning (e.g. “cough”) is thus simply used as a memory helper (**mnemonic**). It is worthwhile to mention that “chiff” is a rather well-known verbal descriptor among organ-builders and often represents an unwanted component of the sound (a *noise*). The introduction of other descriptors like “cough,” “hiss” or “kiss” is meant to contribute to the acknowledgement of the importance and the necessity of certain transient components in the sound of flue organ pipes.

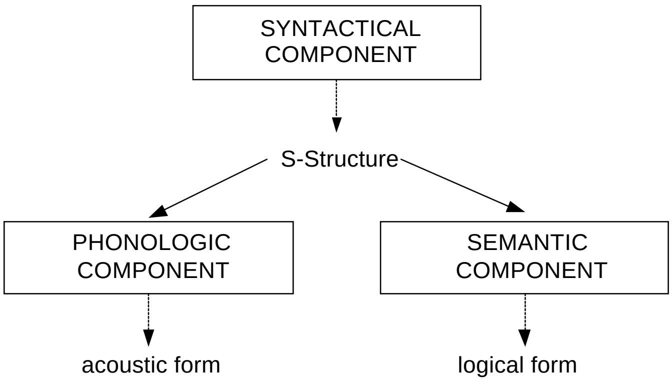


Figure 10. General scheme of a grammar (generative theory), adapted from Ducrot et al. (1995, pg.127). In small letters, the different representations of a statement produced by the grammar. Plain lines indicate an input to a grammar component while dotted lines indicate an output.

Table 15. List of descriptors used for the description of the voicing of flue organ pipes, adapted from M. Yokota (private communication). Revised version of Table 3, p. 11, with emphasis on the type of descriptors used.

Part of the sound	**Verbal descriptors**	**type of descriptors**				
		specific	**extended**	**synesthetic**	**bipolar**	**phonosymbol**
speech or transient	“Chiff”	•				•
	“Cough”	•				•
	“Hiss”	•				•
	“Kiss”	•				•
	Soft vs strong			•	•	
	Slow vs fast				•	
	Short vs long				•	
	Pitch (e.g. octave)	•				
steady state	Amount of fundamental					
	Amount of quint					
	Amount of octave					
	Stringy vs fluty		•		•	
	Round vs sharp			•	•	
	Full vs thin			•	•	
	Light vs heavy			•	•	
	Nasal		•			
	Breathy		•			
	Dirty vs clean			•	•	
	Free vs tight			•	•	
	Floating/spatial vs oppressive			•	•	
general impression	Intense			•		
	Sandy			•		
	Sweet			•		
	Hollow			•		
proportion between	Fundamental and overtones	•				
	Speech and steady state	•				
	Noise and musical tone	•				

Nolle (Figure 6b, p. 16) also reported the use of two other phonosymbols, namely “ping” and “buzz.” Whereas “ping” follows the previous rule of the first part, suggesting then a “xylophone like attack,” (Nolle, 1979) this is definitely not the case with “buzz,” which contains its main sonic information in the last part. Note that the relationship between the symbol “hiss” and its featural signal correspondent was checked in a dedicated experiment (see Chapter IV - and Paper I). Employing phonosymbols can thus be extremely practical in pointing out a particular component of a sound. But even if it often sounds obvious, it is necessary before employing such words, to specify clearly that its possible semantic part is actually used primarily as a mnemonic and that only a restricted part of the word (e.g. first or last part) might imitate the desired sonic component.

We also mentioned the possible use of **metaphors**, used here in the sense of analogies with events belonging to the same sensory modality—and thus a restricted view of its linguistic meaning (Ducrot *et al.*, 1995, pg. 584). Instances of **sonic analogies** do not appear in the descriptors given by the expert but are rather common in the free verbalizations of the participants of our listening tests (see Paper I). Examples of such sonic analogies found are “horn-like, like the sea, like a singer that has a leaky voice.” Other found descriptors such as “strong fabric” or “soft as a cloth” are also metaphors, but their clear synesthetic origin led us to consider them primarily as synesthetic descriptors.

II.3.d - Summary:

- Verbal descriptors of timbre perception do exist but generally suffer from an arbitrary origin.
- Existing lexicons of verbal descriptors are subordinated to specific professional activities where high expertise on sound quality is needed.
- The need for verbal descriptors primarily serves communication purposes between “listeners”, who may be of similar or different types.
- Nevertheless, we can speculate that the internal representation of sounds is dependent on the possible use of verbal descriptors.

II.4 - Sound quality methods and listening tests

Methods used to assess the timbral qualities of musical sounds are often similar to those used in assessing perceived sound quality. Consequently, in the following we will assume that there is no need to distinguish between these two research fields in terms of methodology. A general framework for the study of product sound quality is given in the figure below.

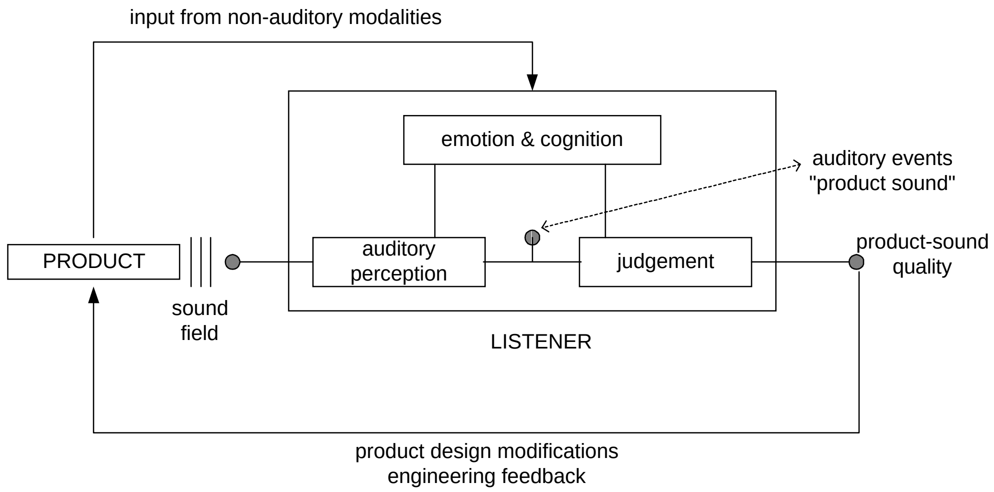


Figure 11. Schematic representation of the psychoacoustical approach to the assessment of sound quality, adapted from Västfjäll (2001b).

In this type of representation, the listener is considered as a “black box” and the context (frame of reference) is not modeled. Nonetheless, from this still very general framework, a great variety of methods have been developed. The problem in the study of auditory perception is that we do not know all its parameters or dimensions. Psychoacoustics research can thus be divided according to goals, the first being to **identify** these dimensions, and the second being to try to **map** them to physical variables. Of the second kind, psychometric functions based on the concept of difference limens have been established. This kind of method falls into the realm of psychophysics, where an isomorphic relationship is postulated to exist between a perceptual attribute and a signal variable. The functional relation between both parameters is then **confirmed** by mapping the signal variable with the mean responses of test participants, thus forming what is called a **normative** variable. Of course, exploratory research generally precedes confirmatory research. A very good example of this research process is illustrated by the two papers on sharpness of steady sounds by von Bismarck. He first made an exploratory study of the possible dimensions of timbre (Von Bismarck, 1974b) and then focused on one particular attribute (sharpness) and tried to map it with a physical variable (Von Bismarck, 1974a).

It is nevertheless important to note that most of the work in auditory psychophysics has been concerned with modeling *loudness* and pitch perception. Even for these rather simple “attributes” (on the existence of which, at least, we all agree), a huge amount of research still did not produce satisfactory models (in terms of prediction). Whereas one might see the source of such difficulties as lying in the complexity of the auditory system (the mechanics of the cochlea, for example), we would rather put emphasis on problems related to the modeling of our perceptual-cognitive judgments. A good review of this issue is presented in Sarris (1999). For a systematic review of sound-quality methods, we recommend papers by Guski (1997) or Bodden (1999). In any case, as our purpose here is to deal with timbre perception, we must admit that this field is rather open for exploration. We thus wish to emphasize **exploratory** methods suitable for the development of our understanding of timbre perception of flue organ pipe sounds.

There is a large number of ways to perform psychoacoustics testing, and we do not intend here to cover the whole range of those that already exist. We nevertheless feel the need to present at least a simple classification that might summarize this variety (see Table 16). Moreover, it is often the case that listening test designs combine several types of stimuli presentations and assessment methods. Paper IV presents a more advanced discussion of pros and cons of the different type of stimuli presentations.

One important concept that does not appear explicitly in Table 16 can be well depicted by the idea that an experimenter has to decide on the limits of the possible answers of the test participants. Bodden (1997) hence talks about free versus forced methods. Forced-choice methods are best symbolized by unidimensional scale (psychophysical) methods that might tend to limit severely the importance of context. In that respect, and acknowledging the fact that this type of method is generally used to build normative psychophysical attributes, the use of such attributes in an artificial listener system might obstruct some parts of the “true” way of listening (Bodden *et al.*, 1998). At the other end, we find the completely free verbalization tests that might reveal some important aspect of perception but might also reflect the personality of the experimenter himself (during the analysis stage). In between these two extremes, one can find several methods varying in the degree of freedom they leave to the test participants. Triggering imagined or memorized auditory events can be rather forced (see e.g. Muckel *et al.*, 1999). It is in fact important to realize that the type and ways of presenting the stimuli will definitely play a role in the results obtained. They might also affect, for example, the type and strength of the physical sources of variability (Toole, 1985).

Table 16. Classification of some basic elements of listening tests design.

<table>
<tr><th>stimuli</th><th colspan="2">stimuli presentation</th><th>input data (obtained from the test participants)</th><th>classes of methods</th></tr>
<tr><td rowspan="9">sounds</td><td rowspan="3">one by one (absolute rating)</td><td>modifiable stimuli</td><td rowspan="2">unidimensional scale (single value output)</td><td rowspan="2">unidimensional psychophysics</td></tr>
<tr><td rowspan="4">fixed stimuli</td></tr>
<tr><td rowspan="2">multidimensional scale (single value output)</td><td rowspan="6">multidimensional analysis</td></tr>
<tr><td rowspan="2">paired comparisons (relative rating, with or without fixed reference)

triadic comparisons</td></tr>
<tr><td rowspan="2">semantic differential (several values output)</td></tr>
<tr><td colspan="2" rowspan="3">multiple comparisons</td></tr>
<tr><td>dissimilarity or similarity ratings</td></tr>
<tr><td>categorization</td></tr>
<tr><td colspan="2">any</td><td rowspan="2">free verbalizations (written or recorded)</td><td rowspan="2">content analysis of verbal descriptions (quantitative or qualitative)</td></tr>
<tr><td>no sound</td><td colspan="2">in memory and imagination</td></tr>
</table>

For our experiments, we have been concerned largely with the exploration of timbre in a “voicing” context. We have used recorded sound files compared in pairs or many at a time. In order to gather a number of verbal descriptors, which could extend the original descriptor of the expert, we collected a number of wordings from various tests and conducted a content analysis of verbal description, both quantitative and qualitative (see Paper II and III).

III - Objective Description of an Experimental Pipe in Terms of Acoustics and Signal Processing

Despite its apparent simplicity, a single flue organ pipe is in itself a complex system the behavior of which is still not completely understood by physicists. Thus, in this chapter we wish only to introduce some of the basic physics of this instrument. A full review of flue pipe acoustics is beyond the scope of this thesis; for a recent overview of related research, the reader is referred to the special issue of *Acustica* on wind instruments (Vol. 86 July/August 2000). The physics of such instruments has been receiving attention in the scientific community since the pioneering work of Helmholtz (1885) and we do not intend here to cover the area in detail but simply to sketch the state of the art. We will hence try to illustrate here the physical description of a flue organ pipe by specific experiments; namely, the analysis of various measurements made on an experimental. This first step will be followed by a detailed presentation of the characteristics of the sound of flue organ pipes by means of a signal processing approach.

III.1 - Physical description of an experimental pipe

A flue pipe can be described as a feedback system, as diagrammed in Figure 12 below.

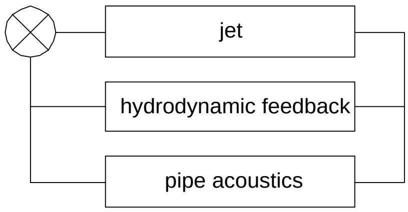


Figure 12. Diagram showing flue pipe functioning, adapted from Verge (1995).

The jet is an unstable hydrodynamic flow that drives an acoustic wave inside the pipe. The interaction between a jet and a resonator is a complex issue that has been described, for example, by Verge (1995). A more complete model is proposed by Fabre et al. (2000) and includes particularly the effects of the pressure reservoir (pipe foot, organ wind channels) and the flue channel configuration specifically studied experimentally by Ségoufin et al. (2000). A discussion of a detailed description of pipe acoustics and its relation to its sound can be found in Miklós et al.(2000). The voicer is principally able to change the configuration of the mouth geometry and hence mainly affects the properties of the jet and its interaction with the pipe acoustics.

In order to examine different configurations of the mouth without having to use different pipes or definitively modify an existing pipe, we built an experimental pipe. Its construction is described in detail in Appendix I : Construction of the experimental pipe, and a photograph is reproduced in Figure 13. In contrast to previous experimental pipes described in the literature, this pipe has a complex mouth shaped similarly to an authentic pipe but offering in addition the possibility of studying different steps of voicing in detail. A step of voicing is represented in this case by a set of three removable pieces, the upper-lip (UL), lower-lip (LL) and languid (LA). At the time of writing, two sets have been built. Measurements made with these two sets follow.

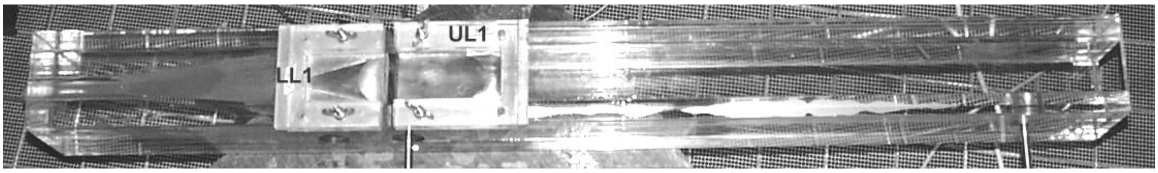


Figure 13. Photograph of an experimental pipe made of Plexiglas with removable mouth parts. The lower-lip (LL1) and upper-lip (UL1) pertaining to Set1 are marked. One can also notice two quarter-inch microphones (MIC1 and MIC2) emerging from the body at the bottom of the picture.

Two major categories of measurements are to be distinguished, namely in passive or active (blown) mode.

In passive mode, it was possible to measure:

- frequency response of the resonator
- end corrections.

In active mode, it was possible to measure:

- internal acoustic field (at two positions)
- foot pressure

In active mode, it was also possible to perform:

- jet visualizations
- binaural recording of the external acoustic field (at one position).

III.1.a - Measurements in passive mode

Each set was constructed according to nominal dimensions of the mouth (decided beforehand by a scaling empirical rule). Each set was then voiced (Appendix I : Construction of the experimental pipe), yielding the mouth dimensions reported in Table 17 and Figure 14.

Table 17. Mouth detailed dimensions for two removable sets.

Parameters	**Set 1**			**Set 2**		
in mm	Left	Center	Right	Left	Center	Right
cut-up (H)	10.8		10.2	11.2		11.3
windway width (T)	0.50	0.64	0.57	0.80	0.70	0.82
languid thickness at front	3.2		3.1	3.00		2.95
languid angle	550			520		
lower-lip thickness	1.00		1.05	1.07		1.06
	no nicks			with nicks		
	no counterface			with small counterface		
upper-lip thickness	0.95					

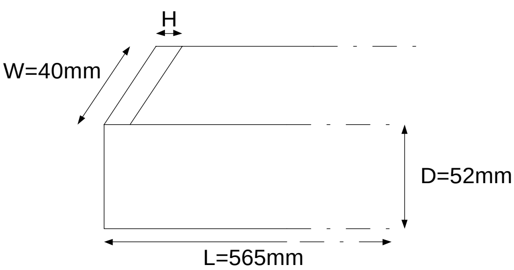


Figure 14. Dimensions of the mouth.

The frequency responses of the pipe (see Figure 16) were measured in an anechoic chamber with the set-up presented in Figure 15.

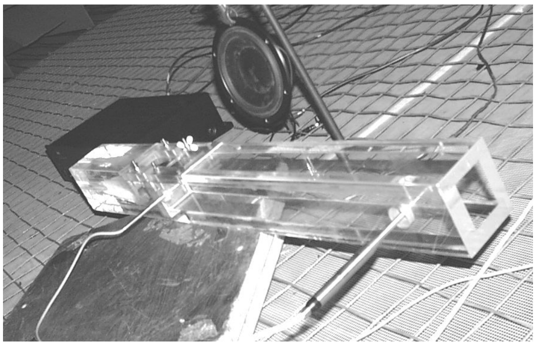
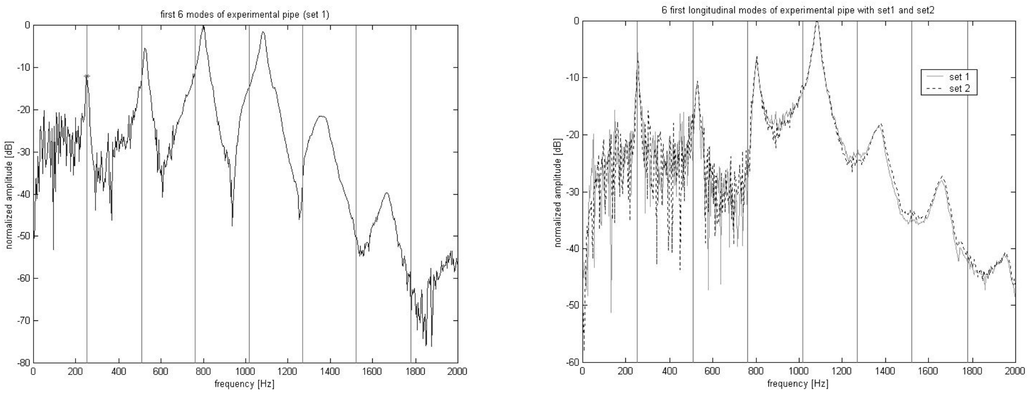


Figure 15. Set-up used for the passive measurements of two mouth sets. A loudspeaker excites the pipe with a pseudo-random signal. The two microphones pick up the filtered response of the pipe.



a) b)

Figure 16. Frequency response of the pipe resonator. (a) Set 1, speaker and microphone close to the mouth region. (b) Set 1 and Set 2, speaker and microphone close to the top (passive ending).

Vertical lines are harmonically placed with a fundamental starting at the first longitudinal mode (with a star on top). M1=254, M2=527,M3=801,M4=1371,M5=1656, M6=1945 [in Hz +/-1.5Hz].

Note that the frequency responses are dependent on the microphone and speaker position, but as far as the frequency peaks are concerned, they are similar. An interesting point that will be further developed in the next section is that the longitudinal modes (M1 to M6) are not harmonically related, due to the frequency dependence of the end-corrections. It is possible to approximate these end-corrections at the top of the pipe and at the mouth (active ending) by means of the equations presented below.

$$F_0 = \frac{c_0}{2 \cdot \left(\Delta L_m + \delta_p + L\right)}$$

Equation 2. Frequency of the first longitudinal mode (which coincides closely with the fundamental frequency of the pipe).

$$c_0 = 331.30 \cdot \sqrt{1 + \frac{T}{273.16}}$$

Equation 3. Speed of sound in function of temperature (346 $m.s^{-1}$ at T=25°C).

$$\Delta L_m = \Delta L_c + \delta_p \ ,$$

Equation 4. Total end-correction at the mouth.

$$\delta_p = 0.6133 \cdot \sqrt{\frac{W \cdot D}{\pi}}$$

Equation 5. End correction of an unflanged pipe (at the passive end but also at the mouth ignoring the constriction).

$$\delta_c = \frac{4H}{\pi} \ln\left[0.5 \cdot \tan\left(\frac{\pi H}{4W}\right) + 0.5 \cdot \cot\left(\frac{\pi H}{4W}\right)\right]$$

Equation 6. End correction added by the mouth constriction.

$$\Delta_c = \frac{1}{k} \arctan\left(\delta_c \cdot \frac{W}{H} \cdot k\right)$$

Equation 7. Equivalent end correction due to the mouth constriction.

$$p_n(x) = \sin\left(\frac{2\pi \cdot f_n \cdot (x + \delta_p)}{c_0}\right)$$

Equation 8. Pressure of the n^{th} harmonic along pipe axis at the passive ending.

From the frequency value of the first mode (M1=254Hz) it is then possible to compute the components of the end corrections. Note that the constriction at the mouth raises the end corrections by a factor of 6. By applying equation 1, we can confirm that the end-corrections are consistent: F0 is very close to M1.

δ_p [meter]	δ_c [meter]	Δ_c [meter]	F0 [Hz]
0.0158	0.0176	0.0682	258

III.1.b - Measurements in active mode

III.1.b.i - Set-up

Measurements taken while the pipe was being blown were made with the set-up shown in Figure 17.

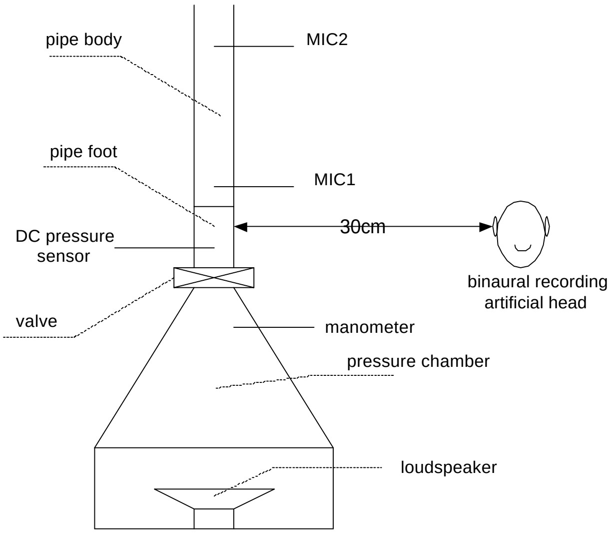
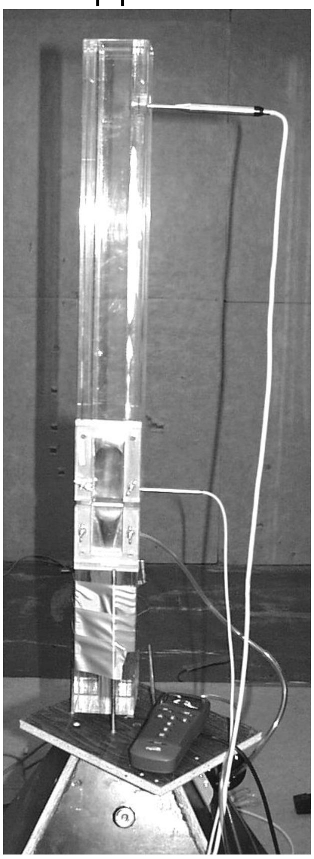


Figure 17. Set-up used for the measurement of the experimental pipe in active mode.

The pipe could be operated at the desired pressure by means of constant pressure input (measured with the manometer). In order to reproduce any pressure fluctuations, it was decided to add a loudspeaker in the pressure chamber. The valve was electromagnetic and was mounted according to Figure 18. As already stated, two microphones were able to pick up the internal acoustic field at two separate locations. In addition, a pressure transducer was able to deliver the pressure inside the foot of the pipe. Moreover an artificial head was placed 30 cm from the mouth of the pipe, a distance which is close to the one occupied by the organ-builder during voicing. All these sensors were linked to a multi-channel acquisition system.[3]



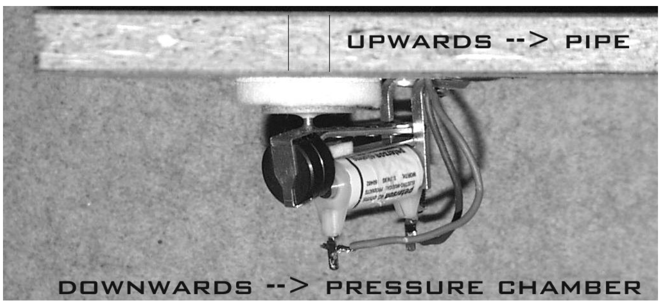


Figure 18. The valve electromagnetically controlled.

Figure 19. Photograph of the set-up. The experimental pipe is placed vertically on top of the pressure chamber via a board under which the valve operates. Two microphones are positioned flush in the pipe's wall. The flexible plastic tube delivers the foot pressure to a transducer. The manometer reads the pressure inside the pressure chamber.

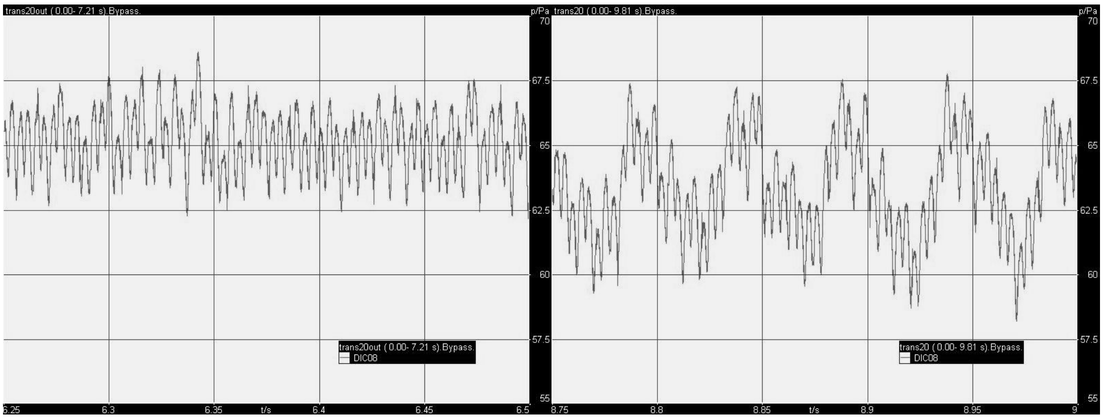
3 Head Acoustics system ®

III.1.b.ii - Measurements

Two measurement sessions were done where the influence of pressure fluctuations and overall properties of the experimental pipe were examined.

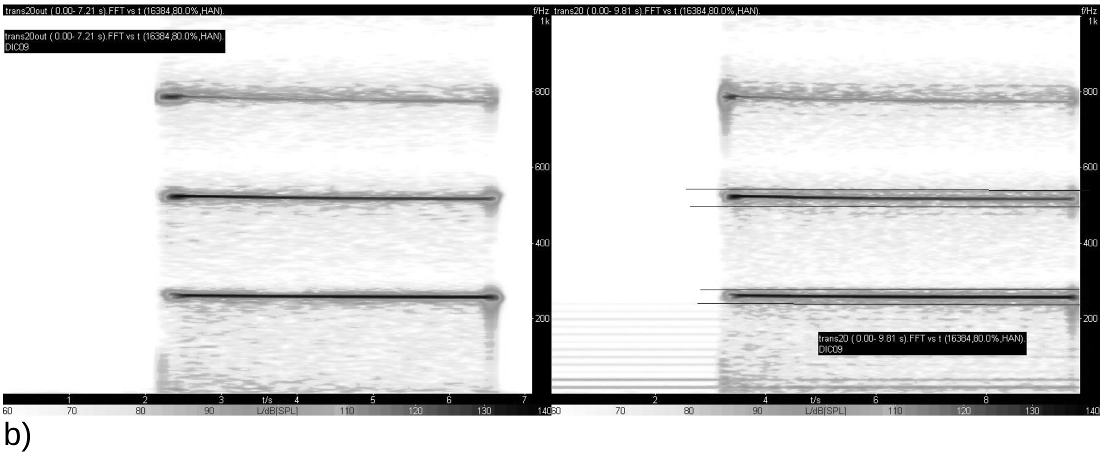
A) PRESSURE FLUCTUATIONS
T=20°C, 53% relative humidity

In order to show the influence of pressure fluctuations inside the pressure chamber, we modulated the input pressure by means of a loudspeaker with a frequency of 20 Hz (frequency low enough not to be heard as such and high enough to be able to be produced by the speaker). A clear difference can be seen in Figure 20 and Figure 21 when comparing screens (a) without any modulations, and (b) with added modulation.



a) b)

Figure 20. Pressure from MIC1, (a) in normal mode of operation, (b) with a 20Hz modulation (brought by the internal loudspeaker)



b)

Figure 21. Spectrograms of the pressure signals from MIC1, (a) without any modulation, (b) with a 20Hz modulation. Modulations appear around the two first harmonics (emphasized with four straight lines).

B) INTERNAL AND EXTERNAL ACOUSTIC FIELDS
T=22°C, 39% relative humidity

A second measurement session showed the possibilities of this set-up for simultaneously recording the foot pressure rise, the internal and external acoustic field for the two sets of mouth pieces (see Figure 23, together with sound samples presented in htmlDoc).

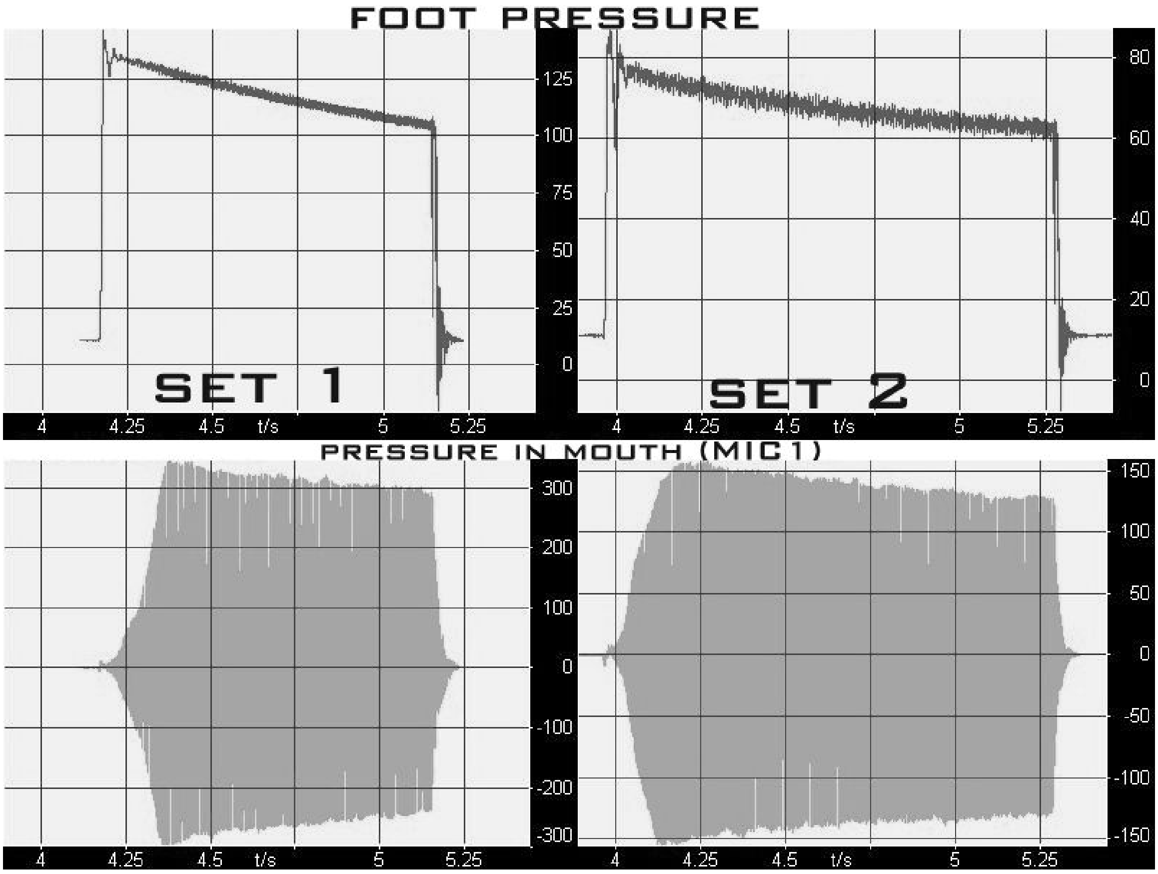


Figure 22. Comparison of the two sets. (x-axis, time [s] and y-axis, pressure).

Moreover, it was possible to check the correspondence between the edgetones (mouthtones) and the foot pressure rise. The pipe resonator was first filled with mineral wool in order to damp all possible activation of acoustic feedback. As can be seen in Figure 23 and Figure 24, the mouthtones activation can be separated into several zones depending on the level of the foot pressure. As expected, the frequency of the edgetone rises with the foot pressure before reaching some kind of saturation. The edgetone is not a simple stimulus but is composed of several harmonics and a wide-band *noise*.

A number of visual representations were also made,4 and short movies are presented in htmlDoc

4 At the Laboratoire d’Acoustique Musicale (Université Paris 6).

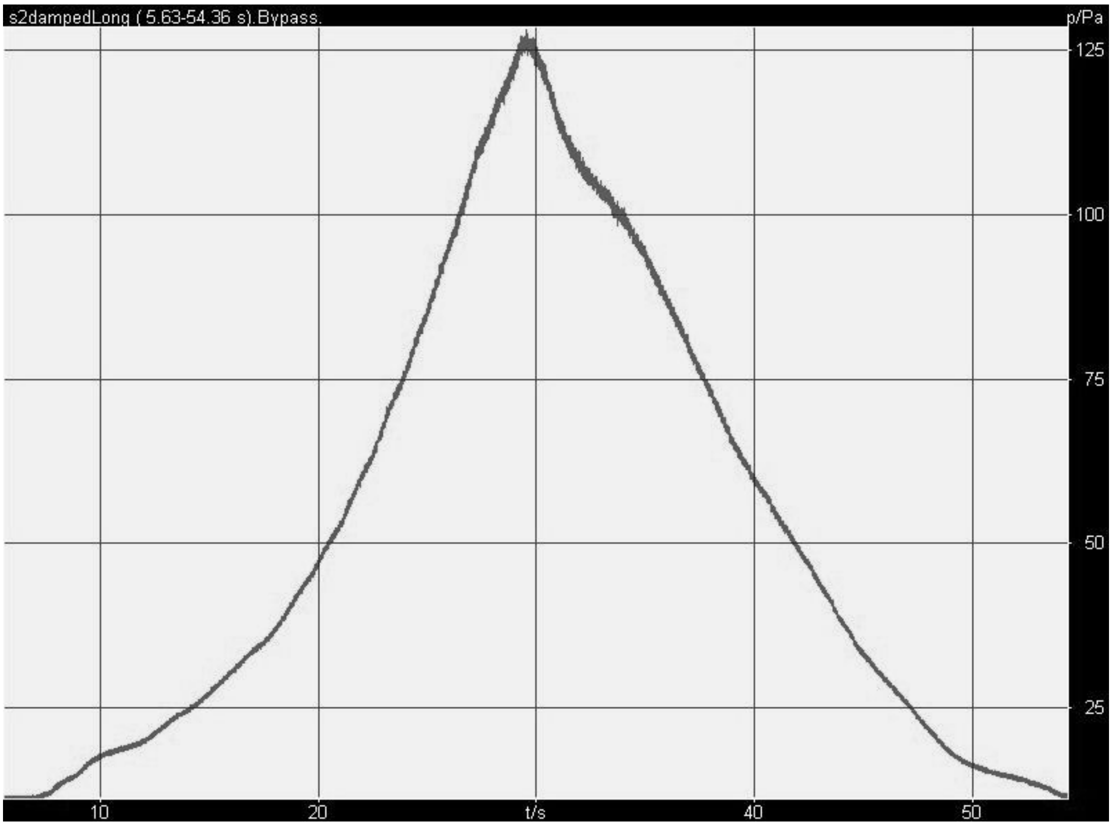


Figure 23. Foot pressure rise with Set 2. (x-axis, time [s] and y-axis, pressure).

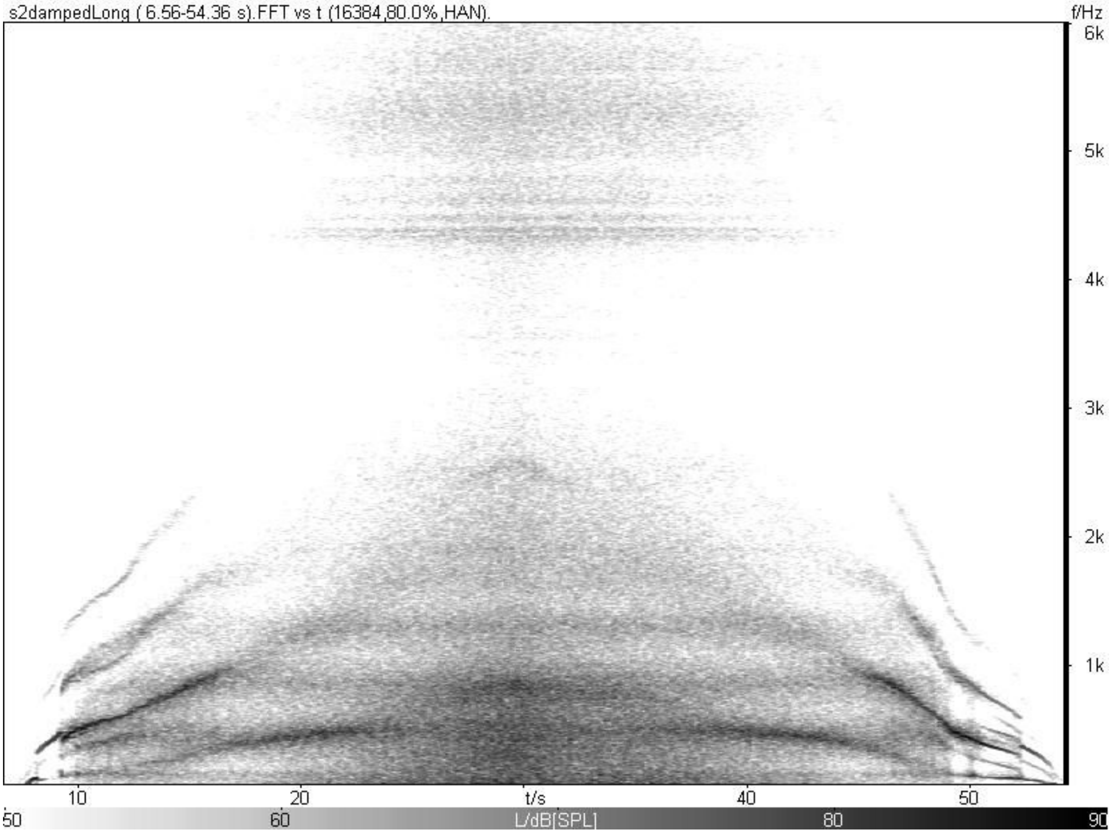


Figure 24. Response at the mouth (MIC1) of the damped pipe with Set 2 to the pressure rise of Figure 23.

III.1.c - Summary

An experimental pipe was built which allows precise measurement of various physical quantities. The removable sets proved to be reliable and offer the unique possibility of making repeatable measurements on the same pipe with different geometry. This is very important to note, understanding the importance of the combined influences of various parameters (Miklós *et al.*, 2000). It was shown that such a pipe can be used to study the mouthtones if properly damped. However, there would be an advantage in a pipe's having a detachable body in order to study the mouthtones separately (Castellengo, 1999). Moreover, it is clear that more than two sets should be studied in order to fully describe the potential voicing trajectories of this pipe.

III.2 - Description of a flue pipe sound in terms of signal processing

As a complement to a physical analysis of the mechanisms of a flue pipe, it is of interest to develop an analysis of the signals produced by such an instrument. The sound of a flue organ pipe recorded in anechoic conditions can be broken down into several parts as depicted in Figure 25.

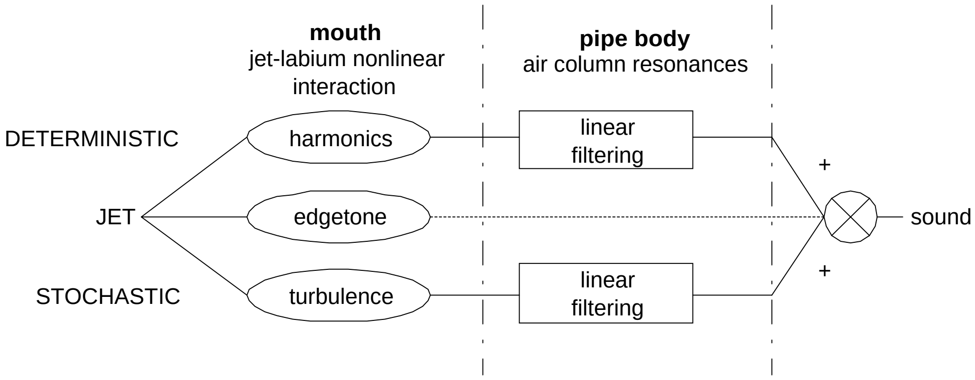


Figure 25. Representation of the sound of a flue pipe in terms of signal processing.

Listening tests and experiments generally provide us with a set of signals corresponding to the sound pressure at different points for different pipes or different settings of the same pipe. All these signals have to be processed in order to extract useful objective information.

We can distinguish three categories of signal treatments:

- Homogenization or pre-processing of the sound files relative to a set
- Separation of periodic and *residual* parts
- Extraction of objective parameters.

III.2.a - Homogenization / Sound pre-processing

III.2.a.i - In the amplitude domain

In general, in order to represent a physical variable, signals need to be calibrated. The International System of Units (SI) is used (however, one must note that organ-builders tend to work using heights of water column to measure the input pressure). We mainly designed listening tests that were not centered on loudness. In that case, the calibration of those sound files is of no importance, but a perceptual level *equalization* relative to the set of sounds under study was applied. When subjects were asked to focus on the transient parts, the decay parts of sounds had to be faded in a standard way. A decreasing linear ramp was used for this purpose. When no calibration is needed, signals are normalized (divided by their maximum of amplitude).

III.2.a.ii - In the time domain

The pressure signal equivalent to a single note can be divided into three main parts. The division into these parts is justified by the fact that very different physical phenomena occur during the transient, stationary and decay parts (see Figure 26), even though, all together, they form a single entity. This segmentation process, though apparently simple, is of crucial importance since we know that our hearing is very sensitive to such transitions. Nevertheless, no precise definition of where to place the articulation points exists. In order to set the segments, we combined visual inspection and arbitrarily chosen durations, using a dedicated computer program[5]. Pollard (1988) describes an automatic segmentation method based on the detection of zero-crossings of the harmonics derivative. Keeler (1972) characterized the mean transient time for 200 organ pipes and normalized it to the number of periods of each sound, which turned out to be less than 50 for all cases. But it should be stated that the transient part is a mixture of several phenomena (see Paper , Fabre (1992)) and its perceptive duration can be different from its physical one (see chapter IV, test on speed of transient). As a rule of thumb, the duration of the transient part was chosen to be of the order of 50 times the period of the fundamental. In order to compute the spectrum of the stationary part of the sound an “artificial” segment was used, containing a suitable number of samples (2^N) for the requirement of the Fast Fourier Transform. A presentation of the variables implied in this process is shown in Figure 26.

[5] Matlab® routine **tick.m** present in htmlDoc.

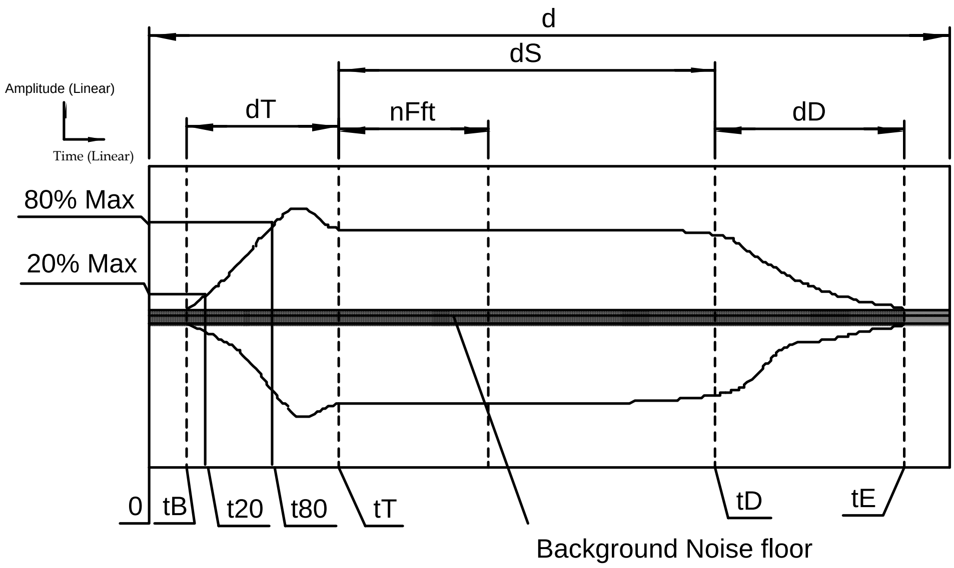


Figure 26. Time segmentation of a single sound.
d=duration, t=time instant, T= transient, S=stationary, D=decay, tB=time of beginning, tE=time of end

Additionally, a synchronization of each sound was made. This was done either by internal synchronization (similar to t20 or t80) or by reference to an external signal (e.g. the trigger signal of the valve or the pressure rise in the foot). For the creation of testing sounds it is also important to control the total duration d.

III.2.b - Separation of periodic and residual parts.

In Figure 26, the domain marked in gray represents the background *noise*. This noise is produced by the recording chain (microphone, amplifier, tape recorder, etc.) or by the environment. The pipe itself is a source of “noise” (the *stochastic* part of Figure 25), and each type of pipe has its particular “noise characteristics.” Besides this “noise” there are the harmonic components that give the sound of a pipe its main pitch and timbre. We believe that this “noisy” part also plays an important role in the distinction of the many different timbres that can be created by an organ. Hence, instead of the term “noise” we prefer to employ the term *residual*.

In order to study the influence of the harmonic (or periodic part) and the influence of the residual on the timbre separately, then, we must achieve a reliable way to decompose the original sound (which should allow no audible differences after reconstruction). In the following section we enumerate and compare several possible methods. In order to compare the algorithms, a “real” sound and test signal were used. The results of this decomposition are also presented in hmtlDoc (audio) and in Appendix III (figures).

III.2.b.i - On the test stimuli

The test signal was constructed by *additive synthesis* and contained some simple features of an organ pipe's sound. It is consequently far from being a “perceptual” replica of a flue pipe sound. It was designed primarily to exhibit simple arbitrary characteristics typical of an organ pipe tone, on which the power of each algorithm could be tested. As the noise and the harmonics are constructed separately and then added, the separation process can then be checked in an objective way.

The different parameters of the test signal are given below:

Deterministic part:
Fundamental frequency 200Hz and 19 overtones -> nbHarm=20
Amplitude scaling of each harmonics (noHarm):
amp(fundamental)=1

Segmentation:
tB=.2 s except for first overtone .1s
dT=10 ms
overshoot of 5%
dD=20 ms

Amplitude of the noise: 0.01 (relative to the amplitude of H1)
tB=.1s
Number of the air-column modes: 20
Q value: 30 for each mode
Frequency of first mode 200 Hz
Inharmonic ratio 5%

A real sound was also chosen to test these methods (sound “s4”). This sound belongs to the set of copies made from a single original old pipe. The sound “s4” was recorded at the beginning of the voicing process.

III.2.b.ii - On the methods

A common generic signal model can be written as:

$$\boxed{\begin{aligned} s(t) &= \sum_{k=1}^{K} A_k(t)\sin[\theta_k(t)] + r(t) \\ \theta_k(t) &= \int_0^t \omega_k(\sigma)d\sigma + \alpha_k + \Phi[\omega_k(t)] \end{aligned}}$$

Equation 9. Signal model

Here A_k represents the K time-varying amplitudes of the sinusoidal components. In the case of the organ sounds, these components are of fixed frequency except during the transient, where some deviations may be observed due to the important variation of phase delays on the jet (Coltman 1976). The phase of the kth harmonic $_k$ is computed via the frequency of the harmonic, θ_k, a possible constant phase shift, α_k, and via an eventual frequency shift controlled by the functional Φ. r(t) denotes the *residual* part, since it may be obtained by subtraction of the deterministic/periodic part from the original signal s(t).

Some methods of separating the *noise* and the harmonic part work on a frame-by-frame basis. The waveform is segmented into constant time windows. A best-fit algorithm such as the least-square method is applied to each window in order to find the best value for Ak and $_k$. At the boundary of each frame, interpolation and correspondence matching of each parameter must be made carefully in order to avoid discontinuities which would distort the resulting sound badly. The residual part is modeled as a source-filter model. In the case of the organ pipe, the resonances of the air-column are modeled by an autoregressive (AR) process by using the Linear Predictive Coding method, Rabiner and Schafer (1978). This strongly reduces the information storage, because only a few coefficients are needed to code the corresponding Infinite Impulse Response (IIR) filter. Note that it is also possible to represent the noise by a sinusoidal model as proposed by McAulay and Quatieri (1986) and implemented e.g. by Fitz et al. (1997). For a high-quality rendering (including noise), the storage-cost is high and would in most cases lead to an “information explosion” Ding (1997).

Note that many of these algorithms come from the speech-signal processing branch of science. Flue pipe sounds and the human voice actually share many common characteristics.

Three methods using the previous principles were used to decompose “s4” and the test signal. We applied the Spectral Modeling System (SMS) method proposed by Serra (1990 & 1997) and another method called QUASAR, Ding (1997). A third method proposed by Guettler (1998) (referred to here as the K.G. method) working on a single FFT was also tried out. In addition, we have implemented a filtering method based on the theory of the heterodyne bank filter complemented with a Fast Fourier Transform (FFT) zoom (see Figure 27 and Figure 28).

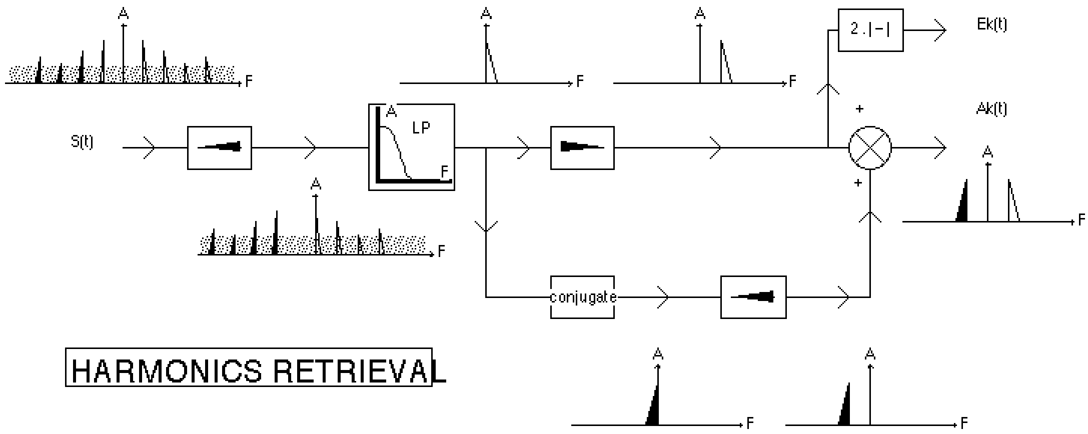


Figure 27. Algorithm used to retrieve one harmonic (periodic, deterministic) part. The original signal s(t) is translated and Low-Pass filtered so that only one component of the conjugate pairs of the harmonic Ak(t) is retrieved. A Hilbert transform is performed in order to get the envelope Ek(t).

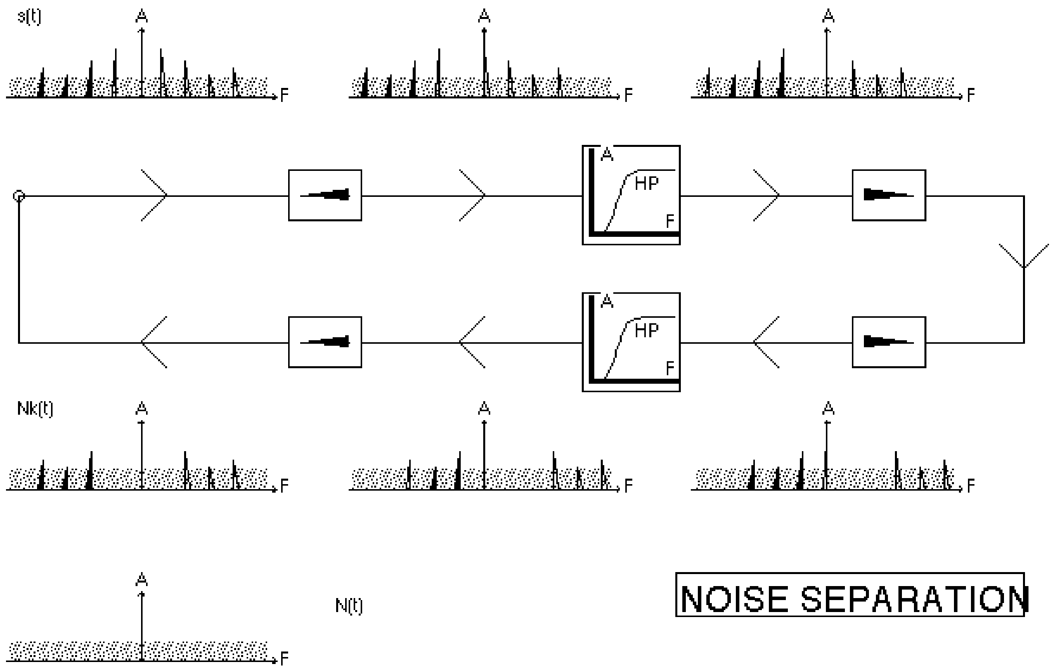


Figure 28. Algorithm used to separate the residual (or noise) from the harmonics. From the original signal s(t), each component of the harmonic conjugate pairs is filtered out sequentially until only the residual Nk(t) is left.

Results of the decomposition of “s4” are presented as sound files (see htmlDoc) and as a set of figures (see Appendix III: Figures from test on separation methods.). Once this separation process has been applied to the whole set of sounds in question, further objective analysis can be achieved. This is described in the next section.

III.2.c - Objective parameters

III.2.c.i - Stationary analysis of the deterministic signal

An FFT is performed on the stationary part of the signals, assuming that we just look at a part of a signal that extends to infinity. The linear amplitudes Sk of the harmonics can then be estimated by detecting the spectral peaks. An alternative way to find the Sk without working on the deterministic signal is to construct a denoised stationary part by time averaging over single-period time frames. The averaged frame is then replicated to form a perfect periodic signal which by FFT gives an estimate of Sk. The spectrum gives us the amplitude of the harmonics—and thus the harmonic spectral *centroid* or spectral center of gravity defined by:

$$Fc = F_0 \frac{\sum_{k=1}^{K} kS_k}{\sum_{k=1}^{K} S_k} \quad \text{, where } F_0 \text{ is the fundamental frequency}$$

$$Nc = \frac{Fc}{F_0} \quad \text{, normalized centroid index}$$

$$Sc = \frac{1}{K}\sum_{k=1}^{K} S_k$$

Equation 10. Definition of spectral centroid.

Fc and Ac are useful to compare “*sharpness*” of deterministic signals of the same *pitch* . Nc is a normalized version of Fc and makes possible the comparison of the harmonic spectral centroïd of sounds of different pitch.

III.2.c.ii - Transient analysis of the deterministic signal

Harmonic envelope

For each harmonic component, a rise time (based upon the synchronization times, t20 and t80) and a time offset (based upon any synchronization time) are computed. In order to reduce the information contained in all the harmonics, we introduce a revised version of the instantaneous harmonic central spectroid proposed by Beauchamp (1982).

$$BR(t) = \frac{\sum_{k=1}^{K} kA_k(t)}{rms[r(t))] + \sum_{k=1}^{K} A_k(t)}$$

Equation 11. Instantaneous Brightness.

Global envelope

The envelope of the original signal is obtained by low-pass filtering with a cut-off frequency lower than F0. Note that the Hilbert transform can only be used to calculate the envelopes of the harmonics since it is only applicable to narrow-band signals, see Hartmann (1997).

III.2.c.iii - Transient analysis of the residual signal

The spectrogram (sonogram) is used. A number of FFT are processed on constant and overlapping time frames. A smoothing window is applied to each frame in order to prevent boundary discontinuity effects.

IV - Listening Test Experiments

IV.1 - Introduction

During the course of this study we designed and administered several types of listening tests. In this chapter we will present three experiments that were aimed at establishing a correlation between subjective and objective attributes of flue organ pipe sounds. Subjective attributes were chosen either from the specifications of the voicer (see Chapter I) or from an extended list of verbal descriptors established by other experiments presented in Papers II and III.

The administration of listening tests is a delicate and critical procedure. In our case, we were interested in exploring and checking the consistency of timbre perception among different type of “expert” listeners like organ-builders and musicians. In order to get these experts to participate in our tests, we had little other choice (for practical reasons) than to ask them to perform the test while they were at a conference or symposium organized by the GOArt work group in Göteborg. Moreover, the requirements of careful and close listening dictated by our tests led us to think of a system that could be used in an individual and independent manner. Our choice thus quickly pointed toward computers, as their flexibility and performance with regard to raw result storage, interactive user interface (e.g. **sonic icon**), and audio processing, as well as analysis facilities (statistical, signal processing) are unequalled. That is why we developed a computer program that could fulfill this task (the first implementations being designed by the present author and programmed by Scholz *et al.*(1999) and the latest being represented by the LISE[6] environment designed and programmed by the present author).

IV.2 - Primary attributes mapping

This first experiment was constructed to find out how sounds s0 to s9 (see Chapter I and Appendix II) of principal pipes being voiced could be perceived. As can be experienced by direct listening (cf. htmlDoc[7]), these sounds present small differences. However, each of these differences is representing a voicing step. Hence, the main hypothesis behind this test is to check whether we can recognize which is the perceptual attribute that the voicer wanted to change during this voicing step. The session test was organized in three different modules:

- a grouping and scaling test (see Paper IV)
- a pair comparison of global difference (see Paper IV)
- a semantic differential test (each sound s1 to s9 was compared with the reference s0).

[6] LISE (Listening Interface for Sound Experiments) is provided and documented in the htmlDoc. It consists of a set of modules written in Matlab®. Each module can perform a specific test such as a pair or a multiple comparison (see Appendix V). A particular module aimed at scaling, categorizing and verbalizing audio stimuli receives special treatment in Paper IV.

[7] Companion html document with audio illustrations (see table of contents).

The scale used for the semantic differential test was made of four attributes:

- speed of the transient
- amount of transient noise
- brightness
- nasality
- balance between the noise (residual) part and the harmonic part.

The test was administrated to 22 participants thanks to a computer interface (see Figure 29). They listened to sound files (sonic icons) through headphones.

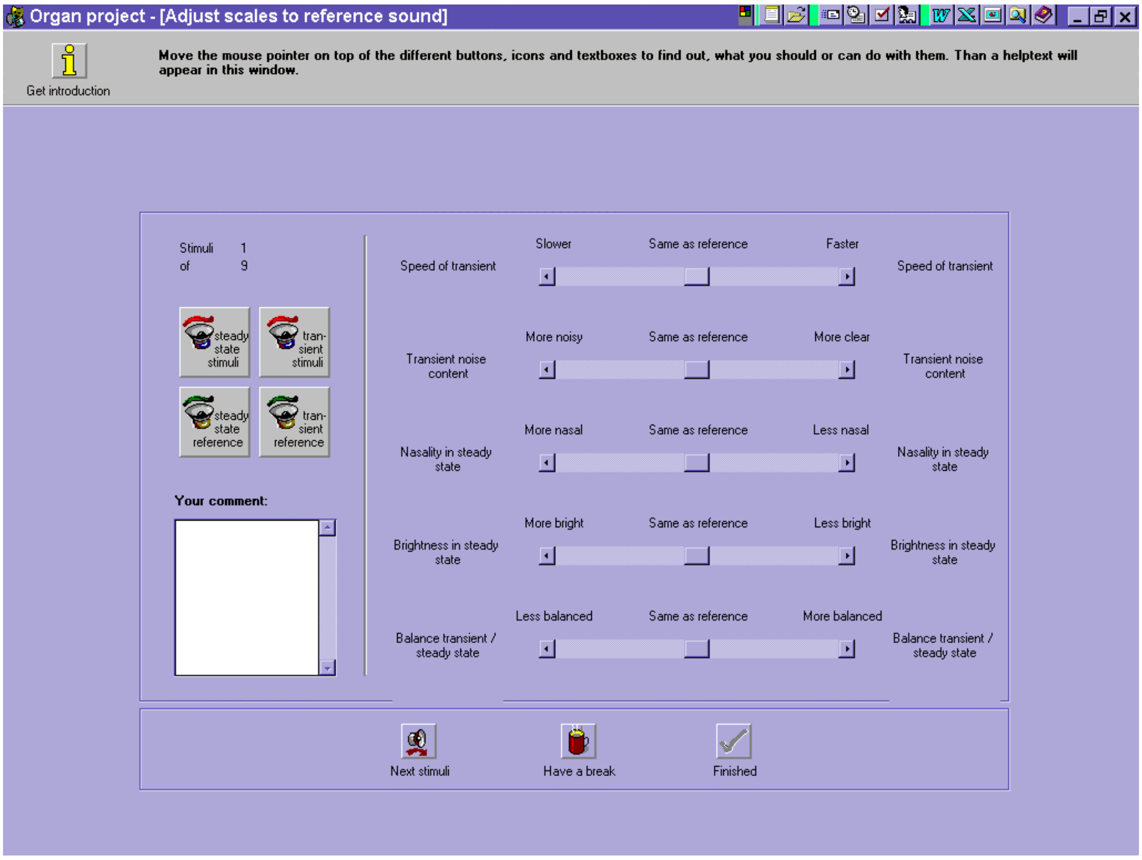


Figure 29. Schematic representation of a Semantic Differential test. In the top left corner, four push buttons allow the participant to listen to the stimuli and the sound of reference and to their transient part only. In the bottom left corner, a text field allows the participant to write his/her comments at any time. The main screen is composed of a set of 5 scales.

In order to check the correlation between the subjective parameters extracted from this test and objective parameters, sounds were processed according to the procedure described in Chapter III. Pitch was calculated on the steady-state part (see Table 18). Global envelopes were calculated. Harmonic and residual parts were separated. Hence, in the following we will describe the treatment of each one of these parameters separately.

Table 18. Pitch of all sound files (obtained with **pitch.m**, see htmlDoc).

s0 : 191.11 Hz : F2 and 6 cents +/-6 cents (0.67) Hz
s1 : 189.76 Hz : F2 and -7 cents +/-6 cents (0.67) Hz

s2 : 189.76 Hz : F2 and -7 cents +/-6 cents (0.67) Hz
s3 : 189.76 Hz : F2 and -7 cents +/-6 cents (0.67) Hz
s4 : 189.76 Hz : F2 and -7 cents +/-6 cents (0.67) Hz
s5 : 191.11 Hz : F2 and 6 cents +/-6 cents (0.67) Hz
s6 : 191.11 Hz : F2 and 6 cents +/-6 cents (0.67) Hz
s7 : 191.11 Hz : F2 and 6 cents +/-6 cents (0.67) Hz
s8 : 192.45 Hz : F2 and 18 cents +/-6 cents (0.67) Hz
s9 : 192.45 Hz : F2 and 18 cents +/-6 cents (0.67) Hz

IV.2.a - Speed and duration of the transient

Participants rated the speed of the transient by comparing their perception of sound s0 (reference) with sounds s1 to s9. The mean result of these ratings is presented in Table 19 and in Figure 31. In order to compare these results with an objective parameter, the duration of each transient was automatically computed by detection of time instants at which 20 and 80 percent of the maximum amplitude of the signal was reached (see Figure 30).

Table 19. Subjective ratings on the "speed of the transient" and objective measure of the transient duration.

sounds	**s0**	**s1**	**s2**	**s3**	**s4**	**s5**	**s6**	**s7**	**s8**	**s9**
voicer		-3	-2.8	-0.6	-7	-1.4	1.2	-2.8	-0.4	2.8
GOArt (mean)		-4.44	-2.48	-0.36	-5.16	-2.15	-0.58	-2.23	-3.34	-1.69
transient duration [ms]	39.0	98.3	44.0	50.6	182.4	44.3	35.6	38.3	74.3	88.0

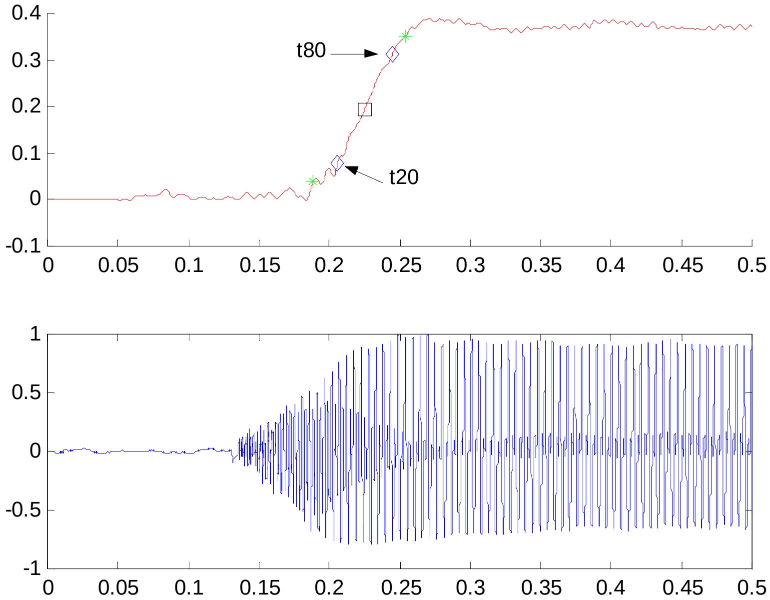


Figure 30. Rising time in the transient part. of sound s0. The upper plot shows the global envelope, while the bottom plot shows the time signal. x-axis is time, y-axis is normalized amplitude.

Of course the rise-time of harmonics is very important in the perception of the transient duration (listen, for example, to s0 and s4 in htmlDoc). But some components of the residual might play an important role as well (listen, for example, to the higher frequency component (around 3Khz) in s0 compared with s6; s0's speech seems to be a bit slower, while this is not easily detectable from the global envelope). Inspection of Figure 31 shows a fairly good agreement between the ratings of the voicer and the computed duration of the transient (except for sounds s7 to s9, which suffer from too high a level of noise at low frequency). Moreover, despite large standard deviations, the mean ratings of the group of 22 participants fit rather well with the previously mentioned data. This might lead us to think that a psychophysical parameter could work for this particular perceptual attribute, but we would maintain a rather pessimistic attitude, however.

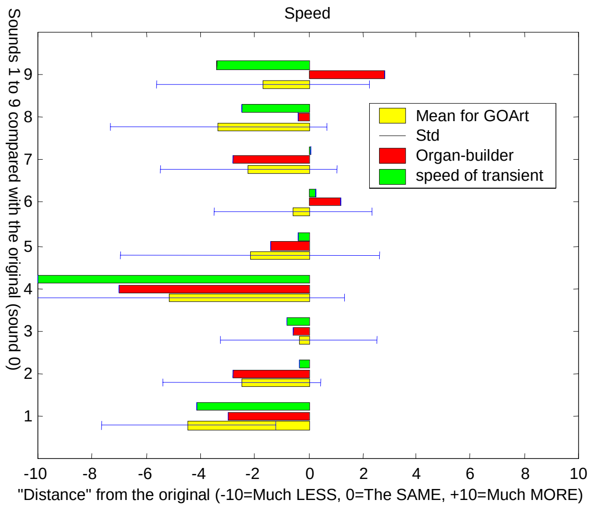


Figure 31. Mean ratings of 22 participants are plotted against the rating of a voicer and an objective parameter for the scale "speed of the transient."

IV.2.b - Amount of transient noise

The same conclusion holds here for the scale of “transient noise amount.” From an inspection of Table 20 and Figure 32, it is rather clear that one can hardly draw any conclusion on the correlation between the ratings of the voicer and the group and the adequacy of an objective parameter such as the proposed rms (root mean square) value of the noise calculated during the transient part.

Table 20. Subjective ratings on the “transient noise” against a proposed objective measure.

sounds	s0	s1	s2	s3	s4	s5	s6	s7	s8	s9
voicer		1	2.6	1	-7.4	-2.8	2.8	-0.4	0.4	5.6
GOArt (mean)		-3.44	-3.21	1.15	-7.63	-3.60	-0.58	-0.84	-1.43	0.87
RMS noise in transient (*10e-2)	1.69	1.44	1.08	1.10	1.37	1.30	1.27	1.96	2.92	2.77

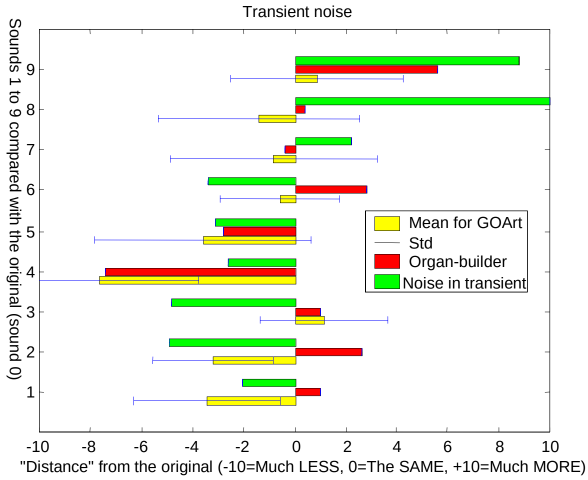


Figure 32. Mean ratings of 22 participants are plotted against the rating of a voicer and an objective parameter for the scale “noise in the transient.”

IV.2.c - Brightness and nasality

The treatment of the perceptual attributes “brightness” and “nasality” offers another perspective, as it is clear that they might overlap, aiming to describe the same thing. illustrates that this is the case for the voicer, whose ratings are consistent across all sounds (nasality and brightness correlate strongly). However, the mean ratings of the group show no apparent connection between brightness and nasality.

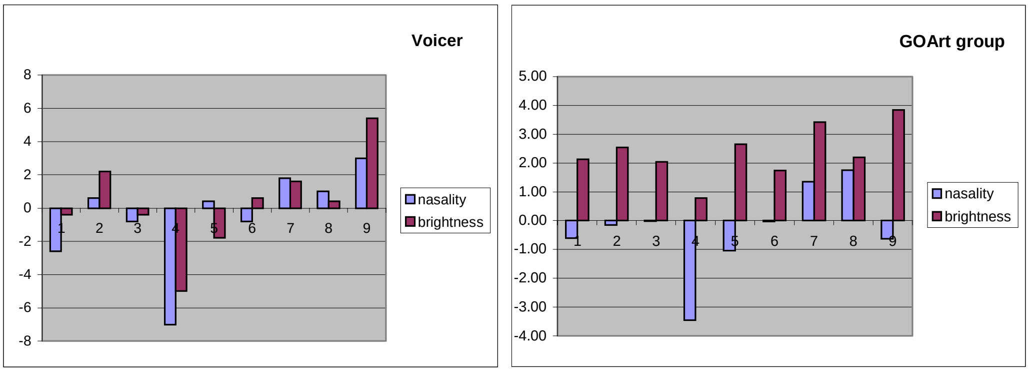


Figure 33. Ratings of the voicer (a) and group (b) of the nasality and brightness of the nine sounds (s1 to s9 in x-axis).

Table 21 indicates the ratings of the test participants as well as a proposed objective parameter computed from the time-varying spectral centroid (mean of the stationary part). Choosing such a physical variable, we thus made the hypothesis that brightness and nasality emerge from the stationary part of a sound. This might not always be the case, but at present we have no example to the contrary.

Table 21. Subjective ratings on “brightness” and “nasality” against a proposed objective measure (mean of the time-varying *spectral centroid*).

group\sounds	**s0**	**s1**	**s2**	**s3**	**s4**	**s5**	**s6**	**s7**	**s8**	**s9**
nasality		-0.60	-0.15	-0.01	-3.45	-1.05	-0.03	1.35	1.75	-0.63
brightness		2.13	2.54	2.05	0.79	2.65	1.75	3.42	2.19	3.85
Voicer										
nasality		-2.6	0.6	-0.8	-7	0.4	-0.8	1.8	1	3
brightness		-0.4	2.2	-0.4	-5	-1.8	0.6	1.6	0.4	5.4
centroid mean	2.277	2.302	2.222	2.219	2.402	2.296	2.330	2.071	2.374	1.981

Inspection of Figure 34 and Figure 35 shows that the centroid correlates better with brightness than with nasality, which is rather encouraging (the spectral centroid was actually introduced in sound analysis by Beauchamp (1982) under the term “brightness”).

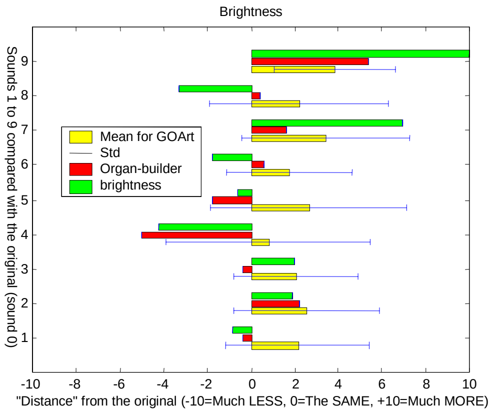


Figure 34. Mean ratings of 22 participants are plotted against the rating of a voicer and an objective parameter for the scale “brightness.”

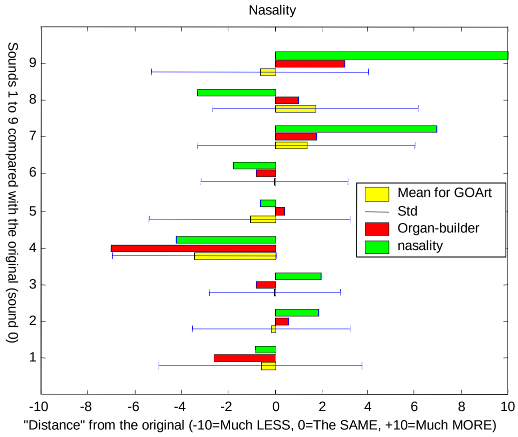


Figure 35. Mean ratings of 22 participants are plotted against the rating of a voicer and an objective parameter for the scale “nasality.”

IV.2.d - Balance between noise and harmonics

Ratings of the “balance between noise and harmonics” are presented in Table 22. Correlations between the ratings of the voicer, the participants and the objective estimate are too low to draw any conclusion on this important aspect of voicing. Probably, just as in the case of the “noise in the transient,” the term “noise” is confusing, covering several meanings. However, it is interesting to note that s7, s8 and s9 have a rather low signal-to-noise ratio (listen to their residual components, especially at low frequency).

Table 22. Subjective ratings on the “balance between noise and harmonics” against a proposed objective measure (Signal to Noise ratio).

sounds	s0	s1	s2	s3	s4	s5	s6	s7	s8	s9
voicer		0.6	-2.6	-0.4	10	-0.2	-1.2	-0.2	-0.6	-3.8
GOArt (mean)		1.55	1.13	-0.41	7.30	4.15	-0.27	-0.22	-0.10	2.26
SNR	28.3	27.5	33.1	31.8	28.7	28.7	34.6	16.8	10.3	14.8

IV.2.e - Features of sounds

In addition to the previous ratings of the five scales, participants were asked to answer a few questions concerning different features of each sound, especially during the transient. The question was: “Do you think that sound (s#) is (or has some) ------ compared to the sound of reference (s0)?” for the descriptors “free,” “tight,” “sweet,” “harsh,” “chiff” and “cough.” Results of this questionnaire are presented in Table 23.

Table 23. Percentage of “yes” answers.

	s1	s2	s3	s4	s5	s6	s7	s8	s9	average
Free	32	61	53	16	21	53	47	58	42	42
Tight	16	33	37	37	16	37	32	53	37	33
Sweet	26	39	37	5	21	32	26	53	21	29
Harsh	53	22	26	79	58	37	42	16	32	40
Chiff	53	61	32	95	79	63	53	63	37	59
Cough	53	33	58	53	74	37	26	47	42	47

In Figure 36, we present what could be called a “featural” profile of a selected number of sounds, namely, s1, s6 and s9, because each of them represents a rather different step of voicing (s1 - first step, s6 - last step, s9 - last step of a copy pipe with added nickings). It is interesting to note that the profile of s6 is quiite close to the mean profile over all sounds. Moreover, this type of representation offers us a way to check what happened during several voicing steps. In fact, between steps 1 and 6, the sound of the pipe got more “free,” more “tight” and had more “chiff,” while “cough” and “harshness” were decreased. This can also be recognized in Figure 37, where the evolution of this type of profile is presented for all sounds.

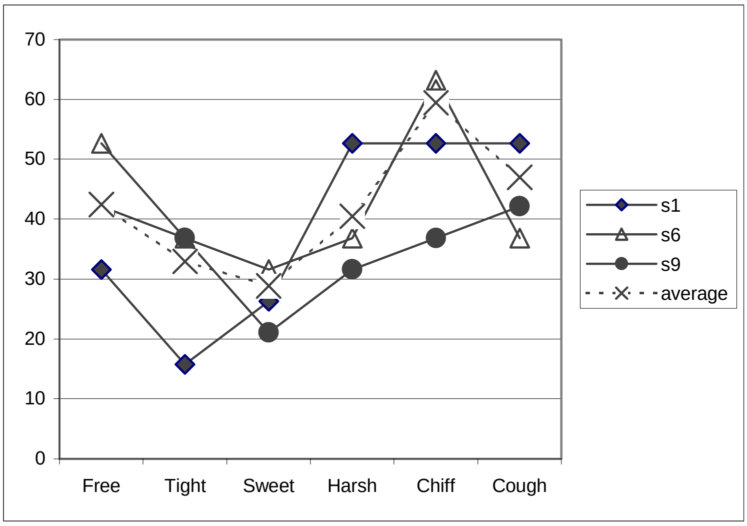


Figure 36. Featural profile of three sounds and the average over all sounds.

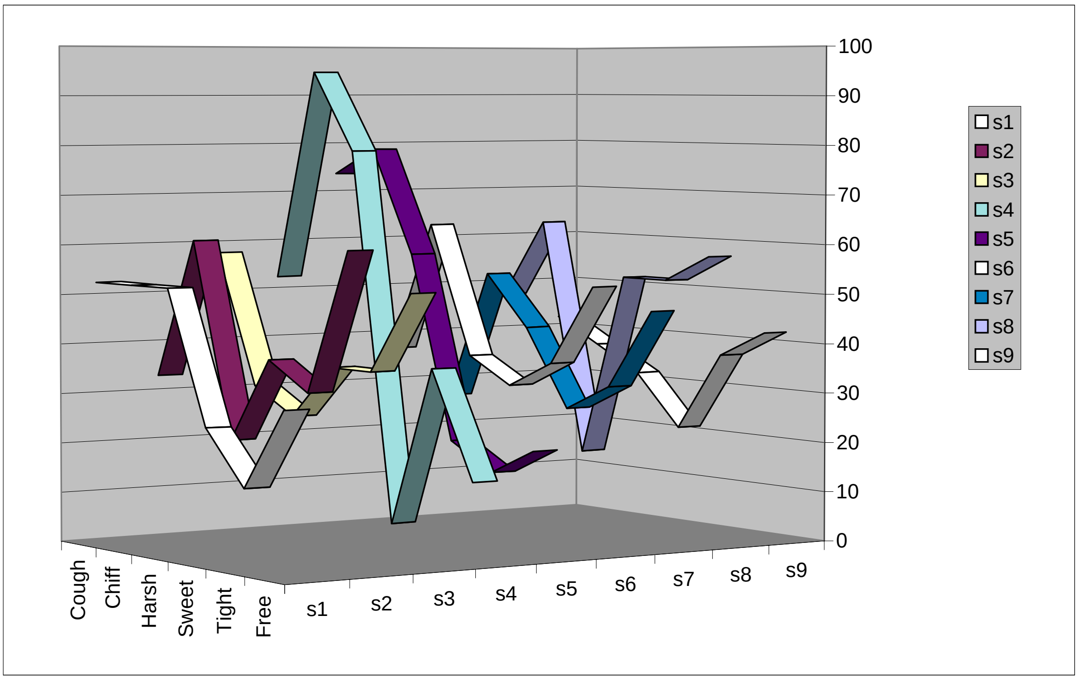


Figure 37. Evolution of the featural profile over several voicing steps.

IV.2.f - Summary

All these preliminary experiments led us to think that it is extremely difficult to find a physical parameter which can render the many facets of a single perceptual attribute (like transient speed), especially when differences are slight. It might thus be impossible to find normative parameters suitable for this kind of task.

We also recognized the importance of a precise description of the transient part in both its harmonic and residual components, especially, though the use of phonosymbols like “chiff” or “cough” (the phonosymbol “hiss” is treated in Paper I and in Rioux (1999). Moreover, the term “noise” should not be used as a descriptor; it covers too many meanings. A way to check whether this type of confusion appears is treated in the next section.

IV.3 - Perception of harmonic vs. residual parts

This test was intended to check the perceptual validity of the harmonic+residual separation. The task presented to the test participants is sketched in Figure 38.

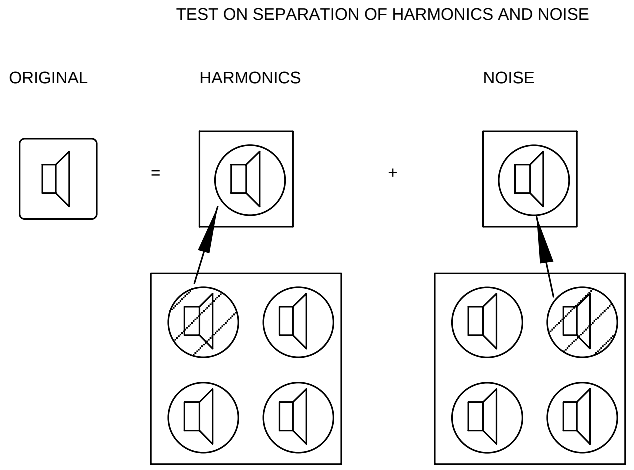


Figure 38. Sketch of the Harmonics+Residual separation test (user graphic interface).

For each original sound appearing in the box labeled “ORIGINAL,” the test participants had to find the corresponding harmonic and residual components. All the sounds appear as “sonic icons.” When pressed, these icons produce a sound. They also can be moved according to the usual “drag and drop” scheme. Sounds were produced by the pipes described in Appendix II and called “G1, G2, G3, g1, g2.”

The raw answers from this test are plotted in Figure 39.

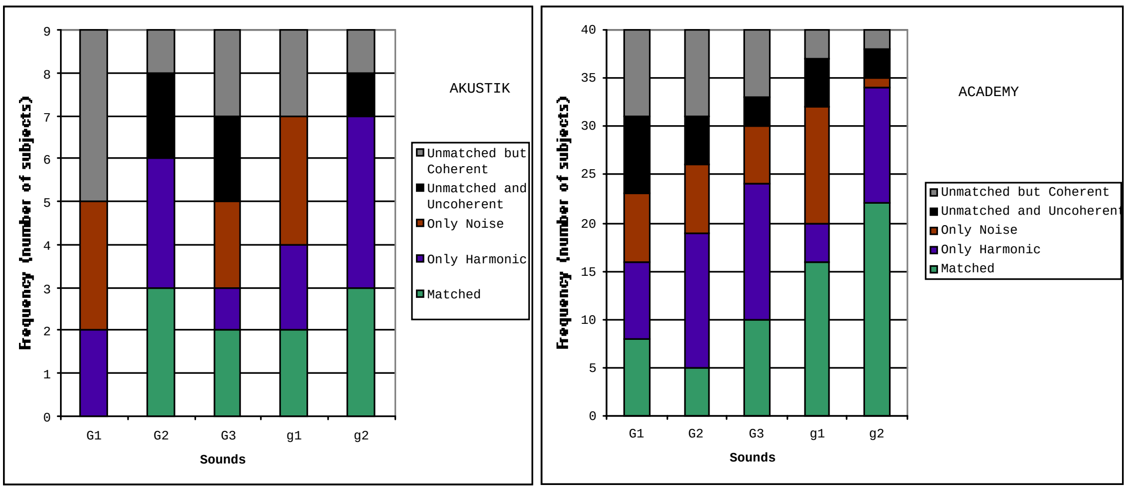


Figure 39. Answers to the separation test for two groups of listeners.

Each column represents the frequency of responses (number of test participants) obtained according to the different types of answers possible for each sound. The types of answers were classified in the following way:

- **Matched**: The harmonic and the residual part correspond exactly to those of the original sound presented.
- **Only harmonic**: Only the harmonic was recognized.
- **Only residual**: Only the residual was recognized.
- **Unmatched and uncoherent**: The harmonic and the residual parts did not correspond to those of the original sound. Moreover, added together, they did not match with any of the original sounds.
- **Unmatched but coherent**: Like the previous case, but now the harmonic and the residual parts corresponded to one of the original sounds.

The answers of the voicer are presented in Table 24. Considering these results, it seems obvious that the harmonics/residual separation is an easy task for an organ-builder. This confirms the fact that a voicer is trained to hear noisy and harmonic components separately, as it is necessary to regulate them independently in a definite way through the voicing techniques. These results also suggest that the separation method employed is perceptually valid.

VOICER	Sounds	TYPE OF ANSWERS	
		Matched	Only Harmonic
	G1	0	1
	G2	1	0
	G3	1	0
	g1	1	0
	g2	1	0

Table 24.The answers of the organ-builder.

In order to extract condensed information on the perceptual validity of such a separation in the two groups of test participants, a nonparametric chi-squared significance test was performed for each sound and for a reduced number of answers (see Table 25). The procedure applied to calculate the Matched and Unmatched answers is as follows:

- **Matched=**Matched+Only harmonics+Only residual.
- **Unmatched**=Unmatched and uncoherent+Unmatched but coherent.

The results of the chi-square test with alpha=0.05 and chi-square critical value=3.84 are shown in Table 25.

a)

ACADEMY test participants	Sounds	TYPE OF ANSWERS		Chi2
		Matched	Unmatched	
	G1	23	17	0.9
	G2	26	14	3.6
	G3	30	10	10
	g1	32	8	14.4
	g2	35	5	22.5

b)

AKUSTIC test participants	Sounds	TYPE OF ANSWERS		Chi2
		Matched	Unmatched	
	G1	5	4	0.11
	G2	6	3	1
	G3	5	4	0.11
	g1	7	2	2.78
	g2	7	2	2.78

Table 25. Contingency tables of the separation test for the two groups of test participants.

For the group of 40 test participants (musicians) (Table 25a), some significant differences for the matched and unmatched answers are found, except for the sound G1. This suggests, then, that even though they are not trained in voicing techniques, the separation task was quite natural for many of these musically trained ears. By contrast, the level of significant difference between the matched and unmatched answers for the group of nine test participants collected in our department of acoustics is low. This may suggest, then, that untrained or nonmusical ears are not so familiar with the separation of pipe sounds into their harmonic and residual parts. Nevertheless, we must emphasize the fact that nine test participants is far too small a group to allow any general conclusion to be drawn.

In conclusion, in this test the separation of the residual and harmonic components of the flue organ pipe sounds presented was globally successful. Even though the few test participants in the second group did not show much agreement with the other group or with the voicer, we may infer that the separation task is perceptually natural and that the algorithm used preserves the residual and the harmonic components as perceived.

IV.4 - Other tests

The experiments described above notwithstanding, a number of other listening tests were conducted. As mentioned earlier, a study was done on the perception of the phonosymbol “hiss” and is reported in this document in Paper I. In addition, following the paradigms described in Chapter II, we developed two experiments dedicated to the construction of a specialized lexicon. These two experiments are presented in Papers II and III. Finally, it needs to be mentioned that a number of further listening tests could have been performed on the experimental pipe presented in the last chapter had time permitted.

Conclusions

We first reviewed and documented the art organ voicing as it is practiced in an organ workshop (see Chapter I). This shows us how refined voicers' techniques can be. The complexity of this task is also reflected in the physical description of the system itself. Most conspicuous difficulties concern the description of jet behavior. It is not surprising, then, to note that the primary part of voicing concerns the mouth area, the shape of which influences the jet development. A review of studies related specifically to the voicing or more generally to the sound quality or the physics of flue organ pipes, shows different approaches. These studies differ mainly in the model used for the pipe. The present study is unique in that it delineates a hybrid model of a pipe (see Experimental Pipe, Appendix I) and moreover integrates the voicer's inputs (his techniques and descriptions).

In order to establish a relation between the empirical knowledge of the voicer and the scientific field, we used and developed several tools and methods and designed experiments and listening tests. In this respect, this thesis is built upon a truly interdisciplinary approach, gathering knowledge and expertise from various domains of science such as acoustics, signal processing, experimental psychology, computer science and linguistics. This approach is necessary in order to explore such a domain as the sound quality of flue organ pipes, and this is also reflected in other sound-quality studies concerning other types of sounds (see Chapter II). We thus both proposed a methodology and found a number of results.

The voicer listens to the sound of a single pipe in an extremely detailed way. A voicer's ears perform a very fine analysis of the characteristics of the transient vs. stationary parts and periodic vs. stochastic parts (see Chapters III and IV). All these features are important in order to set the pipe to its final timbre, even if these features do not appear as such in the context of an audience (e.g. the turbulence noise decreases quickly with distance). It was found that in the case of a rating task concerning exhibited attributes, not only could objective parameters not be found to match subjective ones precisely, but ratings from a group of 22 participants and ratings of the voicer also did not match well. In our opinion, this is not surprising, considering the fact that our vocabulary for the description of sound is rather poor and lacking in precision. We thus put more effort into non-forced methods, such as free grouping and scaling (see Paper IV) and decided to create a list of verbal descriptors suitable for the description of flue organ pipe sounds. This list was created through a four-step process (see Paper II and III):

1. interviewing the voicer and making a first list of verbal attributes
2. making a first listening test and using qualitative analysis of subjects' comments (Paper II)
3. reformatting the initial list (see Appendix IV)
4. performing a factor analysis of a new listening test based on the revised list (Paper III).

Moreover, a description of the sound quality of flue organ pipes that is capable of being communicated can be established only if a list of descriptors is available together with analysis and recordings. This is illustrated by the multimedia document accompanying this thesis.

Perspectives and Future Work

It turned out to be very difficult to follow in a rational way the geometrical adjustments occurring during a "real" voicing session on a "real" pipe. For this reason a hybrid pipe was built. The main feature of this experimental pipe is to provide repeatability and "comparability" of different voicing steps. The mouth of this pipe is made so that the languid, the upper-lip and the lower-lip are built in interchangeable frames. The pipe was built in Plexiglas to ensure precise measurement of the relative position of the pieces and so that visualization via Schlieren technique and a subsequent qualitative analysis of the jet behavior could be made.

Three types of parameters appeared to be important in this framework. Each of these types can be represented in a corresponding space: a verbalization space (perception being verbalized), a voicing space (geometrical and pressure parameters) and a physical-signal oriented space (issued from signal analysis). A voicing session can be represented by a trajectory inside each of these parameter spaces. An attempt has been made to map together these three trajectories. In other words, we tried to answer the question, given one trajectory, is it possible to reconstruct the other two? With the experimental pipe, it is possible to synthesize a simplified voicing trajectory.

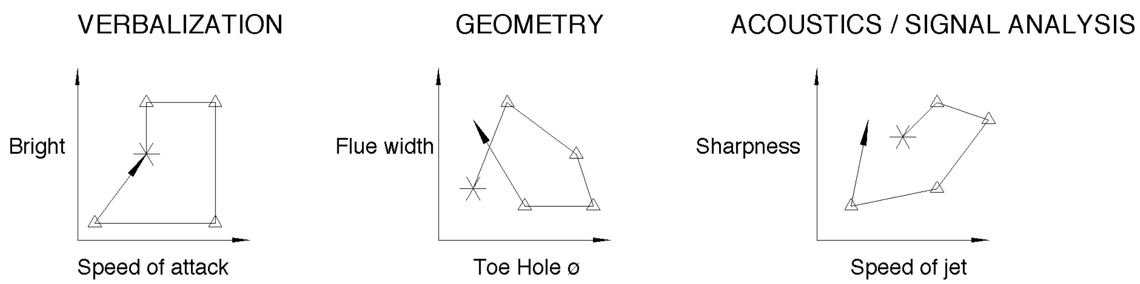


Figure 40. Three apparent types of description of a voicing trajectory.

Outcomes of this research work could be applied to actual voicing (see Figure 2) or to pedagogical requirements in an organ workshop.

We have managed to present a lexicon composed of descriptors in a hierarchy structured according to a set of classes of attributes. It is possible, moreover, to find hypothetical relationships between these attributes and signal characteristics of the stimuli, but at this time no firm correspondence has been established. This would represent a major challenge for future work. It is thought that the strategy of approaching physical parameters in a way that has to do with combining a verbal description and a requirement of precise signal characteristics is of great interest in a number of fields in acoustics ranging from noise control to physically informed musical sound synthesis or management of sound databases.

We think as well that the attempt we have presented here to create a standard of verbalization for a particular set of sounds could be extended to some other classes of sound.

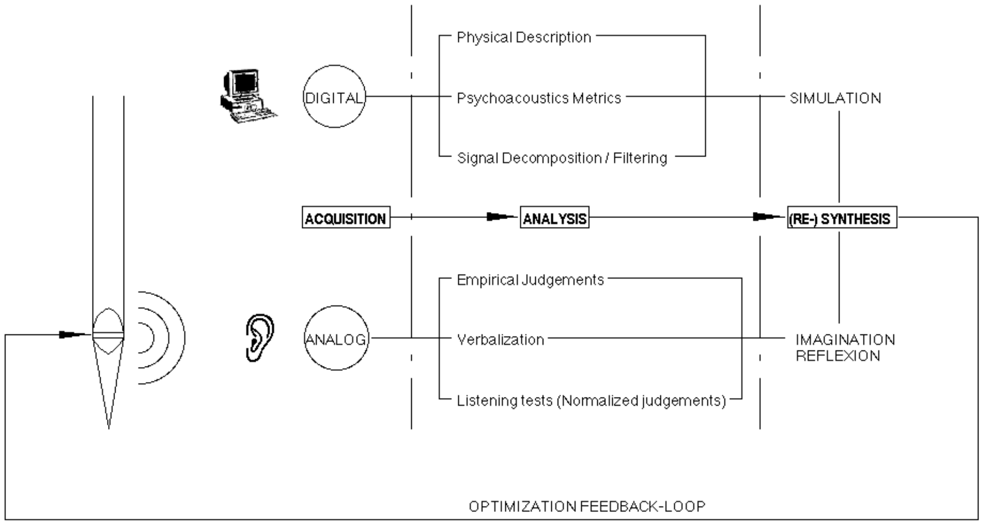


Figure 41. A possible integration of digital technologies in the voicing process.

ARTICLES

Paper I

Methods for an objective and subjective description of starting transients of some flue organ pipes – integrating the view of an organ-builder

Acustica 86(4) July/August 2000, 634-641, Special issue on wind instruments

Vincent Rioux
Chalmers Room Acoustics Group, Chalmers University of Technology,
Department of Applied Acoustics, SE-41296 Gothenburg, Sweden.

Summary

The transient sound from an organ pipe is a very important component in deciding the pipe's timbre. This paper addresses the issue of verbal description of transient sounds of flue organ pipes. This problem must be solved when, as in our case, one wishes to document, understand and possibly assist the work of an organ-builder in the process of voicing organ pipes. Therefore, we explain the procedure used to map the relationship between verbal descriptors and quantitative acoustic analyses. The study is restricted to single flue organ pipes (no chords). A procedure integrating both objective and subjective methods is presented in order to develop such a description. After an introduction on basic voicing principles, a list of verbal descriptors used by the voicer is presented and serves as a basis for listening tests and signal analysis. Several types of listening tests are described. Computer implementation of these tests is shown to be particularly useful. Signal analysis using time-frequency methods is employed for separating harmonics and for representing the transient components of the signal. Results are also reported on the perception of the speed of the transient and of the sound “hiss”.

Keywords

Musical acoustics, sound quality, flue organ pipes, voicing techniques, transient sounds, psychoacoustics, listening tests, *onomatopoeia*, signal processing, time-frequency analysis, harmonic/residual decomposition.

1. Introduction

This article presents some aspects of the work done in collaboration with expert organ-builder Munetaka Yokota. The work was triggered by a project concerning the reconstruction of a 17th century Baroque organ (in the Nya Örgryte church, Gothenburg, Sweden). A group of musicians and musicologists (the Gothenburg Art Center or "GOArt") decided on the construction of such an instrument by the organ workshop of Mr. Yokota. The Chalmers University of Technology was also asked to investigate several aspects of organ building (acoustics, fluid dynamics and metal properties). It was then possible to perform studies both in the organ workshop and in the church and to benefit from the knowledge of trained listeners during various conferences held at GOArt.

Organ pipes of Baroque organs generally display a special character at their onset. This onset, also called "speech" by organ-pipe voicers, is often qualified by *onomatopoeia* such as "chiff". The task of the voicer is to harmonize the sound of pipes relative to other pipes of the same or different registers (cf next section). Voicing techniques are used to tailor the onset, but due to the complexity of this task (for both technical and aesthetical reasons), these techniques are still under discussion within the world of organ building itself.

From the point of view of acoustics, several authors have studied transient phenomena of flue organ pipes. For example, Fletcher [1], Fabre [2] and Verge [3] carried out extensive theoretical studies, but there still remains the need to integrate this work into the practical knowledge of the voicer. Recently Castellengo [4] published a set of important observations on the different components of the so-called mouth-tones and pipe-tones. Closer to the area of organ building, Angster questioned the way organs should be documented [5] and proposed an acoustical description of several steps of voicing [6]. Nolle [7] also studied some voicing steps on an experimental pipe and was able to map a voicing "trajectory" with some verbal descriptors.

Throughout many of these publications, however, one may notice that organ builders are seldom consulted or not considered as the source of information for how to *describe* these sounds. Therefore, we wish to emphasise the necessary inclusion of the voicer's perspective in the effective characterisation of the pipe timbre. First objective and subjective methods are introduced. Then some results will be discussed.

2. Voicing techniques: an overview

Voicing techniques refer to geometrical changes made by the organ builder on each pipe of each *stop* (registers). Techniques have been developed which allow extremely fine variations in the sound quality of organ pipes. The voicing techniques are based on empirical (or trial-and-error) knowledge which has been accumulated, refined (and sometimes forgotten) over centuries. These techniques also went through changes according to acoustical and musical aesthetics of different time periods. Applying these principles, the voicer changes the geometry of the pipes according to the following desired transformations of the sound:

1. modification of the timbre:
- adjustment of the stationary part (relative levels of harmonics, noise content)
- adjustment of the transient/speech (speed, noise character, harmonic content)
2. adjustments for *loudness*,
3. tuning.

One must emphasise the fact that the sound production of flue organ pipes is dependent on many interconnected factors. These interdependencies make it difficult to delimit precisely where voicing begins and ends in the pipe-making process. One could for example integrate choices of body or foot geometry, or even the type of metal used for the pipe walls, as these choices will crucially influence further refinements. But in the following, the term "voicing" will be employed when it concerns geometrical modifications of the mouth and possibly of the toe-hole (size of bore) (Figure 1).

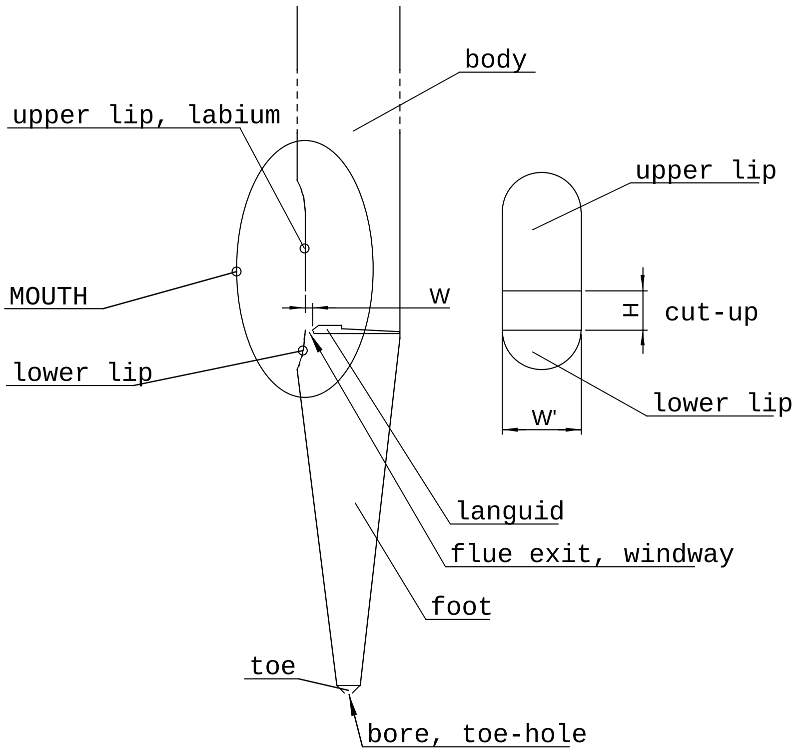


Figure 1. Parts of a flue organ pipe.

Voicers have designed a set of tools in order to modify the mouth area. Among many variables the following geometrical parameters are of primary importance:

- dimension of mouth opening (H/W' ratio)
- ratio of cut-up to wind-way width (H/W)
- relative position of the upper-lip to the middle of the flue exit
- proportion between the areas of the foot hole and the flue exit

These adjustments affect both the steady and transient waveforms. It is not the scope of this article however to present in detail how modifications of such geometrical parameters alter characteristics of the sound (for more details on this topics see [5, 6,7]). The static and dynamic pressures in the foot are also of great importance for the sound produced by the pipe. The onset of the pressure rise in the foot is directly correlated with the initial velocity of the jet and thus to the starting transient [8]. This foot pressure should then be regarded as an important component of voicing.

3. Objective description of starting transients

The physics of sound generation by flue organ pipes is very complex. The main problems lie in the description of the interaction of an unstable turbulent jet with an edge and an acoustic field. Currently, numerical simulations, though capable of producing somewhat realistic sounds [9,10], are still not able to reproduce the whole range of potential timbre with the extremely fine precision required by the voicing art. We emphasise therefore the study of sounds from real pipes.

3.1. Time-frequency analysis

In order to analyse the complex transients of flue organ pipes, we choose to work in the time-frequency domain, following the early works of Castellengo [11]. The spectrogram is a graphical representation of the short-time Fourier transform of the sounds, and used for a first investigation. Other transformations like wavelets or variants of the Wigner-Ville transform [12] are performed when a better resolution is needed. All these transformations may contain a model of the signal but do not take into account any perceptive model. For a normative representation of the sound, we use the cochleagram proposed by Lyon [13] and implemented by Slaney [14].

During the starting transient several competing phenomena are superimposed. The development of the periodic (or harmonic) part is of the order of 50 periods of the fundamental [15]. It is this part of the starting transient that has received the most attention in the existing literature. But in some cases, the smooth evolution of these harmonics is somehow perceptively "low rated", compared to the quick bursts of energy appearing first at the initial transient phase, which is of the order of 10 periods of the fundamental. A schematic of a typical transient of a flue organ pipe is presented in Figure 2 as a time-frequency plot where one can recognise various components of the onset.

Tonal bursts are produced by an impulsive excitation of the passive, inharmonic modes of the air-column (plane and higher non-plane modes). The amplitudes of these tones may be slightly affected by structural modes [16]. In parallel, it is most probable that an edge-tone [17] can easily develop during this initial part, as the jet is not yet forced to oscillate on an air-column mode and thus able to develop a strong hydrodynamic feedback. The components of the early part of the transients that do not belong to the harmonic set are referred to as "mouth-tones" and have been thoroughly discussed by Castellengo [4]. The wide band noise created by the turbulence and filtered by the air-column is important in the starting transient, since it is not yet masked by the strong periodic signal. It thus appears as a burst of noise.

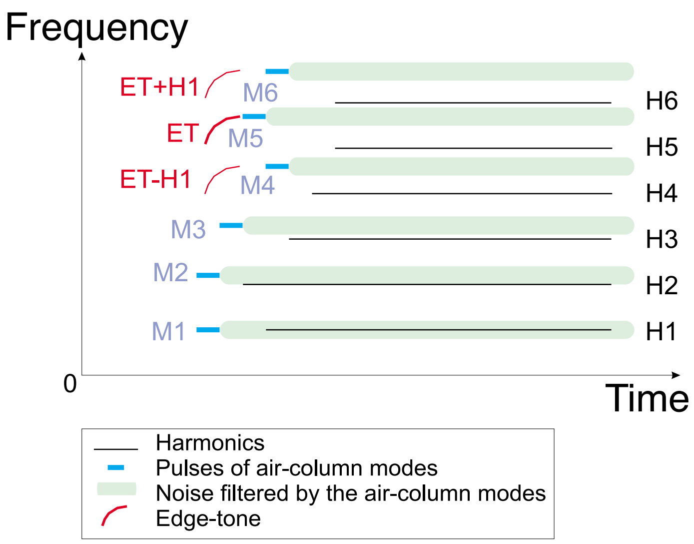


Figure 2.Time-frequency representation of a typical sound of a flue organ pipe. Straight equally spaced lines (H1 to H6) represent the six first harmonics. The thick and short segments (denoted M1 to M6) represent the modes of the impulsively excited air-column. The curved portion noted ET represent the transitory edge-tone whose frequency varies proportionally with the speed of the jet. This edge-tone may also be modulated by one harmonic (here, the fundamental). The thick, shaded and long segments represent the noise associated with the turbulence which is filtered by the modes of the air-column. Note that everything except the harmonics H1-Hn is called the residual in the following.

3.2. Harmonics / residual decomposition

As illustrated in Figure 3, a flue organ pipe sound is composed of two significant parts, a periodic (or harmonic) part and a residual part.
The periodic part contains the sum of the time-dependent harmonics, characterised by their arrival times, their rise-rates, and their relative amplitudes [18].

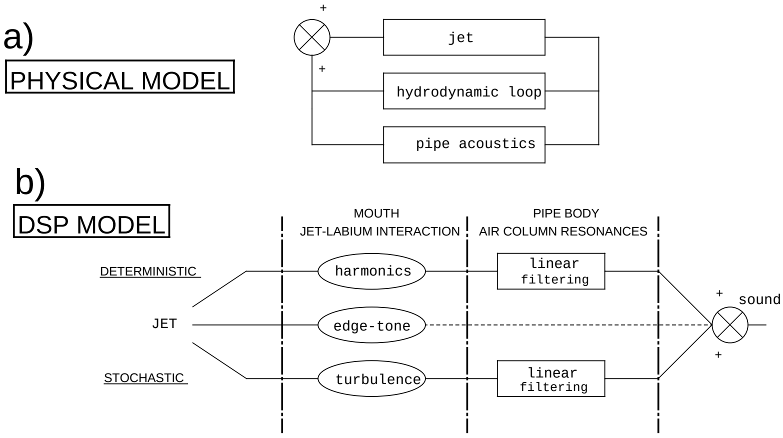


Figure 3. Sketch of two simplified models of a flue organ pipe. In a), a simplified physical model stresses the components of the feedback loop system. The jet is the source of energy which counterbalances the different terms of loss (vortex shedding, wall boundaries and vibration and radiation). In b), a signal processing model is presented. The jet oscillating back and forth around the labium is the source of two kinds of basic signal. The deterministic part is due to a non-linear interaction between the steady jet and the labium. The stochastic signal is due to the turbulent part of the jet.

As stated in the introduction, even though the periodic part is the core of the sound, giving it its pitch and basic timbre, the residual part (the bursts plus the stationary noise) is often underestimated but nevertheless recognised as giving each instrument its unique character. As a matter of fact, several verbal descriptors used by the voicer refer to this residual (see Table I).

Several authors [19, 20, 21, and 22] have proposed various methods to decompose musical signals. These methods work for a large class of signals possibly ”affected” by pitch changes and vibrato (i.e., both frequency and amplitude).

But, considering the relative uniform character of a single flue pipe sound, we preferred to implement a simple and robust method [23]. Harmonics are individually filtered with the same infinite impulse response (IIR) filter (high-pass or low-pass to obtain the periodic or residual respectively). The relative simplicity of the method makes it easy to automatically process several sounds with a known fundamental frequency. In our framework, this decomposition is used as a first step toward computation of normative parameters. It is also useful as a listening tool, as it helps untrained listeners to focus on certain components of the original sound. With this tool we are also able to check whether the voicer has developed a similar decomposition process when listening to the sound (see Section 5: Experiments).

4. Subjective evaluation: Listening tests and methods

Complex phenomena present in the transient are often verbally described using *onomatopoeia*, e.g., "buzz", "chiff", "ping" [7,4] or "cough", "hiss", "spitz”. The voicing terms used by organ builder Munetaka Yokota are shown in Table I.

Note that for these descriptors, only the first part (e.g. "hi" in "hiss") is likely to imitate a specific character of the transient. The consonant part would thus describe the noise bursts and the vowel part would emphasise the frequency region of the prominent tonal bursts. An investigation on the term "hiss" is presented in section 5.

Table I. List of descriptors used during the voicing.
• indicates the 'presence of'
o emphasises the 'approximate' character of the associated physical parameter
The descriptors in grey concerns mainly the residual part (transient or steady).

Part of the sound	Descriptors	Onoma-topoeia	Antonym	Physical parameters
speech or transient	Chiff	•		
	Cough	•		
	Hiss	•		
	Kiss	•		
	Soft Vs strong		•	•
	Slow Vs fast		•	•
	Short Vs long		•	•
	Pitch (e.g. octave)			•
steady state	Amount of fundamental			•
	Amount of quint			•
	Amount of octave			•
	Stringy Vs fluty		•	o
	Round Vs sharp		•	o
	Full Vs thin		•	
	Light Vs heavy		•	
	Nasal			o
	Breathy			o
	Dirt Vs clean		•	
	Free Vs tight		•	
	Floating/spatial		•	
general impression	Intense			
	Sandy			
	Sweet			
	Hollow			
proportion between	Fundamental			•
	Speech and steady state			
	Noise and musical tone			o

To possibly standardise such a list of verbal descriptors, it is necessary to relate such idiosyncratic items to the actual physical behaviour of the organ pipe, mirrored in the various components of the sound. In order to verify the matching between various objective parameters and the subjective description of the organ builder, we designed several types of tests and constructed a computer program to administer the tests and collect data [24, 25].

Sound samples were digitally recorded (DAT recorder) at 30 cm from the mouth of the pipes and played by a PC sound-card through headphones. Musicians and musicologists attending the GOArt conferences were asked to participate in these tests. Sounds used for a particular test were always chosen to be of the same pitch and same *loudness*. For each test, loudness of each stimulus was perceptively equalised across the set, by the test-leader, using a graphical interface like the one shown in figure 4.

The decay part of sounds was normalised by fading-out with a constant slope (so that only the onset could be heard) and harmonics were filtered out.

Three kinds of test methods were used depending on the way stimuli were compared to each other (by groups, by pairs or individually). A specific test was also designed for assessing basic perceptive features of harmonic extraction.

For all these tests, special care was taken to allow the subject to write comments about any stimuli at any time.

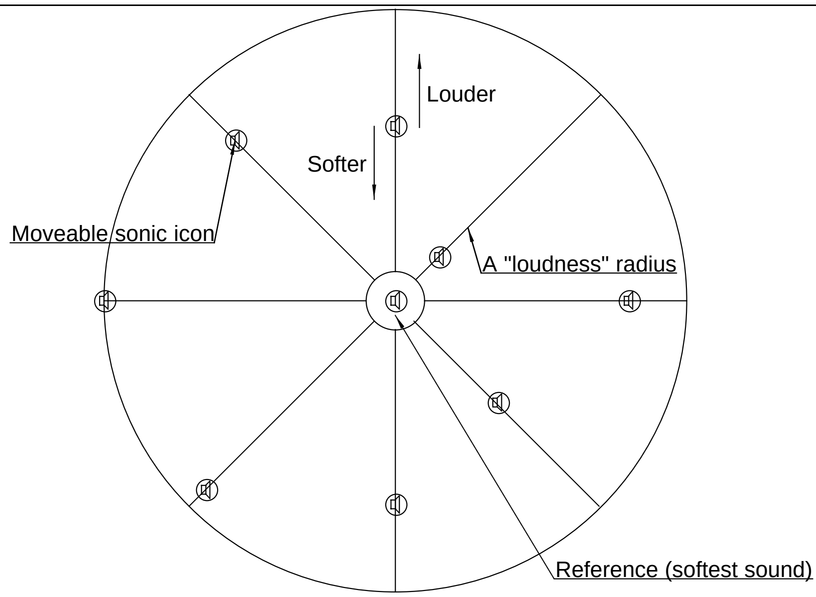


Figure 4. Sketch showing the user interface for loudness equalisation. Each sound is represented as a *sonic icon* and can be moved along a radius. The user first determines which of the sounds is the weakest and sets it as a reference in the middle of the circle. The other sounds can then be adjusted to the same loudness by changing their positions on their radii.

4.1. Grouping test

A grouping test was implemented in which subjects were asked to place *sonic icons* on a 2D plane (see Figure 5). The distance between two icons is interpreted as a measure of dissimilarity. Subjects can deliberately group sounds by selecting them, thus defining categories. This approach is particularly well suited to the beginning of a sequence of tests, as it allows the subjects to familiarise themselves with the sounds and thus decrease the precedence effect.

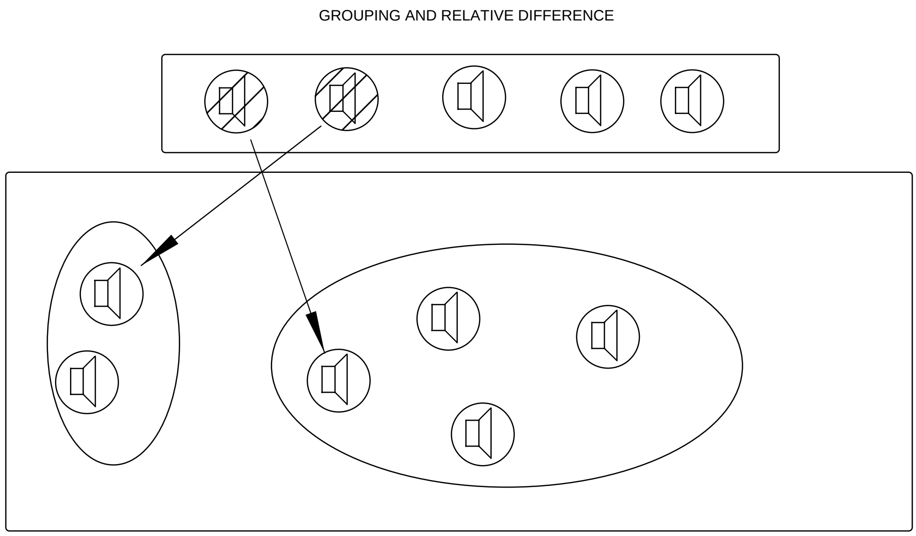


Figure 5. Sketch showing basic outline of the grouping test user interface. The arrows symbolise "drag and drop".

4.2. Pair comparison tests

Two types of pair comparisons were used. These tests differed by the way subjects were asked to judge the dissimilarity of the presented pair of stimuli [26]. First, in a "difference test", the subjects were asked to rate the global difference between all possible pairs of stimuli taken from the set (see Figure 6, top).

Second, a "scaling test" (see Figure 6, bottom), can be regarded as a more detailed version of the former. The subject is asked to focus on a particular descriptor of the sounds (e.g., "hiss") and to rate the relative perceived difference between the pair of sounds presented. Note that in this particular case, the comparison was always made against a reference, the number of combinations being then reduced to the number of sounds of the set. The reference was the sound of a historical pipe, and the other sounds were modern copies of this historical pipe with slight, but deliberate differences in voicing. In that case, the reference had a natural justification.

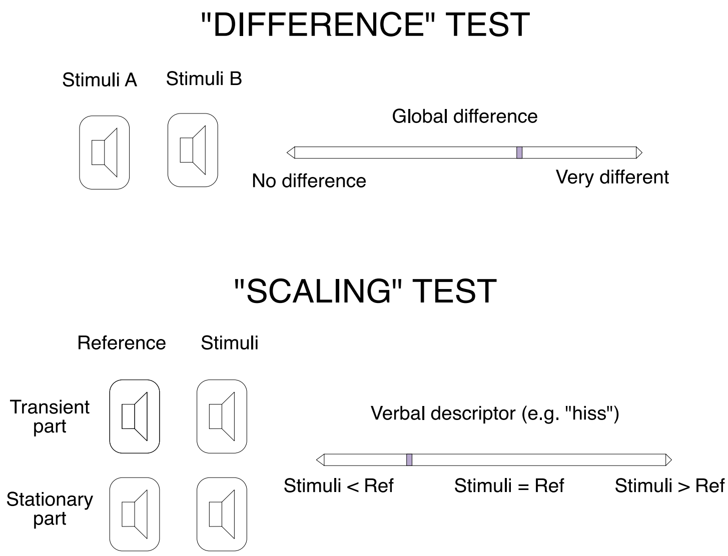


Figure 6. Sketch of the user interface of two pair comparison tests. On top, a "global difference" test. On bottom, a "scaling" test is shown where several *attributes* are to be rated. This test includes a reference stimulus.

4.3. Multiple comparison test

Multiple comparison tests (see Figure 7) were also implemented and carried out in the same session as the pair comparison tests. When a test subject starts a pair comparison, it takes a certain (unknown) amount of time before the subject is accustomed to the span of the scales in relation to the given set of sounds. This effect, referred to as the ”*learning effect*”, is much reduced in a multiple comparison test where all the sounds are presented at once. The drawback however is that the task becomes more complex for the subject.

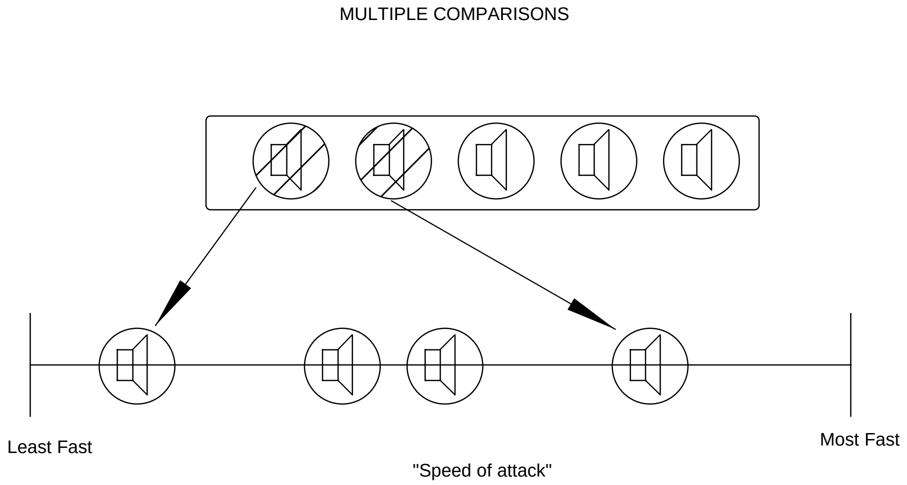


Figure 7. Sketch of the user interface of a multiple comparison test. Each "*sonic icon*" contained in the rectangular upper box is to be placed on a scale.

4.4. Separation test

Finally, a new test was devised to check the validity of our harmonic/residual decomposition. All the sounds to be tested are pre-processed and their harmonic and residual parts were stored separately. For each sound appearing in the "original" box, the subject must find the correct harmonic and residual parts from the sets placed in the lower boxes (see Figure 8). Once a choice has been made, a new “total” sound is introduced and the subject is asked to find its corresponding residual and harmonic parts.

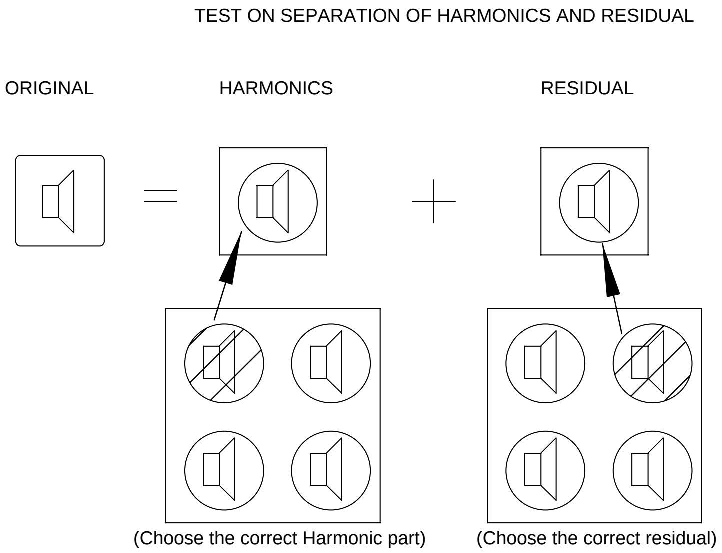


Figure 8. Sketch of the user interface of the ”separation test”.

5. Experiments

Two test-sessions were made during a conference and a symposium organised by the Gothenburg Organ Art Center. It was possible to gather between 20 and 40 participants for each session in less than a week, thanks to the flexibility of the computer test program. As an illustration, sounds that were presented to the subjects during these listening tests are available online [25].

5.1. First session

The first exploratory session concerned general aspects of flue organ pipe sounds [27]. A historical pipe (principal 4' G sharp) served as a reference and several modern copies were made. Sounds corresponding to successive voicing steps of one of the copies were recorded in the organ workshop. The nomenclature is as follows: *s0* is the sound for the original pipe, *s1-s6* refer to sounds of the copied pipes through subsequent steps. A grouping test, a scaling test and a global difference test were presented to 40 subjects. Concerning the transient part, test participants were asked to focus on the perceived "speed of the starting transient" and to compare it to the reference during a scaling test (see Figure 9).

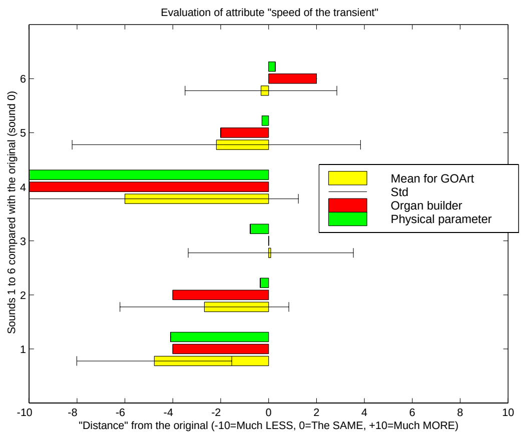


Figure 9. Comparison of perceptual ratings for the *attribute* "speed of the transient" through different steps of voicing. The "speed of transient" characterises the duration of the onset. If the duration is short, the "speed of the transient" is high. Sounds are ordered along the vertical axis. The horizontal axis represents the amount of difference for one sound compared to the reference (point zero). Two ratings are presented: one for musically trained subjects (mean and standard deviation for the GOArt group) and one for the voicer. A physical parameter computed from the duration of the envelope of the harmonic part is also shown.

The global difference test was analysed by means of multidimensional scaling techniques [28, 29]. Results from this test are projected on a two-dimensional space (Figure 10). The two dimensions explain 38% of variance which is a moderate figure but notice that the answers of the voicer and of the musicians correspond closely (i.e., the dotted lines are rather short and never cross). Dimension 1 is correlated with the ”speed of transient” attribute. Although the amount of noise (“noisy” character) is a possible choice, by lack of a suitable physical attribute an interpretation for dimension 2 is still out of reach.

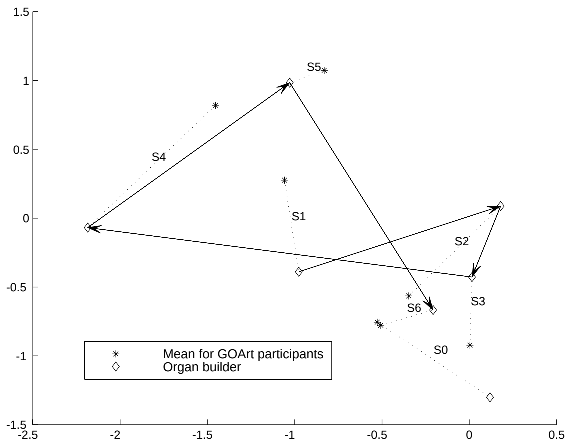


Figure 10. MDS representation of the ”global difference” test. S0 is the reference sound. S1-S6 are different sounds recorded during the voicing process. The distance between the voicer and the other participants is plotted in dotted lines. The arrows represent the ”perceptual” path of the voicing process.

This first test session focused on a reference pipe (a historical one) and its modern copies. This first set of tests involved many steps of voicing and several attributes. The next session consists of studying several “finished” pipes that have a unique character described by few attributes.

5.2. Second session

In a second session, we thus chose to compare three pipes of the same pitch but of different stops, and with slightly different transients. The best verbal descriptor that could explain these differences was said to be ”hiss”, according to the organ-builder.

Pipes were digitally recorded in an anechoic chamber. Ideally, the recordings should convey the same sounds as they were received by the organ builder while voicing. In practice, we used monophonic recordings made at 30 cm from the mouth of the pipe (see Figure 11). Sounds are named G1,G2 and G3 [25].

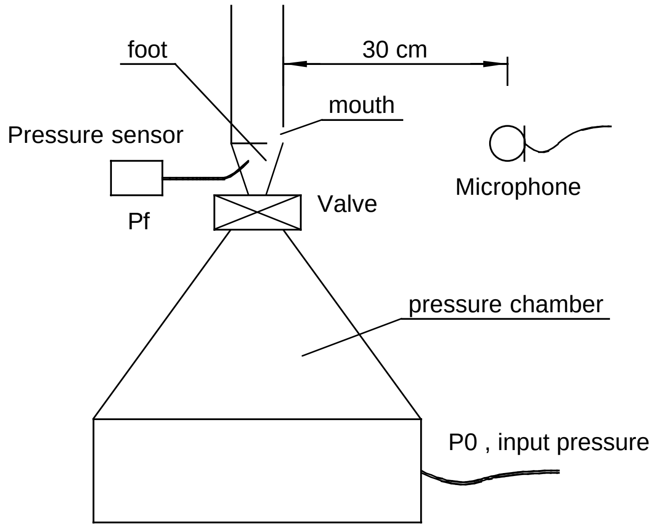


Figure 11. Experimental set-up used in an anechoic chamber for the recordings.

In order to check the validity and generality of using the *onomatopoeia* “hiss” we compared results of subjective tests with time-frequency analysis [30]. Listening tests were based on pair comparisons and multiple comparisons. Subjects were asked to rate both the "total" sounds and their associated residuals. While the answers of the voicer were consistent through all varieties of tests and sounds (residuals as well as complete sounds), the answers of musically trained listeners were less consistent but closer to the voicer’s answers when listening to the residual parts alone. This, combined with the results of a ”harmonics/residual separation test” leads us to conclude that this separation is perceptually helpful and valid and that it helps the listener to focus on the starting transient bursts.

This study also proved that onomatopoetic descriptors such as ”hiss” can be useful to describe particular aspects of the starting transients. The same tests were performed with and without the possibility of listening to a recorded version of the voicer saying the word ”hiss”. The answers of the subjects were closer to the voicer’s when this last possibility was offered. In fact, a comparison of the time-frequency analysis of the sounds pronounced by the voicer (Figure 12.b) and by a pipe (Figure 12.a) clearly shows that the same prominent features appear for the starting transient of the pipe and for the “hi-“ part of “hiss”.

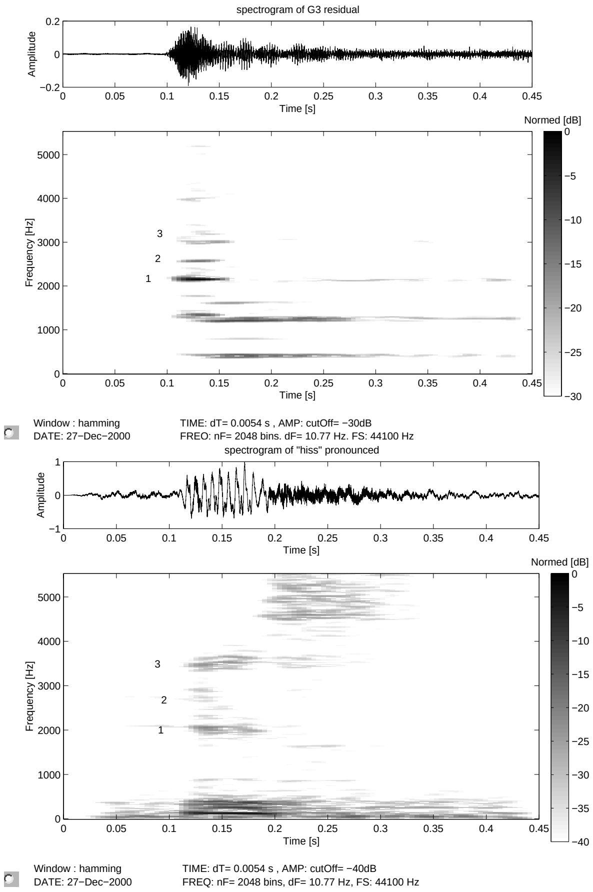


Figure 12. Comparison of time-frequency features of the "hiss" descriptor. a) Modal frequency distribution [12] of the transient of a pipe (referred to as G3, F0=412 Hz). b) Spectrogram [13,14] of the word "hiss" pronounced by the organ builder. The three numbered components are common to the two spectra. This illustrates the utility and appropriateness of such onomatopoetic descriptors.

6. Conclusion

In this paper a number of tools and methods for objective and subjective description of flue organ pipe sounds are presented. The methods are shown to be useful for the assessment of starting transients of flue organ pipes. The interdisciplinary character of the approach is underscored by the combined use of results coming from formal listening tests and systematic analyses of sound signals. A computer-based program was designed and proved to be particularly practical and flexible (a large number of tests could be executed in a short time). In addition, harmonics/residual separation coupled with time-frequency analysis is shown to be of major interest when studying such transitory phenomena. Use of onomatopoeia for the description of transitory events deserves more attention as clear evidence of its potential use is demonstrated in the case of the “hiss” sound.

7. Future work

Generality of use of onomatopoeia for description of the onset of flue organ pipes must be systematically checked for a larger set of pipes. The list of verbal descriptors will also benefit from a generalisation so as to produce a standardised lexicon. As we stated in section 2 on voicing techniques, geometrical parameters were not mapped with perceptual or signal parameters. In order to explore this mapping, we wish to study an experimental pipe with an adjustable mouth area.

Furthermore listening tests methods should be refined so as to e.g. cast both scaling and grouping tests and integrate loudness equalisation of stimuli by each subjects.

9 Acknowledgements

This work was done within the framework of the project ”Changing processes in North European Organ Art 1600-1970 – Integrated Studies on Performance Practice and Instrument”, taking place at GOArt, Gothenburg University, Sweden. The work was supported by the Bank of Sweden Tercentenary Fund, Chalmers University of Technology and Gothenburg University.

The author particularly wishes to thank organ-builder Munetaka Yokota for his collaboration, enthusiasm and constant support. Thanks are also due to Mendel Kleiner, Daniel Västfjäll and Matthias Scholz as well as to all the participants, musicians and musicologists who took part in the listening tests. The author is grateful to the reviewers for their comments on the earlier draft.

8. References

[1] Fletcher, N.H. : Transients in the speech of flue organ pipes – A theoretical study. Acustica **34** (1976) 224-233.
[2] Fabre, B. : La production du son dans les instruments à embouchure de flûte: Modèle aéro-acoustique pour la simulation temporelle. Thèse de l’université du Maine (1992) 152-179

[3] Verge, M.P., Fabre, B., Mahu, W.E.A., Hirschberg, A., Van Hassel, R.R., Wijnands, A.P.J., de Vries, J.J. and Hogendoorn, C.J. : Jet formation and jet velocity fluctuations in a flue organ pipe. J. Acoust. Soc. Am. **95** (1994) 1119-1132.
[4] Castellengo, M. : Acoustical Analysis of Initial Transients in Flute Like Instruments. Acta acustica **85** (1999)
[5] Angster, J. and Miklos, A. : Documentation of the Sound of a Historical Pipe Organ. Applied Acoustics **46** (1995) 61-82.
[6] Angster, J. : Effect of voicing steps on the stationary spectrum and attack transient of a flue organ pipe. International Symposium on Musical Acoustics. Book 2. Edinburgh (1997) 285-294.
[7] Nolle, A. W. : Some voicing adjustments of flue organ pipes. J. Acoust. Soc. Am. **66** (1979) 1612-1626.
[8] Nolle, A.W. and Finch, T.L. : Starting transients of flue organ pipes in relation to pressure rise time. J. Acoust. Soc. Am. **91** (1992) 2190-2202.
[9] Verge, M.P., Hirschberg, A. and Caussé, R. : Sound production in recorderlike instruments. II. A simulation model. J. Acoust. Soc. Am. **101** (1997) 2925-2939.
[10] Kropp, W., Svensson, P. and Kleiner, M. : Time domain formulation of the sound generation mechanism of flue organ pipes, with auralization. Proc. of the Nordic Acoustical Meeting. Helsinki, Finland. (1996). 415-422.
[11] Castellengo, M. : Paramètres sensibles d'un tuyau d'orgue à embouchure de flûte. Bulletin du Groupe d'Acoustique Musicale. Paris **42** (1969).
[12] Pielemeier, W. and Wakefield, G. : A high-resolution time-frequency representation for musical instrument signals. J. Acoust. Soc. Am. **99** (1996) 2382-2397.
[13] Lyon, R. F. and Carver, M. : An analog Electronic Cochlea. IEEE Transactions on Acoust., Speech and Signal Processing. Vol. **36** no 7. (July 1988) 1119-1134.
[14] Slaney, M. Auditory Toolbox: A Matlab Toolbox for Auditory Modeling Work. http://www.interval.com/~malcolm/pubs.html (1994)
[15] Keeler, J.S. : The attack transients of organ pipes. IEEE Trans. AU. **20** (1972) 378.
[16] Kob, M. : Influence of wall vibrations on the transient sound of a flue organ pipe. Acustica **86** (2000) 642-648.
[17] Elder, S.A. : Edgetones versus pipetones. J. Acoust. Soc. Am. **64** (1978). 1721-1723
[18] Pollard, H.F. and Jansson, E.V. : Analysis and Assessment of Musical Starting Transients. Acustica **51** (1982) 249-262.
[19] McAulay, R. J., and Quatieri, T. F. : Speech Analysis/Synthesis Based on a Sinusoidal Representation. IEEE transactions on Acoustics, Speech, and Signal Processing **34** (1986) 744-754.
[20] Serra, X. and Smith III, J. : Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic plus Stochastic Decomposition. Computer Music Journal **14** (1990) 12-24.
[21] Fitz, K. and Haken, L. : Sinusoidal Modeling and Manipulation Using Lemur. Computer Music Journal **20** (1996) 44-59.
[22] Ding, Y. and Qian, X. : Processing of Musical Tones Using a Combined Quadratic Polynomial-Phase Sinusoid and Residual (QUASAR) SIgnal Model. Journal of the Audio Engineering Society **45** (1998) 571-584.
[23] Rioux, V., Kleiner, M. and Yokota, M. : Hamonics + noise extraction: a tool for the description of flue organ pipe sounds. Proc. of the Nordic Acoustical Meeting. Stockholm (1998).

[24] Scholz, M. Rioux, V., Västfjäll, D. and Kleiner, M. : Psychat---A program for psychoacoustical tests. CD of Proceedings of Forum Acusticum. Berlin, Germany (1999)
[25] Sound files and analyses of stimuli used in this paper, as well as a set of Matlab routines for sound analysis can be found online at: http://www.ta.chalmers.se/HomePages/Vincent/index.html
[26] Guilford, J. Psychometric Methods. Mc Graw.Hill. New-York (1954) 154-176.
[27] Rioux, V., Yokota, M., Västfjäll, D., Scholz, M. and Kleiner, M. : Preliminary study of an organ builder's perception of a flue pipe sound. Proceedings of the International Symposium on Musical Acoustics. Leavenworth (1998)
[28] SPSS Statistical Analysis software http://www.spss.com Chicago USA.
[29] PC-MDS Multidimensional Statistics Package. Brigham Young University. Provo, Utah USA.
[30] Rioux, V., Västfjäll, D., Yokota, M. and Kleiner, M. : Noise Quality of Transient Sounds: Perception of ”hiss” in a Flue Organ Pipe. CD of Proceedings of Forum Acusticum. Berlin, Germany (1999)

Paper II

PAPER II

Verbal Descriptions of Flue Organ Pipe Sounds (I): A Qualitative Analysis

Accepted for publication in Musicae Scientae (2001)

Vincent Rioux, Daniel Västfäll

Chalmers Room Acoustics Group, Department of Applied Acoustics,
Chalmers University of Technology

Abstract

This paper presents a lexicon of verbal descriptions of flue organ pipe sounds derived from organ-builder expertise. The purpose of this lexicon is primarily to enhance communication between and among organ builders, acousticians and musicians. A qualitative analysis was made of data collected from a listening test in which a number of specialists on organ performance, music theory, history and building took part. The 40 participants made ratings and gave comments to ten recordings of pipe sounds through the voicing process. By comparing an analysis of participants’ comments with a primary list of descriptors defined by an organ builder, an attempt is made to extend this primary list to a more standard lexicon. A temporary classification of these descriptors is proposed and is reviewed in a companion paper oriented towards a quantitative analysis of this type of data.

1. INTRODUCTION

This work is part of a current research program held between Chalmers University of Technology and Göteborg Organ Art Center (GOArt). Its aim is to develop methods of investigation adapted to the ”real” conditions of the secular organ-building craft in order to optimize and develop empirical techniques used for the voicing process of flue organ pipes. Following this paradigm, it is logical to consider the organ builder as the main source of information concerning voicing techniques and also as a primary reference for perceptual judgments of pipe *sound quality* (Rioux 1999). The main question within this framework is: *How and why do **historical** pipes sound different from "modern" ones?*

Before answering such a question, it is necessary to properly define the basic characteristics of this type of sounds. We thus propose here a procedure to develop a lexicon adapted to the description of flue organ pipe sounds.

It may be worthwhile to state that such lexicons defining aspects of perception have successfully been elaborated in the field of food industry, for example (Noble 1987). Whereas in acoustics, 30 years of research on timbral aspects have not converged to firm agreement. The problem of eliciting verbal descriptors has been addressed by, for example, Bismarck (1974b) and Samoylenko et al. (1996).

In parallel, studies concerning the essential attributes of timbre of musical sound (Grey 1977 or McAdams 1999) or non-musical sound (Zwicker 1999, Moore 1997, Bismarck 1974a) remain too general to be directly applied to the special process of organ voicing. In organ building, working on timbral aspects is essential but often avoided in treatises because of its complexity. This knowledge is supposed to be transmitted by a master-to-disciple procedure and/or by personal experience. Some builders simply refuse to describe sounds and some authors (Pelto 1995, Monette 1999, Nolle 1979) employ verbal descriptors but generally not founded on an objective (agreed) ground.

It thus seems reasonable to postulate that only very focused studies on a certain type of source (here, flue organ pipes) under certain conditions (here, several steps of voicing) can lead to a precise description of a limited sound collection.

2. VOICING TECHNIQUES

According to Rioux (1999a; 1999b), the voicing techniques refer to the geometrical transformations performed by the organ builder on each pipe of each *stop* (register). These techniques have been developed to change the sound quality of organ pipes with an extremely fine precision. They are based on trial-and-error methods which have been accumulated, refined (and sometimes forgotten) over centuries. Applying these principles, the voicer changes the geometry of the pipes according to the following desired transformations of the sound (Rioux 1999a):

- (pitch) tuning
- loudness adjustments
- modification of the timbre
- adjustment of the stationary part (relative levels of harmonics, *noise* content)
- adjustment of the transient or speech part (speed, noise content, un/voiced sound, harmonics rise)

These geometrical transformations mainly affect the so-called "mouth area" shown in Figure 1.

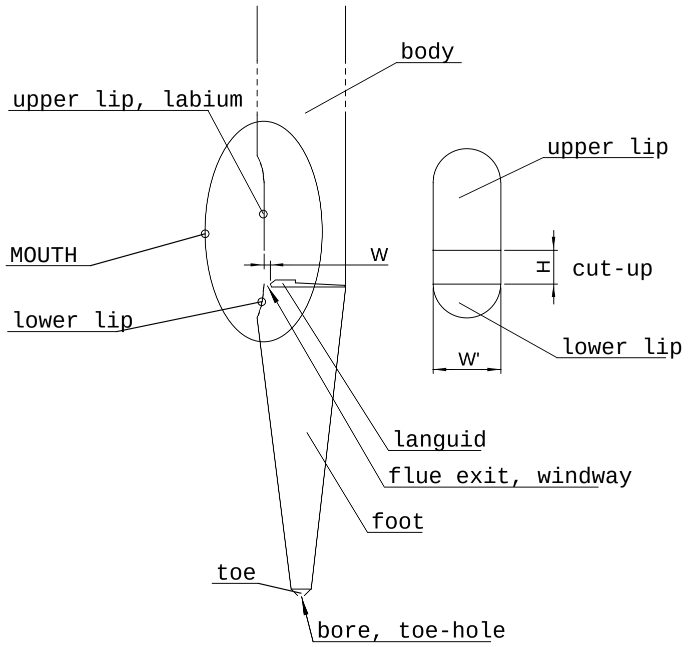


Figure 1. Schematic flue organ pipe

The current paper describes qualitative judgments of the perceived sound quality of ten different pipes during the voicing process. Earlier an analysis has been performed of quantitative ratings of these pipes (Rioux et al. 1998). The current paper extends the results from this previous research. Below, a description of the selection and recording of the ten pipes is given.

3. STIMULI SELECTION AND RECORDING

The following summary is taken from Rioux et al. (1998). A 4-foot G-sharp pipe from a "principal" stop was taken out of an organ made in the church of Jonsered (Gothenburg, Sweden) in 1783 by Perh Schiörlin. Three copies of each pipe were made while varying the thickness of the wall, amount of tin-lead alloy, or presence of nicking (small cuts made on the languid edge). The main goal from the voicer’s point of view was to try to make each copy sound as close to the original, regardless of the difference in material or the construction of the pipes, and thus to use the many resources of the voicing techniques. The pipes operated on a voicing bench were recorded (DAT Sony TCD D-10) at 20 cm from the pipe's mouth (close to the voicer’s ear). Ten sounds were recorded with the same electret microphone (Panasonic wm-063Y epoxy) but with an added Bruel & Kjäer wind-screen for sounds 7 to 9 in order to avoid saturation. All sounds were equalized to the same perceptual *loudness* so that timbral aspects could be extracted directly.

The ten final sounds are presented in *Table 1.*

Table 1. The set of stimuli is composed of 10 sounds. The first sound comes from an original pipe. Seven sounds were obtained from a copy while being voiced. Two copies with wall material changes provided two other stimuli.

Nomenclature	Pipe		Desired effect	Comments / process
sound 0	#1	Original pipe		
sound 1	#2		Pitch tuning	length was cut
sound 2	#2	Voicing	Fastening the speech	the languid is knocked down
sound 3	#2	Steps of a	Equilibrate the harmonic content of the steady part	cut-up raised
sound 4	#2	Geometrically	Increase loudness	opening up the windway
sound 5	#2	Reconstructed	Correction/compensation	windway down + languid lifted
sound 6	#2	Pipe	Final adjustment / cleaning	windway down + languid lifted
sound 7	#2		Final adjustment / cleaning	
sound 8	#3		Thick wall copy	
sound 9	#4		Nicked copy	

While following the work of the organ builder, interviews, recordings and notes concerning different kinds of pipe construction (from scratch to finalization) were compiled (Rioux et al. 1998; Rioux 1999a; 1999b). This first exploratory phase led to the discrimination of a large number of verbal descriptors. The list of descriptors was used for the analysis and interpretation of comments in the current study *(see Table 2)*.

Table 2. Original list of subjective descriptors used by the organ builder, and frequencies of appearance within subjects and comments. Descriptors are classified according to the sound part they refer to and categorized according to several parameters. These parameters can also be signal-based. "Noisy" refers to stochastic components. "Tonal" refers to periodic components.

Sound parts	**Verbal descriptors**	**Categorization parameters**				**Frequencies**
class	Packed in categories	Noisy	Tonal	Onoma-topoeia	Antonym	Occurrences
	Chiff					5
speech	Cough					-
	Hiss					-
or	Kiss					-
	Soft vs. strong					5
transient	Slow vs. fast					2
	Short vs. long					1
	Pitch (e.g. octave)					1
	Amount of fundamental					7
	Amount of quint					2
	Amount of octave					-
	Stringy vs. fluty					-
	Round vs. sharp					1
	Full vs. thin					1
steady state	Light vs. heavy					3
	Free vs. tight					9
	Floating/spatial vs. oppressive					-
	Dirt vs. clean					2
	Nasal					3
	Breathy					2
	Intense					3
general	Sandy					-
impression	Sweet					3
	Hollow					2
	Fundamental and overtones					6
proportions	Speech and steady state					5
	Noise and harmonic tone					3
Total Frequencies						66
	Indicates the presence of the parameter					
	Assumes the presence of the parameter					

4. METHOD

4.1. Participants and data collection

Forty trained listeners (musicians, musicologists, organ builders, specialists of organ performance, music theory, history and building) participated in a listening test. The test was performed during an annual research conference on different aspects on organ music and organ building in Gothenburg. The main purpose of the study was to compare the acoustical quality of different individual organ pipes during different steps of the voicing procedure. Ratings were performed of the dissimilarity of the different pipes on different descriptors given by the organ builder. These data are reported elsewhere (Rioux et al. 1998; Rioux 2000a). Additionally, participants were encouraged to give their own written description of each organ pipe in comparison to a reference pipe (the "original"). Tests were administered on a computer and at any time, subjects could add free comments in a text box (Scholz et al. 1999, Rioux 2000b). All sound samples were presented to participants in a randomized order. In all, 26 of 40 participants made at least one comment for one of the pipes.

4.2. Qualitative analysis

Verbal data were analyzed following the paradigms exposed by Denzin et al. . An alternative approach such as the "Repertory Grid Technique" can be found in Berg et al. (1999). It is noteworthy that all the descriptors were collected during a multiphase listening test ranging from free categorization (where subjects must elicit their own constructs) to semantic differential (where subjects had to focus on provided constructs).

Data were analyzed in three ways:

A) Correspondence to the original list

A list of subjective descriptors of different organ pipe sounds was assembled under the supervision of an organ builder. The descriptors in this list *(cf. Table 2)* comprise terms frequently used by the organ builder. Four main classes of descriptors were found. They relate to the interconnection of different time segments of a sound, namely, the transient part (or speech) and the stationary part. For musical instruments such as the organ, the decay part is often of very little perceptual importance, since the reverberation of the room would hinder the perception of its characteristics.

These classes may not be mutually exclusive (the same descriptor can sometimes be applied in two classes) but in this original lexicon they appear to be so. In order to limit the scope of the list, descriptors were chosen not to relate to any conveyed emotions, affects, judgment or taste.

The four classes are described below:

1. *Transient*. This class focuses on the early part of the sound, the onset or transient, where a large amount of energy is released and builds up to the steady state. The transient class is divided into several categories such as *onomatopoeia*. These words are sound-like in nature, for example *chiff, cough, hiss*, and *kiss*. Recent findings indicate that there is a direct acoustical link between the

frequencies used when pronouncing the early part of the word "hiss" and the organ pipe that possesses the perceptual quality of "hiss" (Rioux 2000a). Furthermore, several *antonyms* describing the transient behavior of the pipe can be formulated. Examples of such words are: *soft-strong, slow-fast, short-long*. These descriptors refer to both the physical (slow-fast) and perceptual (e.g. soft-strong) qualities of the transient. An additional indication of both the physical and perceptual character of the transient is its *pitch,* which can be conceived as either physical (frequency of the tone) or psychological (cognitive representation of the musical scale) (Krumhansl 1990).

2. *Steady state* concerns the part of the sound between the onset and the decay. Again, the organ builder distinguishes several categories of the steady state. The first category concerns the "physical amount of": *Fundamental, Quint, and Octave*. Antonyms include *Stringy-fluty, Round-sharp, Full-thin, Light-heavy, Dirty-clear, Free-tight, Floating/spatial-oppressive*. Also a category with reference to physiology and especially the voice (actually one pipe in the organ is called "vox humana", the voice of man) can be formulated. Terms of this category are *nasal* and *breathy*.

3. *General impression*. This category refers to a global quality of a sound. Included terms often entail reference to nearly emotional qualities *(intense, sweet*) or analogies to sensations (*sandy, hollow*).

4. The voicer also describes the *proportion* between qualities of the sound, such as the proportion between *fundamental and overtone, speech and steady state,* and *noise and musical tone*.

In a first analysis, subjects' comments were coded according to this list. A table of frequencies of appearance of each adjective was built. Only words that had a correspondence to a word in the list were included.

B) Extension of the original list

Subjects' comments were coded according to content by word or meaning (whole sentence) and categorized according to similarity. Existing categories as well as new categories were used or created.

C) Summaries for each sound

Comments for each sound were analyzed and summarized, to assemble the subjective character of each sound in the voicing process in comparison to the original pipe.

5. RESULTS AND DISCUSSION

Correspondence to the original list. In all, 220 words or sentences were coded and collected from the comments of 26 participants on 10 sounds. These 220 units were categorized according to the classes of the original list *(see Table 2)*. Frequencies were then obtained from the occurrences of words or sentences. Words or sentences that did not fit into any category were omitted. In total, 66 out of 220 units (30%) were categorized. In Table 2, the original list of descriptors and related frequencies for participants' comments is displayed.

As shown in Table 2, not all descriptors in the original list are used by the participants to describe the 10 pipes. However, descriptors with reference to the physical description are frequently used (e.g. fundamental, proportion between fundamental and overtone), totaling about 35% of the 66 cases. Descriptors describing both the onset, the steady state, and the proportion between the onset and the steady state were used. However, as noted earlier, the original list covered only 30% of the responses and comments, which indicates a need for an extended classification list.

Extension of the list of descriptors. The remaining 154 of the 220 original coded descriptors were analyzed on the following principles. (a) Frequency: words that were used frequently were given more weight than single occurrences. (b) New descriptors were coded by similarity to existing categories. (c) In the addition of new descriptors that did not fit existing categories, new subcategories were formed. Descriptors found earlier in one of the original categories were transferred to a new category if applicable. Not all descriptors used by the subjects are displayed in the categories. In *Table 3*, the new list is displayed. Descriptors in italics are new or added to the original list.

As displayed in Table 3, all original categories were extended by at least one descriptor. Furthermore, two new classes and eight new categories were formed from participants' responses.

For the transient class a category called *sound analogy* was formed. This category contains sound-like words such as "SSS-noise" and "bå-sound" (scandinavian å, like in b-oa-t). These words differ from the original list of onomatopoeic words in the sense that they make direct reference to the sound (noise and sound are used at the end). Also a new onomatopoeic word, *spit*, was added. It is believed that this word reflects a very fast transient sound. However, only one respondent reported this word. For the steady-state class, a new category named *physiology* was added. This category contains words that link the perceived sound quality to the human voice (words like *breathy, throaty,* and *singing*). An additional category, interpreted as *wind analogies*, contained several descriptors describing the wind component or wind quality of the sound. Examples of such words are: *airy, leaky, flowy, reedy*. Clearly, the descriptors in this category refer to wind systems and the sound quality of such systems. Also within this category can be found words such as *reedy* which refer to other wind instruments. Several descriptors within the extended list refer to other instruments (a good example of this is the descriptor *horn-like,* which was used by several participants to describe the tone quality of one of the pipes).

Table 3. Extended list of subjective descriptors

Class	Categories	Descriptors	Comments
Transient		onomatopoeia	
		Chiff	
		Cough	
		Hiss	
		Kiss	
		Spit	
	Sound analogy		
		"*bå-sound"*	
		"SSSS-noise"	
	Antonyms		
		Soft-strong	
		Weak-strong	
		Slow-fast	
		Slow-quick	
		Short-long	
		Aggressive- gentle	*(cf. general impression)*
	Physical	Pitch	
		Overflowing transient	
Steady state		Amount of:	
		Fundamental	
		Quint	
		Octave	
		Overtones	
	Antonyms		
		Stringy-fluty	
		Round-sharp	
		Full-thin	
		Light-heavy	
		Dirt-clean	
		Dirty-clear	
		Free-tight	
		Loose-tight	
		Floating/spatial-oppressive	
		Floating- suppressed	
	Physiology		
		Nasal	*reference to human voice*
		Breathy	
		Throaty	
		Singing	
		Spitting	
	Wind analogies		
		Airy	
		Leaky	
		Flowy	
		Reedy	
		Fluffy	
		Wooly	
		Windy	
		Floppy	
		Wind noise	
		Wind-rich sound	
	Perceived roughness		
		Raspy	
		Harsh	
		Rough	
	Metaphor		
		Horn-like	
		Like the sea	
		Like a singer that has a leaky voice	
		Soft like cloth	
		Strong fabric	

Class	Categories	Descriptors	Comments
General impression			
	Intense		
	Sandy		
	Sweet		
	Hollow		
Antonym:			
	Energy		
		Aggressive-gentle	*(cf. speed of transient)*
		Weak-strong	
		Tensed-relaxed	
	Spaciousness		
		Empty-full	
		Wide-narrow	
		Open-narrow	
		Introvert (turned inward)-extrovert/open	
	Complexity/clarity		
		Complex-simple	
		Complex-easy	
		Clear-diffuse	
		Noisy-clear	
	Timbre quality	*Bright-dark*	
		Light-dark	
		Warm-cold	
		Little/much character	
		Feminine-masculine	
		Pleasant-unpleasant	
General Impression of tone quality			
	Thin		
	Pressed		
	Forced sound		
	Dense		
	Deep		
	Raw		
	Old		
	Lyrical		
	Dull		
	Convincing		
Proportion between			
	Fundamental and overtone		
	Speech and steady state		
	Noise and musical tone		
	Fundamental-noise		
	Principal-noise		
	Noise-sound		
	Noise-harmonics		
Noise/overtones-principal/fundamental:			
	Disconnected		
	Well-balanced		
	Integrated		
	Connected		
	Related		
Noise or fundamental/principal:			
	Diffuse-distinct		
	Unstable-stable		
	Focused-unfocused		
	Defined-undefined		

Additionally, the category of physiology is closely connected to the category of wind analogies since the human voice is a wind system. This means that several of the descriptors within the physiology list could be categorized in the ”wind” category as well. However, it seems that the physiology category, even though related to the wind analogies, is qualitatively different. The new *metaphor* category also mainly refers to wind qualities of human voice qualities of the pipe (like a singer…, like a horn… etc.). Moreover, this category is related to onomatopoeic descriptions since it attempts to illustrate a phenomenon by referring to "hands-on" experiences. A category for the description of the *perceived roughness* of the harmonics and *noise* or the musical tone was formed. This category contains three adjectives: *raspy, harsh, rough*. It has been formalized under the category of steady state or general impression, but can probably also be used to describe the transient quality of the organ pipes. Further, it is quite interesting to note that roughness is a frequently used measure of product sound quality (Blauert & Jekosch 1997) and that a psychoacoustic metric to estimate subjective roughness has been developed and implemented (Zwicker & Fastl 1999).

An extensive list of antonyms was added to the general impression class. It is divided into four components: (1) *Energy*. Words reflecting the perceived energy or power, such as *weak-strong*. (2) *Spaciousness*. In this category, words refer to the tonal spaciousness. Words such as *open-narrow* and *introvert-extrovert* can be found here. (3) *Complexity/clarity*. Complex-simple and clear-diffuse are examples of adjective pairs that have been categorized to reflect perceived tonal clarity. (4) The fourth category is a general *timbre quality* category, which contains antonyms like *bright-dark* and *warm-cold*. Also adjective pairs of little/much character are to be found in this category. A category of general impression has been labeled *general impression of tone quality*. This category comprises adjectives that refer to the physical description (e.g. thin, forced) as well as adjectives describing more subjective properties (e.g. lyrical, dull) The three categories of spaciousness, clarity and timbre quality correspond to three of the general factors obtained by Gabrielsson and co-workers in factor-analytic studies on sound quality of loudspeakers (Gabrielsson & Lindström 1985).

Four new descriptions of the proportion between fundamental and overtone have been added to the original list: fundamental-noise, principal-noise, sound-noise, harmonics-noise. All these are clearly more focused on the noisy part in comparison with the original items on the list. However, it is hard to determine whether respondents refer to ”musical noise” or ”random noise” (stochastic signal). The current results can be interpreted as both, where the word pair sound-noise would refer to random, unwanted noise and the remaining pairs would refer to musical noises. Two additional categories were also formed: (1) *Relationship between noise/overtones and principal/fundamental*. Words in this category refer to how well noise and fundamental are connected, integrated or related to each other. This category is clearly linked to the previous one, but gives additional information in that it specifies the relationship or proportion between, for instance, fundamental and overtones. In respondents’ comments, these two categories were often combined to express this relationship. (2) The second category describes *Quality or character of noise or fundamental/principal*. Comments given in this category often form antonyms such as diffuse-distinct, unstable-stable, defined-undefined and focused-unfocused. Words in this category were used in conjunction with either the noise or the fundamental throughout all comments. This category is possibly one of the more important for the organ builder, since it seems to reflect the current state of pipe sound quality during the voicing process. It should also be noted that this category resembles the general-impression antonyms when no specific reference to the noise or the fundamental is given.

Summaries of comments for each pipe. As a final analysis, comments for each pipe in comparison with the reference pipe were summarized and interpreted. For each pipe, most prominent descriptors and comments are given. Respondents' comments have been compared over all different sounds, and an attempt to qualitatively characterize each sound from these data is made below.

s1. Narrow, noisy, suppressed, weak, tight, leaky, flowy.

s2. Smooth, warm, weak, slow transient, more fundamental than reference, transient disconnected, pleasant, airy, relatively undefined transient, quite close to the reference.

s3. Clear, nice, pleasant, good, more fundamental than reference but overall very similar, quick speech, free sound.

s4. Unpleasant, not beautiful, much chiff, unbalanced, undefined, unstable, needs improvement, not ready, bad, disastrous, harsh, much tension, too hard speech.

s5. Noisy, raspy, more wind noise than reference, less integrated than reference, transient not connected to fundamental, unfocused, diffuse, much noise in comparison with fundamental, empty, rough, too much (random) noise connected to the sound, dirty.

s6. Somewhat clearer than reference, sweeter principal, quiet balanced, not too bad but could be improved, a bit more fundamental and noise than the reference but still similar, lyrical, very close to the reference, overall too loose, dull, quite relaxed, light.

s7. Unpleasant, overall too forced, rather hollow, not enough defined, less noisy but more fundamental than the reference, many overtones, some air in the sound, balanced, horn-like, tense but balanced sound, not as free as the reference.

s8. Quite nice, less overtones than the reference, moderate chiff, good definition, cleaner and darker than the reference, warm in steady state, lyrical, good, stable, intense, round sound without many overtones, very good except the ”spit” or spitting in the transient, clear, tight and sweet, strong.

s9. A bit harsh, too penetrating and narrow, raspy, hollow, unbalanced, ”bå-sound”, unpleasantly tight, nasal, too noisy tone, too much noise.

Overall, agreement among respondents concerning the main quality (good-bad/positive-negative) exists. The sound qualities of the pipes are in some cases quite clearly differentiated on the basis of the summaries of qualitative judgments. However, most participants focus on different aspects of description when giving qualitative judgments. Interestingly, both absolute judgments of the sound quality (e.g. clear) as well as relative judgments of pipe sound quality (clearer than the reference) are being made by the respondents.

Finally, a comparison with previous results from quantitative analysis (Rioux et al. 1998) shows some correspondence with the summaries of qualitative judgments. In *Figure 2*, a two-dimensional multidimensional scaling (MDS) configuration is shown. The analysis reflects similarity judgments of the ten sounds.

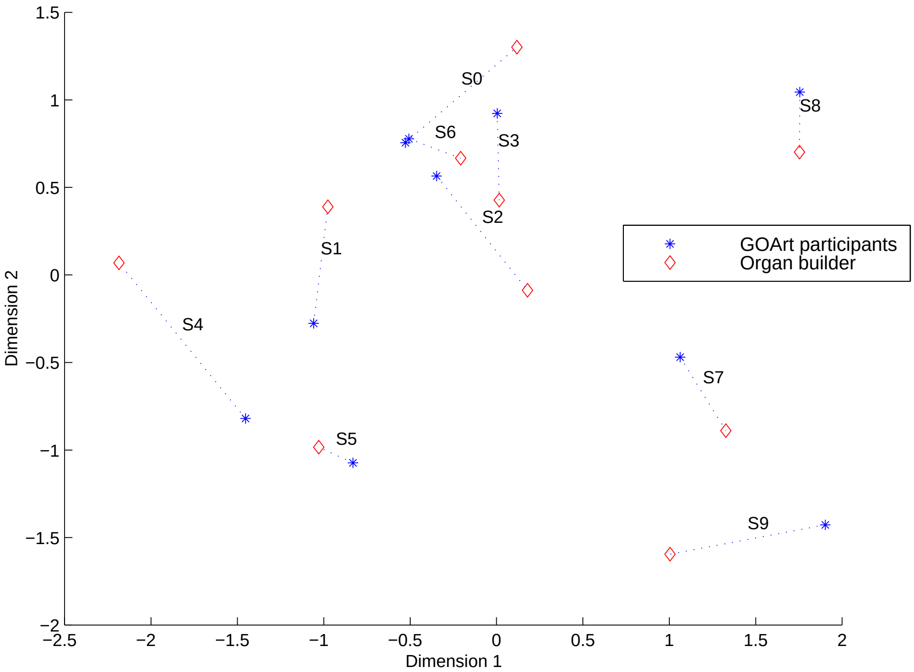


Figure 2. Two-dimensional MDS solution of participants' and organ builder's dissimilarity ratings of 10 organ pipe sounds. Taken from Rioux et al. (1998). No clear evidence was found for the interpretation of the two dimensions. Dotted lines represent the agreement between the voicer and the other test participants. As shown, the segments are always short and never intersect, demonstrating a good agreement. Moreover, the graph clearly displays three main groups of stimuli: (s0,s6,s3,s2), (s7,s8,s9) and (s1,s4,s5).

The projection of the ten pipes s0-s9 in the space reflects similarity on two dimensions of judgment. Sounds projected close to each other are perceptually similar on these two dimensions of judgment. Consequently, sounds far apart are dissimilar. Compared with the qualitative judgments, sounds 2, 3, and 6 are similar to the reference s0. This can also clearly be found in the qualitative judgments where s2, s3, and s6 are said to be very close or similar to the reference. Sounds s1, s4, s5, s7, and s9 are described overall as having lower quality than the remaining pipes, which is also reflected in the MDS configuration as the second dimension. Sound s8 is judged as a very good sound but different than the reference with respect to amount of overtones and transient, which may explain the difference from the reference s0 on dimension 1 in the MDS configuration. Pipes s1, s4, s5, s7, and, s9 are also qualitatively different, which is reflected in the MDS configuration.

6. CONCLUSIONS

The current article focused on qualitative judgments of perceived sound quality of flue organ pipe sounds where data were analyzed in three ways. First it was shown that the list of descriptors used by the organ builder to describe the sound quality of flue organ pipe sounds did not cover all aspects of participants' comments. This is not surprising since participants were asked to freely give any type of comments. However, it indicated that the current mapping could be extended. Hence, secondly, participants' comments were categorized and a new extended classification list was created. It is believed that this list, or a compromised version of this list, can be used for further studies of the perceptual quality of flue organ pipes. It is also hoped that this extension and classification can help the communication of voicing terms. Finally, comments for the nine different sounds were compared and summaries of prominent descriptors and comments for characterizing each pipe were given. Surprisingly, quite clear summaries were obtained for each pipe and from this analysis the pipes could be perceptually differentiated. Furthermore, the qualitative analysis was in good agreement with previous quantitative results, which clearly strengthens the interpretation. Future steps would be to relate respondents' comments to physical description of the sounds and to confirm/disconfirm the current classification of subjective descriptors by investigation of the factor structure.

REFERENCES

Berg, J. & Rumsey, F. (1999). Spatial Attribute Identification and Scaling by Repertory Grid Technique and other Methods. *Proc. of the AES 16th International Conference,* pp. 51-66.

Blauert, J. & Jekosch, U. (1997). Sound-quality evaluation – A multi-layered problem. *Acta Acustica, 83,* 747-743.

von Bismarck, G. (1974a). Sharpness as an Attribute of the Timbre of Steady Sounds. *ACUSTICA, 30,* 159-172.

von Bismarck, G. (1974b). Timbre of Steady Sounds: A Factorial Investigation of its Verbal Attributes. *ACUSTICA, 30,* 146-158.

Denzin, N., K. & Lincoln, Y. S. (1994). *The Handbook of Qualitative Research Theories and Issues*. SAGE Publications, Thousand Oaks/London/New Delhi, 1-17.

Gabrielsson, A. & Lindström, B. (1985). Perceived sound quality of high-fidelity loudspeakers. *Journal of the Audio Engineering Society, 33,* 33-53.

Grey, J. M. (1977). Perceptual evaluation of synthesized musical timbres. *JASA*, *61,* 1270-1277.

Krumhansl, C. (1990). *The Cognitive Foundation of Musical Pitch*. Oxford University Press, New York.

McAdams, S. (1999). Perspectives on the Contribution of Timbre to Musical Structure. *Computer Music Journal*, **23:3**, 85-102

Monette L. G. (1992). *The Art of Organ Voicing*. New Issues Press. Western Michigan Uiniversity. Kalamazoo, Michigan.

Moore, B. C. J. (1997) *An Introduction to the Psychology of Hearing*. 4th edition, Academic Press.

Noble, A. C., Arnold, R. A., Buechsenstein, J., Leach, E. J., Schmidt, J. O. & Stern, P.M. (1987*).* Modification of a Standardized System of Wine Aroma Terminology. *Am. J. Enol. Vitic.,* Vol. 38 No. 2, 143-146.

Nolle A. W., (1979). Some voicing adjustments of flue organ pipes. *J. Acoust. Soc. Am.*, *66:6,* 1612-1626.

Pelto, P. (1995). Four Voicing Techniques. *The Organ Yearbook,* XXV.

Rioux, V., Yokota, M., Västfjäll, D., Scholz, M., & Kleiner, M. (1998). Preliminary study of an organ builder's perception of a flue pipe sound. In *Proceedings of Inter. Symp. On Musical Acoustics 1998, Leavenworth, USA, pp. 174-180.*

Rioux, V. (1999). *Use of objective and subjective methods for evaluation of flue organ pipe sounds – An interdisciplinary study of the voicer's techniques.* Report F99-01, Department of Applied Acoustics, Chalmers University of Technology.

Rioux, V. (2000a). Methods for an objective and subjective description of starting transients of some flue organ pipes – integrating the view of an organ-builder. *Accepted for publication in Acta Acustica*.

Rioux, V. (2000b). Sound files and analyses of stimuli used in this paper, as well as a set of Matlab routines for sound analysis can be found online at: http://www.ta.chalmers.se/HomePages/Vincent/index.html

Rioux, V., Västfjäll, D., Yokota, M. & Kleiner, M. (1999c). Noise Quality of Transient Sounds: Perception of 'Hiss' and 'Cough' in a Flue Organ Pipe. In *Proc. of Forum Acusticum*. Berlin.

Samoylenko E., McAdams S. & Nosulenko V. (1996*).* Systematic Analysis of Verbalizations Produced in Comparing Musical Timbres, *International Journal of Psychology*, *31 (6)*, 255-278.

Scholz, M., Rioux, V., Västfjäll, D. & Kleiner, M. (1999). Psychat---A program for psychoacoustical tests. In *Proceedings of Forum Acusticum, 1999*.

Zwicker, E. & Fastl, H. (1999). *Psychoacoustics –Facts and models*. Second edition. Springer Verlag, Berlin, Germany.

Paper III

PAPER III

Verbal Descriptions of Flue Organ Pipe Sounds (II): A Quantitative Analysis

Accepted for publication in Musicae Scientae (2001)

Vincent Rioux, Daniel Västfäll

Chalmers Room Acoustics Group, Department of Applied Acoustics,
Chalmers University of Technology

Abstract:

In a companion paper (Rioux & Västfjäll 2000), a qualitative analysis of verbal descriptors of flue organ pipe sound yielded a list of possible descriptors. The present article deals with a subsequent dimensional analysis of data collected from a semantic differential test based upon these descriptors. From a quantitative analysis the initial "qualitative" categorisation is reviewed and enhanced. By inspecting the factor structure of the descriptors, groups of related descriptors were formed and suitable prototypes of each group were proposed. We then show how, from original data extracted from free comments, a detailed classification of descriptors can be made.

1. INTRODUCTION

Although instrument-builders and acousticians may both develop an extremely refined analysis of sounds, instrument builders will mainly use an sensory-empirical approach while acousticians mainly will follow an experimental-theoretical approach. Combining the two views requires work on the possible interface between these two approaches. The aim of this study is to extend a previous verbalisation corpus (thought as one of these possible interfaces). From a previous experiment dealing with the recognition of *micro*-aspects of sounds of flue organ pipes by musically trained and untrained subjects, a wealth of verbal comments, descriptions and *attributes* were provided. These comments were analysed qualitatively in a companion article (Rioux & Västfjäll, 2000). The methodology used to achieve a list of verbal descriptors of flue organ pipe sounds can be decomposed in six steps:

a) Collection of a list of descriptors from an expert organ-builder
b) Collection of free comments from trained subjects on a set of sounds
c) Derivation of an organised list by qualitative analysis

d) Reduction of the number of descriptors and sounds
e) Subjective experiment on the reduced set (descriptors, stimuli)
f) Derivation of a revised lexicon from the study of the factor structure of the data

The first three steps were carried out in Rioux and Väsfjäll (2000). The focus of the current article will be on the last three steps. Steps d) (reduction of the number of descriptors and sounds) and e) (subjective experiment on the reduced set (descriptors and stimuli) will be described in the method section. They are intimately linked to the way the listening test was constructed and performed. Step f) (derivation of a revised lexicon from the study of the factor structure of the data) is the backbone of this study and will thus occupy the main part of this article.
It is worthwhile to note that several attempts have been made to create list of *attributes or descriptors* for general impression of sound, timbre or reactions to sound. Gabrielsson et al. (1979), Bech (1999), Chouard et al. (1999), Susini et al.

(1999), Parizet et al. (1999), Västfjäll et al. (2000), Maffiolo (1999) and Guyot (1996) for example, have pursued research on sound quality for the hi-fi, automotive and aeronautic industry where verbal feedback from consumers was thought to be a valuable source of information for sound designers. Grey (1974), Von Bismark (1974) and more recently MacAdams (1999) have focused on a more theoretical approach of musical timbre description. It is our hope that bringing the results of all these researches could help an instrument-maker (the "sound designer") to integrate feedback from musicians (the "consumers"). In all these works, experiments and methods are numerous. For a review, see Beck (1999) or closer to the context of the present article, Rioux (2000).

Before proceeding, a distinction should be made between what will be in the following referred to as descriptors and attributes. Descriptors are considered to be basic entities of verbal description encompassing words (adjectives), group of words (metaphors) and *onomatopoeia*. Attributes carry information about a class of descriptors. In this respect an attribute will share the same properties as a prototype of a group of descriptors, summarising then a common feature shared by a collection of descriptors. It has been found that some attributes (like *loudness*, *sharpness* and roughness) can possibly be correlated with particular signal-features of a sound but we are dealing here with so fine adjustment of timbre that a quantification of our results in terms of correlation with signal features is not available at the present time.

2. METHOD

2.1 Listening test construction

The final list reported in paper I (Rioux & Västfjäll, 2000) displays 99 verbal descriptors. In a study of dimensionality and item inter-correlation, each of this descriptors represents a measured variable which takes a value on a scale. Moreover, these 99 descriptors were obtained from a listening tests performed on 10 sounds. In order to release the load of a listening test made on 99 scales for 10 stimuli (about one thousand evaluations), the number of verbal descriptors were reduced down to 85 and the number of sounds down to 5. The length of the test was then reasonable, lying between 40 minutes and one hour. Reducing the list of descriptors was achieved by the authors by rating the applicability and uniqueness and thus omitting items that either was a) believed to be inapplicable to the organ sounds or b) that were synonyms or close to items already in the list. The main part of the 14 words omitted was excluded on basis of their overlap with other words.

Reducing the number of sounds was both made on agreement between the authors and on inspection of multivariate analysis plots. These plots showed on figure 1 and 2 were obtained from a pair comparison test. In this test subjects rated the global perceived dissimilarity between each pairs (45) of stimuli (10). Figure 1 shows a MDS analysis of the obtained dissimilarity matrix while Figure 2 shows a Tree analysis of the same matrix (Barthélémy et al., 1988 and Guénoche et al. 1999). MDS analysis supports mainly a dimensionality point of view of the data set (cf. MacAdams (1999) and Grey (1974)) while a Tree analysis supports a categorical interpretation of the same data (cf. Maffiolo(1999), Guyot(1996)).

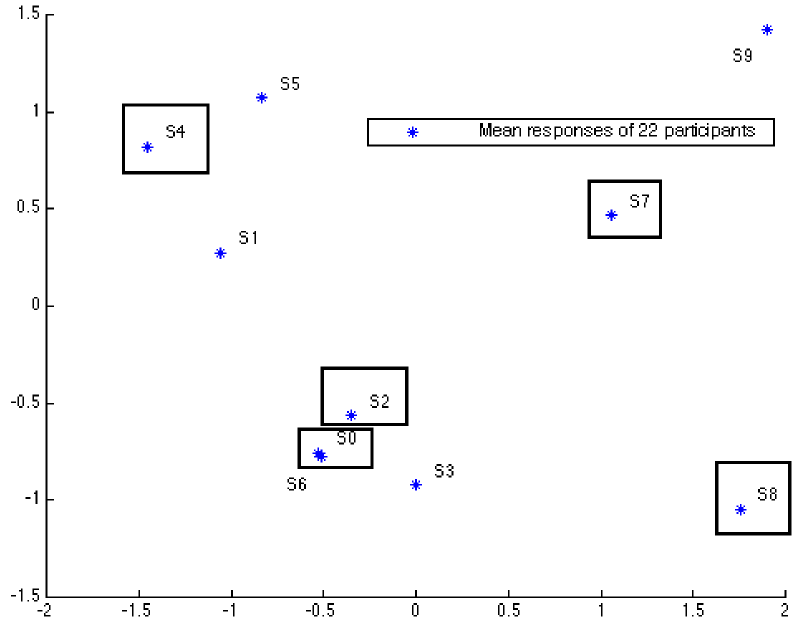


Figure 1. Projection of the first 2 dimensions of a Multidimensional Scaling analysis made on the dissimilarity matrix obtained from a pair comparison test (cf paper I). Distances between stimuli (s0 to s9) represents a measure of their dissimilarity. Here, the significance of dimensions is not of particular interest. This plot clearly shows two groups {s0,s2,s3,s6}, {s1,s4,s5} and some isolated stimuli {s7,s8,s9}. (s0,s2,s4,s7,s8) is the reduced set of sounds chosen.

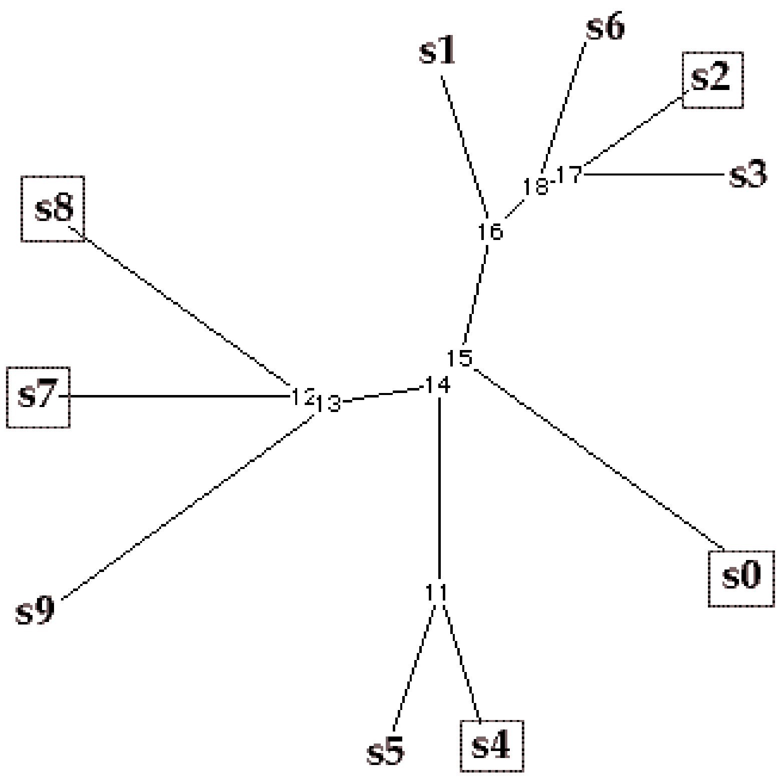


Figure 2. Tree analysis of the same dissimilarity matrix than used for the MDS representation displayed on figure 1. This plot shows the hierarchical structure of the data set. Four groups emerge clearly {s0}, {s1,s2,s3,s6}, {s4,s5} and {s7,s8,s9}. (s0,s2,s4,s7,s8) is the reduced set of sounds chosen.

In order to reduce the set of sounds we chose the sounds which were prototypical for each of the groups found in the analyses. The tree analysis provided help in selecting the groups while the MDS analysis provided with a good support for the evaluation of the dissimilarity between each stimuli. It is noteworthy to state that combining these two methods proved to be surprisingly useful. The final set of sounds retained was then constituted by the sound of reference s0, and then (s2,s4,s7,s8).

2.2 Participants and procedure.

15 undergraduates and graduates at Chalmers University of Technology participated on a voluntary basis. Participants were considered to be naive. They came from different countries such as Sweden, France and USA. Participants arrived individually to the laboratory and were seated in front of a computer. Participants were then instructed by a male experimenter on the task and the use of the computer program and questionnaire. The computer was used as a wave-player (using a PowerPoint™ presentation sheet). Sounds were played through headphones. Subjects were presented with all sounds once so that they could get a grasp of the domain of variation of the timbre quality. With sound examples[8], they also were briefly introduced to the concept of noise vs. harmonic components and transient vs. stationary part.

Participants were given the following instructions on the first page of the questionnaire:

"You will be asked to rate the amount of each quality defined by its descriptor, that you find in a particular sound. All presented sounds are coming from organ pipes". Descriptors are classified as following:
`general`: the general impression
`steady`: perception of the steady or stationary part (the "middle" part of a sound)
`transient`: perception of the starting transient (also called "speech" or "attack")
`noise/tone`: a qualitative comparison between the noise and the tone [listen to the separation of harmonics and *noise*]

Participants were instructed that they could put a mark between 0 (no amount) and 10 (maximum amount) for each descriptor and for all sounds. In addition, participants were instructed that if a descriptor did not "fit" or was adequate for a particular sound they may put mark 0. All descriptors were presented in English. The task description and a sample of the scales is given in Table 1. Participants completed the ratings and were then debriefed and thanked for their participation.

[8] sound examples are presented online at: www.ta.chalmers.se/homepages/vincent/index.html

Table 1. Sample of a questionnaire.

Welcome to this listening test. In order to be able to characterize verbally sounds, we need to know which descriptors are the most suitable. In order to do that, you will be asked to rate the amount of each quality defined by its descriptor, that you find in a particular sound. All presented sounds are coming from organ pipes. Descriptors are classified as such: `- general`: the general impression `- steady`: perception of the steady or stationary part (the "middle" part of a sound) `- transient`: perception of the starting transient (also called "speech" or "attack") `- noise/tone`: a qualitative comparison between the noise and the tone [listen to the separation of harmonics and noise] You may give a mark between 0 (no amount) and 10 (maximum amount) for each descriptor. Note that if you think that a descriptor doesn't "fit" to a particular sound you may put mark 0. example: *SOUND 1* *general impression* *raw* --------------------- |__*5*__| *(0-10)* *noisy* ------------------- |__*7*__| *(0-10)* *deep* -------------------- |__*1*__| *(0-10)*

All variables (cf. Table 2) were presented under a simplified structured scheme based on the physical parts of the sounds they describe. Those are:

- starting transient versus stationary parts (decay part not considered here)
- periodic/harmonic/deterministic signal versus probabilistic/noise-like/*stochastic* signal
- global description of the sound as a single object

Attributes may also be classified by their semantics/linguistic/function such as analogies ("like a horn", "bright"), onomatopoeia ("chiff") or purely analytical ("amount of overtones"). Only the first simplified structure was proposed to the subjects in order to help them focus their attention. This was certainly necessary as some of these attributes definitely become useless without a precise context.

Table 2. List of 85 descriptors.

general impression	steady state	transient
aggressive	airy	chiff (sounds like)
bright	breathy	cough (sounds like)
clear	bright	hiss (sounds like)
cold	clean	spit (sounds like)
complex	clear	agressive
convincing	cold	fast
dark	dirty	gentle
deep	dull	long
defined	floppy	short
dense	flowy	slow
diffuse	fluffy	soft
distinct	fluty	strong
dull	free	weak
easy	full	**noise/tone**
focused	harsh	connected
forced	horn-like	disconnected
gentle	leaky	integrated
light	loose	related
noisy	nasal	well-balanced
old	oppressive	
pleasant	reedy	
pressed	rough	
raw	round	
relaxed	sandy	
simple	sharp	
stable	singing	
strong	spitting	
tensed	stringy	
thin	thin	
undefined	throaty	
unfocused	tight	
unpleasant	warm	
unstable	windy	
warm	wooly	
weak		

3. RESULTS

First to test if all attributes discriminated between sounds repeated analysis of variance (ANOVAs) was performed on each attribute with the five sounds as within factor. The ANOVAs yielded highly significant effect for all attributes (all F $p<.05$). Since the attributes discriminated between sounds participants ratings of all attributes for all sounds were submitted to Principal Component Analyses (PCAs) for each of the parts[9]. For the general impression category (comprising 35 items

[9] Combining variance due to individuals and sounds may inflate error variance but is still a commonly used method of increasing observations and ability to generalize factor patterns in factor analytic studies (Osgood et al, 1957; Smith & Ellisworth, 1985). However, to control for possible influences of error variance separate factor analyses and MDS was performed for each sound. Since these analyses yielded highly comparable dimensional patterns it is concluded that that combination of individual and stimuli variance yielded a negligible influence in the present case.

see Table 3) a six dimensional solution accounted for 65 percent of the variance. After Varimax rotation, the scree criterion indicated that a six factor solution was applicable. As can be seen in Table 3, descriptors denoting simplicity (*simple, defined, focused, distinct and clear*) loaded on the first factor.

Table 3. PCA on the "general impression" class of descriptors.
Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 30 iterations.
TQ = Tonal Quality / NQ = Noise Quality / GQ = Global Quality

Descriptors	Factor loadings for dimensions						Mean	StD	Categories Prototypes Attributes
	1	*2*	*3*	*4*	*5*	*6*			
DEFINED	**.88**					.15	3.9	2.4	TQ/NQ: simplicity
FOCUSED	**.84**	.12	.15		-.13	.24	3.3	2.1	
CONVINCING	**.75**	-.22		.12		.34	4.3	2.4	
DISTINCT	**.68**					-.29	3.3	2.3	
CLEAR	**.67**		.24	.12	-.20	-.11	4.2	2.2	
SIMPLE	**.61**	-.20	.27	-.13	-.22	-.32	4.6	2.8	
FORCED		**.91**		.11	.12		3.9	2.3	TQ: tension
TENSED		**.89**		.17		-.13	4.9	2.2	
PRESSED		**.88**			.16		5.2	2.6	
LIGHT			**.77**		-.23	.12	5.1	2.5	TQ: clarity+weakness
BRIGHT	.21	.12	**.72**	-.20	.23		4.0	2.3	
GENTLE	.24		**.71**	.30			4.7	2.5	
EASY	.13	-.44	**.65**				3.9	2.7	
WEAK	-.33		**.64**	-.29		-.38	3.5	2.2	
THIN		.17	**.56**	-.20		-.55	5.0	2.4	
RELAXED	.11	-.44	**.53**	.20		.42	4.6	2.4	
DARK		.17		**.79**			3.2	2.0	TQ: darkness+strength
DEEP	.27	-.14	-.20	**.73**	.11		3.0	2.3	
WARM	-.28		.24	**.69**		.19	5.1	2.2	
STRONG		.49		**.69**	-.15		4.7	2.8	
DENSE	.20	.29	-.32	**.43**		.40	4.7	2.5	
DIFFUSE					**.79**		4.9	2.5	TQ/NQ: complexity
UNDEFINED	-.44	.14	.28		**.64**	-.15	4.1	2.4	
UNFOCUSED	-.39		.28		**.63**	-.11	3.5	2.1	
RAW		.37		-.11	**.63**		4.5	2.4	
NOISY	-.38	.25	-.11	-.25	**.42**	.22	4.7	2.5	
COMPLEX	-.12	.24		-.15	**.41**	.41	3.4	2.6	
UNPLEASANT	-.16	.53		.13	.12	**-.58**	4.7	2.3	APPRECIATION
PLEASANT	.41	-.46	.38			**.51**	2.8	2.7	
UNSTABLE	-.16	.29			.13		3.1	2.6	GQ: simple/complex
STABLE	.46					.25	3.2	2.3	
AGGRESSIVE	***.36***	***.32***	-.24		***.47***		6.2	2.5	
COLD	.10		.15				3.5	2.8	
DULL	-.25	.26					4.9	1.9	
OLD	.13			.18	.15		2.6	2.3	

Descriptors denoting tension (*forced, tensed, pressed*) loaded on the second factor. On the third factor descriptors denoting, clarity (*light, bright, gentle, easy, thin, relaxed*) loaded. The fourth factor contained descriptors such as *dark, deep, warm, dense* and *strong* and may be interpreted as a darkness/strength factor. It is clear that factor three and four are related to general timbre factors retrieved in factor studies (Gabrielsson & Lindström 1985; Grey, 1977; Bregman, 1990). On factor five a mirror factor of the first simplicity factor was retrieved containing descriptors such as *complex, diffuse, undefined, unfocused, noisy and raw* and may thus be interpreted as a complexity factor. The last factor was a bipolar pleasantness-unpleasantness factor (valence factor) usually retrieved in factor analytic studies of individuals reaction to objects and stimuli (Osgood et al., 1957; Västfjäll et al., 2000; Russell, 1980). *unstable* and *stable* were placed in a category concerning the Global Quality. Finally the descriptors*, aggressive, cold, dull* and *old* did not load meaningfully on any factor. However, Table 3 shows a tentative classification of these attributes.

Next a PCA with Varimax rotation was performed on participants ratings of the steady state descriptors (34). The scree criterion indicated seven factors together accounted for 62 percent of the variance. As may be seen in Table 4, descriptors denoting <u>leaking noise</u> (*floppy, leaking, loose, spitting, dirty*) loaded on the first factor.

Table 4. PCA on the "steady state" class of descriptors.
Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 18 iterations.
TQ = Tonal Quality / NQ = Noise Quality / GQ = Global Quality

Descriptors	Factor loadings for dimensions							Mean	StD	Categories Prototypes Attributes
	1	*2*	*3*	*4*	*5*	*6*	*7*			
DIRTY	**.79**			-.13	.38			4.8	2.2	
SPITTING	**.78**	.14			.20			4.2	2.3	
FLOPPY	**.64**							3.1	2.1	NQ: leaking
LEAKY	**.59**	.12	.40					4.0	2.6	
LOOSE	**.49**	.16	.22	.38		-.44		4.1	2.4	
BRIGHT	.12	**.78**	.11	.27				3.0	2.4	
SHARP	.17	**.78**			.16	.19	-.11	3.8	2.7	
STRINGY		**.58**		-.26		.21		4.3	2.4	TQ:
NASAL	-.40	**.56**	.18		.20			1.8	2.3	sharp
AIRY	.20	**.50**	.47	.40				4.2	2.3	
THIN		**.49**	.30	.14		-.23	-.46	3.3	2.5	
WINDY		.14	**.83**					4.1	2.7	
BREATHY	.15		**.76**	.25	.10			3.1	2.5	NQ: fluctuation
FLOWY		.14	**.42**	.20		.26	.41	4.6	2.6	
FREE		.12		**.79**	-.13			3.7	2.4	
ROUND	-.11	-.16	.25	**.68**	.12		.35	5.1	2.6	TQ:
SINGING	-.13	.27		**.57**	.12	.48		4.6	2.9	clarity
CLEAR	-.36	.34		**.41**	-.38	.23	-.12	3.4	2.7	
HARSH	.21	.20	-.14		**.76**		-.12	4.9	2.7	
ROUGH	.27	.10			**.71**	-.21	.19	3.8	2.6	NQ: rough
SANDY	.10		.27		**.69**	.26		3.1	2.7	
TIGHT		.25				**.77**		4.2	2.6	
FULL			.11	.18		**.62**	.56	4.3	2.5	TQ:
OPPRESSIVE		-.12	.13	-.34	.27	**.79**		3.4	2.7	tension
WARM				.13		**.64**		3.5	2.5	
CLEAN	-.41	.33	.19	.31	-.34	***.78***	.14	3.9	2.7	
COLD		.17	.37			***.59***	.12	2.0	2.1	
WOOLY						***.49***	.16	2.4	2.5	
FLUFFY	.13		.21	.25	-.32	.12	**.78**	2.5	2.1	
REEDY	.23	.17	.17		.14	.17	**.78**	4.0	2.4	
FLUTY	-.11	.21	.30				**.58**	4.2	2.7	
DULL	.19		-.15	.19	.36	-.40	**.56**	5.0	2.1	
HORN LIKE		.15	-.17			.20	**.50**	4.4	2.3	
THROATY	.22		.14		.14		**.49**	2.8	2.5	

On the second factor, descriptors denoting <u>sharpness</u> loaded (*bright, sharp, stringy, nasal, airy, thin*). The third factor was interpreted as a <u>noise fluctuations</u> factor containing descriptors such *as windy, breathy, and flowy*. The fourth factor was interpreted as a <u>clarity</u> factor containing descriptors such as *free, round, singing and clear*. The fifth factor reflected <u>noise roughness</u> (*harsh, rough, sandy*). The sixth factor was a <u>strength+dark</u> factor comprising descriptors such as *tight, full, oppressive, warm*. The remaining descriptors did not load meaningfully on a single dimension and were thus left for themselves…

Finally, participants ratings of the transient descriptors (13) was submitted to a PCA with Varimax rotation. Table 5 shows means, Standard deviation (Sd) and factor loadings for the transient descriptors.

Table 5. PCA on the ”transient” class of descriptors.
Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 7 iterations.

Descriptors	Factor loadings for dimensions				Categories Prototypes Attributes
	1	*2*	*3*	*4*	
SLOW	**.93**		.21		DURATION
LONG	**.93**		.21		
SHORT	**-.86**	.36			
FAST	**-.70**	.54			
AGGRESSIVE		**.83**	-.37		aggressiveness
COUGH	-.12	**.67**	-.19		
CHIFF		**.61**	.31	-.16	
SPIT	-.21	**.46**		.11	
SOFT	.23		**.86**		softness
GENTLE	.12	-.11	**.84**		
WEAK		.10	.43	**-.75**	STRENGTH
STRONG	.11	.47	-.12	**.67**	
HISS		-.17	.34	**.67**	

A four-dimensional solution accounted for 71 percent of the variance. As may be seen in Table 5, descriptors denoting duration/speed of transient (*slow*, *long*, *short*, *fast*) loaded on the first factor. On the second factor onomatopoetic descriptors loaded (*cough*, *chiff*, *spit*). However also the descriptor aggressive loaded on this factor suggesting that the onomatopoetic descriptions cough, chiff and spit are linked to aggressiveness. This interpretation seems reasonable since all these subjective descriptions are related to a quick, violent release of energy in the transient. The third factor contained the two descriptors *soft* and *gentle* and was interpreted as a softness factor. The fourth and final factor contained the descriptors *weak* and *strong* and was interpreted as a strength of transient factor. However, also the onomatopoetic description *hiss* loaded on this factor suggesting that the subjective descriptor *hiss* is linked to strength of transient.

The bipolar descriptors concerning the relationship between *noise* and tone was not submitted to dimensional analyses since they were expected to vary along one dimension.

In sum, the results indicate that stable categories of descriptors of different physical parts of flue organ pipe sounds can be retrieved. The original list of 85 descriptors was reduced to in all 17 dimensions. A question now is how these 17 dimensions are related to the classes of descriptors retrieved in a qualitative analysis (Rioux & Västfjäll, 2000).

4. DISCUSSION

4.1. Correspondence with qualitative analysis

GENERAL IMPRESSION

From the qualitative data we derived a *general impression class* which included the following categories: *energy*, *spaciousness*, *complexity/clarity*, *timbre quality* and *tone quality*. On table 3, a revised categorisation is proposed.
Note that the category *spaciousness* was not considered when building the list of descriptors in order to avoid confusion between the acoustic space (all recordings were monaural!) and the geometry of the pipe (narrow or wide scale).
Looking at table 3, three main categories appear tension, complexity/simplicity and darkness+strength/clarity+weakness. Even if the main characters of the previous categorisation are conserved, many changes must be made.
A new category called tension is created. The *energy* and *timbre quality* appear to be so strongly correlated that they now form a unique quantitative category called darkness+strength/clarity+weakness.
The *timbre quality* and *tone quality* are somehow redefined for clarifying reasons. Indeed, and as it is strongly supported by our classification of descriptors, we will from now on, consider that the timbre quality is a combination of a tone and noise quality. From this step, we realise that we can reinterpret the quantitative categories as affecting the tonal quality (TQ), the noise quality (NQ) or both at the same time. This gives us the opportunity to understand the complexity/simplicity as a perceived balance between the noisy and tonal parts. *Strong* is correlated with tension (.49) but still loads more with the darkness+strength category suggesting that tension is a valid category but not totally orthogonal to a strength (or energy) category. *aggressive* both load on NRJ and timbre quality.

STATIONARY PART

From the qualitative analysis of free comments, 6 categories of descriptors were found: *amount of tonal component*, *antonyms*, *physiology*, *wind analogies*, *perceived roughness* and *metaphors*. The first category (amount of tonal component) was not presented during the listening test as its interpretation in terms of signal feature is straightforward. So as the last categories (metaphors) for the opposite reason.
In table 4, results from the PCA analysis are presented and 6 categories clearly appear. The *wind analogies* and *perceived roughness* share close correspondence to the noise fluctuations and noise roughness. But the *wind analogies* category comprises 2 descriptors of the leaking noise category as well.
The descriptors of the later *physiology* category are scattered among different categories in the new classification and consequently does not seem to be a relevant category for this type of description. This should not be too surprising as human vocal system and flue organ pipes share a quite common physical structure. And thus sounds produced by these two systems should share the basic same categories.
Finally, the category *antonym* quite naturally doesn't appear in the new table confirming the fact that descriptors should be classified by the sound property they describe more than by their own grammatical property. Instead of the *antonym* category we then found 3 categories concerning mainly the tonal quality : sharp, clarity and tension.

9 descriptors (*clean, cold, wooly, fluffy, reedy, fluty, dull, horn-like and throaty*) did not load significantly on any dimension, suggesting that they should be used with care as they can't be associated to a single attribute but rather to an association of attributes.

TRANSIENT PART

The qualitative analysis of descriptors concerning the transient part used four categories *onomatopoeia*, *sound analogy*, *antonyms* and *physical*. As there was only 4 descriptors contained in both categories *sound analogy* and *physical*, these descriptors were not used for the quantitative analysis. Instead, 13 descriptors extracted from the categories *onomatopoeia* and *antonyms* were submitted to the factor analysis. As described earlier, the obtained classification (cf. table 5) is composed of four categories which concerns duration, aggressiveness, softness and strength. Finding names for these categories was straightforward and considering the physical complexity of transients this can be quite surprising. However we showed previously (cf. Rioux 2000) that onomatopoeia can very well describe a particular feature of a transient.

4.2. Remarks

Very few descriptors were used to express noise quality in the general impression class whereas they were numerous in the steady state class.
Comparing the classification obtained from the qualitative analysis to the results displayed on table 4 and 5 (descriptors for steady state and transient) allows us to notice that the factor analysis provides directly with a signal-oriented classification of the descriptors whereas the qualitative analysis provided a more syntactical analysis of these descriptors. This of course has to do with the fact that the classification during the qualitative analysis was done by the authors themselves and thus could mainly be achieved according to the standardised classification of linguistics. Such a factor analysis thus offers us the possibility to present a normative / objective classification of descriptors.
A summary of this classification is shown on figure 3.

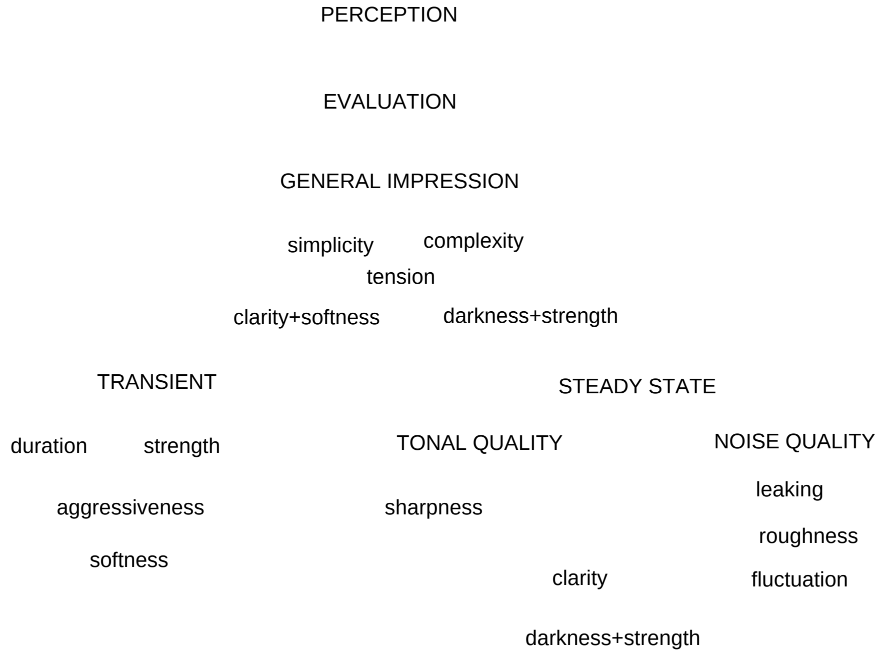


Figure 3. Schematic representation of the structure of attributes found to be suited for a verbal description of flue organ pipe sounds.

CONCLUSION

Outputs of a semantic differential listening-test were used as inputs of a Principal Component Analysis. An original lexicon based on free verbal comments was thus validated and structured by inspecting the various factor loadings. In order to complete the proposed classification it would be needed to correlate descriptors and their attributes to signal features using signal processing methods.

REFERENCES

Barthélémy & Guénoche (1988). *Les arbres et les représentations des proximités.* Paris. Masson.

Bech, S. (1999). Methods for subjective evaluation of spatial characteristics of sound. *Proceedings of the AES 16th International conference 1999 april 10-12*, pp. 487-504.

Bregman, Albert (1990). *Auditory Scene Analysis*. MIT press, Cambridge MA.

Chouard, N. & Hempel, T. (1999) *A semantic Differential Design Especially Developped for the Evaluation of Interior Car Sounds*. Proc. of Forum Acusticum, Berlin.

von Bismarck, G. (1974). Timbre of Steady Sounds: A Factorial Investigation of its Verbal Attributes. *ACUSTICA* 30, 146-158.

Gabrielsson, A. & Sjögren, H. (1979). Perceived sound quality of sound reproducing systems. *J. Acoust. Soc. Am., 65,* 1019-1033.

Gabrielsson, A. & Lindström, B. (1985). Perceived sound quality of high-fidelity loudspeakers. *Journal of the Audio Engineering Society, 33,* 33-53.

Grey, J.M. (1977) . Perceptual evaluation of synthesized musical timbres. *JASA*, 61, pp 1270-1277.

Guénoche, A. & Garreta, H. (1999). Quelle confiance accorder à une représentation arborée?. *Proceedings of the "Journées de la Société Francophone de Classification", Nancy 14-17 septembre 1999*.

Guyot Frédérique, (1996) *Étude de la perception sonore en termes de reconnaissance et d'appréciation qualitative: une approche par la catégorisation*, Ph.D. thesis, Université du Maine, pp. 70-80.

McAdams, S. (1999). Perspectives on the Contribution of Timbre to Musical Structure. *Computer Music Journal*, 23:3, pp. 85-102

Maffiolo, V. (1999). *De la caractérisation sémantique et acoustique de la qualité sonore de l'environnement urbain*. Thèse de doctorat de l'Université du Mans.

Osgood, C. E. (1957). *The measurement of meaning.* Urbana, Ill.: University of Illinois Press.

Parizet, E. & Nosulenko, V. (1999). Multi-dimensional listening tests: Selection of sound descriptors and design of the experiment. *Noise Control Eng. J.* 47 (6) pp. 227-232.

Rioux, V. (2000). Methods for an objective and subjective description of starting transients of some flue organ pipes - integrating the view of an organ builder. *Accepted for publication in Acustica to be issued in autumn 2000.*

Rioux, V. & Västfjäll, D. (2000). Verbal Descriptions of Flue Organ Pipe Sounds (I) : A Qualitative Analysis.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*, 1161-1178. Smith, C. A., & Ellisworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology, 48,* 813-838.

Susini, P., McAdams, S., Winsberg, S. (1999). A Multidimensional Technique for Sound Quality Assessment, *ACUSTICA*, 85:5, pp. 650-656
Västfjäll, D., Friman, M., Gärling, T., & Kleiner, M. (In press). The measurement of core affect: A Swedish self-report measure derived from the affect circumplex. *Scandinavian Journal of Psychology*.

Paper IV

PAPER IV

Subjective Evaluation of Multidimensional Auditory Stimuli: A 2D Scaling Method Using An Interactive Computer Interface

Submitted in september 2000 to Behavior Research Methods, Instruments, and Computers

Vincent Rioux[1], Daniel Västfjäll[1 & 2]

[1] Chalmers Room Acoustics Group, Department of Applied Acoustics, Chalmers University of Technology, Göteborg, Sweden.
[2] Department of Psychology, Göteborg, Sweden.

ABSTRACT

We present a new method for assessing the perceived global difference between multidimensional auditory stimuli. This alternative to classical methods performed with pen and paper (such as paired comparison), takes full advantage of the recent development of human/computer interfaces. Moreover, it can be combined with other listening tests modalities such as verbalization and categorization of stimuli. Auditory stimuli are represented by graphic icons. The task of the test participants is to match a visual organization of these icons (on a surface) with the relative perceived differences of their corresponding auditory stimuli (expressed as distances). We present as well, an original statistical analysis method suited to an interpretation of such raw output data.

The second half of this paper deals with a specific application of this method to the perception of musical timbre, distinguishing between an expert and a group of semi-expert subjects. Results are promising and show good agreement with conventional methods such as paired-comparisons.

KEYWORDS: psychoacoustics, computer interface, multidimensional analysis, musical timbre, verbalization.

1. INTRODUCTION

This paper presents a new computer-based technique for the study of auditory perception. It was originally developed for the study of the evaluation of timbre of musical sounds by the auditory system. We first present some theoretical background for the context of musical timbre and then describe the new test method and one of its applications.

Investigations of the multidimensional perceptual space of musical timbre have been carried out by various researchers. For instance, McAdams (1999), following the early work of Grey (1977), performed a Multidimensional Scaling of dissimilarity ratings of 18 synthesized timbres and found three main dimensions correlated with acoustical parameters, namely the spectral *centroid*, spectral flux and attack time. Von Bismarck (1974), using a semantic differential approach (based on absolute judgements), found that "sharpness" (a close companion of the central spectroid) was a reliable verbal attribute for the discrimination of 30 steady sounds equalized in *pitch* and *loudness*. Samoylenko et al. (1996) worked on free verbalizations and produced what was probably the first systematic classification system of musical timbres. The problems exposed in such works can be summarized in the following points:

- Perception of musical timbre is a complex multidimensional construct.
- Quantitative (e.g. pair comparisons) as well as qualitative (e.g. free verbalizations) analyses should ideally be combined in order to grasp this complexity fully.
- Inter-subject differences are difficult to establish because the dimensions of musical timbre are not yet elucidated.
- It seems difficult to perform experiments based on difference limens or Just Noticeable Differences (JND) except for intensity (for a review of loudness JND's see Hartmann, 1997), as it is difficult to exhibit stimuli varying with only one parameter (perceptual dimensions of timbre are inter-correlated).

Consequently, the study of perception of musical timbre is a complex task. We believe that, considering point 3 (the amount of unknown variables), we have to use some exploratory method. Points 1 and 4 emphasize the fact that we might ask participants to discriminate stimuli at a global level where contributions of the dimensions of each of the stimuli are summed up. The need for a specification of the multidimensional aspect of musical stimuli can be satisfied through verbalization and grouping of stimuli (point 2). The following presents a method based on this paradigm.

2. WHY A NEW METHOD?

Table 1 lists and comments on a panel of existing methods together with their advantages and drawbacks.

A distinction is made between methods based on a number of stimuli fixed by statistical or psychophysical criteria (fixed comparisons) and methods that are not (multiple comparisons).

Paired (or triad) comparisons (compare stimuli A and B) are well known and widely used (e.g. David, 1969) and best satisfy the test leader eager to maintain strict control over the response configurations. Nevertheless these methods do not exclude problems due to this relative lack of flexibility. Two-to-two pair comparisons (compare D to C as A to B) emphasize the necessary relativistic approach when examining relationships among perceptions (cf. McAdams, 1999 and Levin, 2000) but put more constraints on the design of stimuli.

Table 1. Comparison of different test methods in terms of their respective advantages and drawbacks.

	fixed-comparisons		**multiple-comparisons**		
				multiple scaling	
	pairs or triads	2 to 2	categorization (grouping)	2D scaling (computer based)	1D rating[10] (or ranking)
pros	Statistically controllable suitable for small differences in stimuli	- suitable for the description of two families of stimuli[11]	allow elicited constructs	preserve familiar 2D ordering no *learning effect* low edge-effects categorization can be included no need for a reference	reduce scale edge-effects eliminate time-dependence due to order of presentation
cons	only pre-defined constructs (except for global difference) tedious because very repetitive learning effects scale edge-effects[12] fuzzy reference	- restrictions on the design of stimuli		may become tedious because offers infinite combinations subjects need to be familiar with computers	may not reveal multidimensiona l differences

Methods of multiple comparisons (compare A, B, C, D, E,…) encompass all the possible methods in which all pair-type comparisons of the set under study are to be made "at once" and not sequentially. There is then absolutely no constraint on the order of appearance of the stimuli (they are all comparable at the same time). More precisely, it is up to the participants to make judicious comparisons and possibly repeat them whenever they find it necessary.

An example of a categorization test is sketched in Figure 1.

[10] Difficult with pen and paper when the number of stimuli exceeds 4 or 5.

[11] For example, one might be interested in knowing if it is possible to find an orange O2 that differs from O1 just as apples A1 and A2 are dissimilar.

[12] A subject may not interpret the edges of a scale in a consistent way for all comparisons.

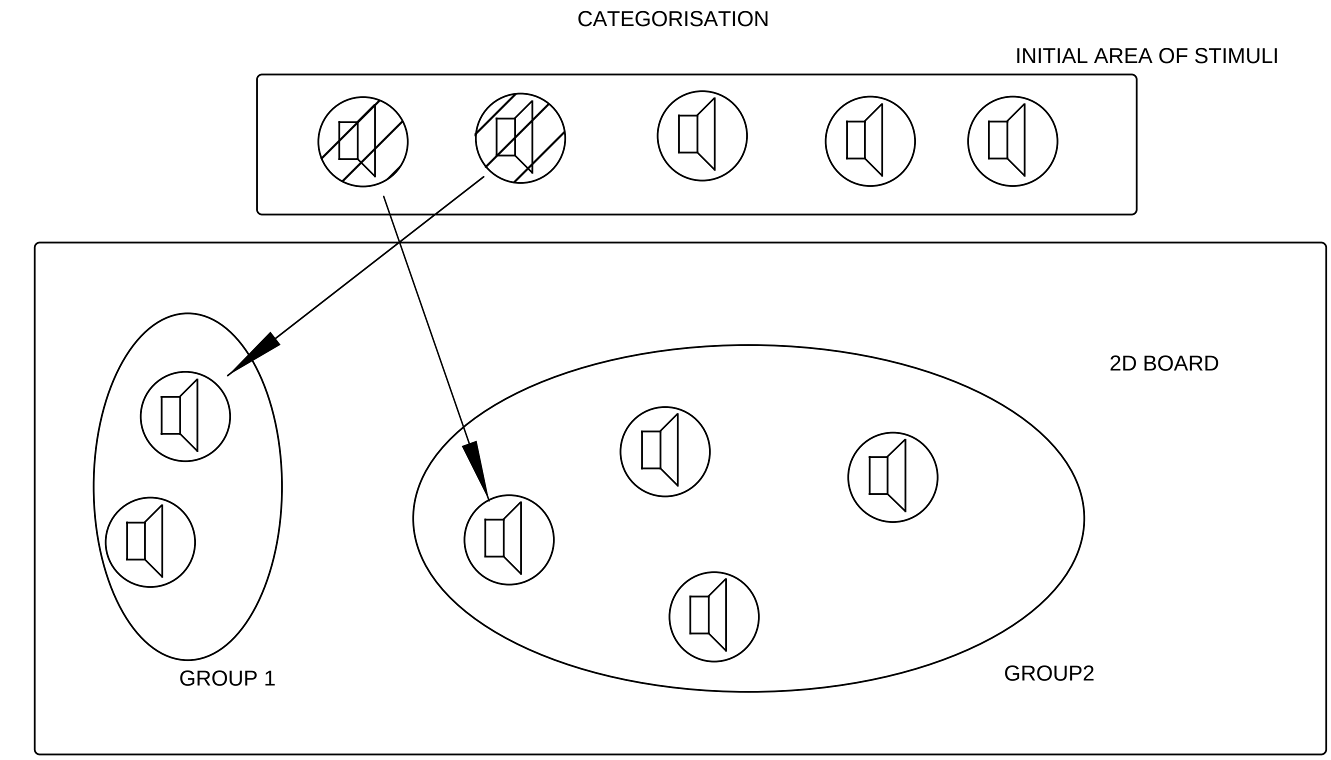


Figure 1. Schematic representation of a simple categorization task performed with a computer interface. Stimuli are symbolized by loudspeakers and can be drag-and-dropped on the working board. Groups are constituted by delimiting a frontier with the mouse. In this particular case inter-stimuli distances are disregarded.

The method, labeled "2D scaling" in Table 1 and sketched in Figure 2, combines the advantages of the 1D ratings and offers the possibility of categorization. A detailed description of the analysis of a 2D scaling test is presented in the next section.

It can be argued (David, 1969, p.10) that multiple scaling methods can provoke tedious pair-type comparisons, especially when inter-stimuli differences are small. This might be the most severe drawback of such a method, but is certainly better than retaining the scale-span (space) and order of presentation (time) dependence inherent to the paired comparison.

Traditional methods used up to now to assess the perceptual differences of auditory stimuli can also be classified according to the amount of constraints placed on the participants. This is mainly reflected in the way subjects are asked to focus on the various aspects of multidimensional stimuli. For example, if the test leader mentions scales, then the participants have no chance to propose their own constructs. On the other hand, free categorization tests leave the participants with complete freedom but make greater demands on the analyst. The possibility of integrating the advantages of both rating methods and categorization methods should be regarded as a decisive step forward allowed only by recent developments in human-computer interfaces.

3. PRESENTATION OF THE METHOD

We developed a "2D scaling and grouping" test in order to suppress memory-order effects and decrease edge effects of scales. For reasons of flexibility, multimedia abilities and ease of storing results, we had to implement this type of test using computers. A typical layout for this new type of test is shown in Figure 2.

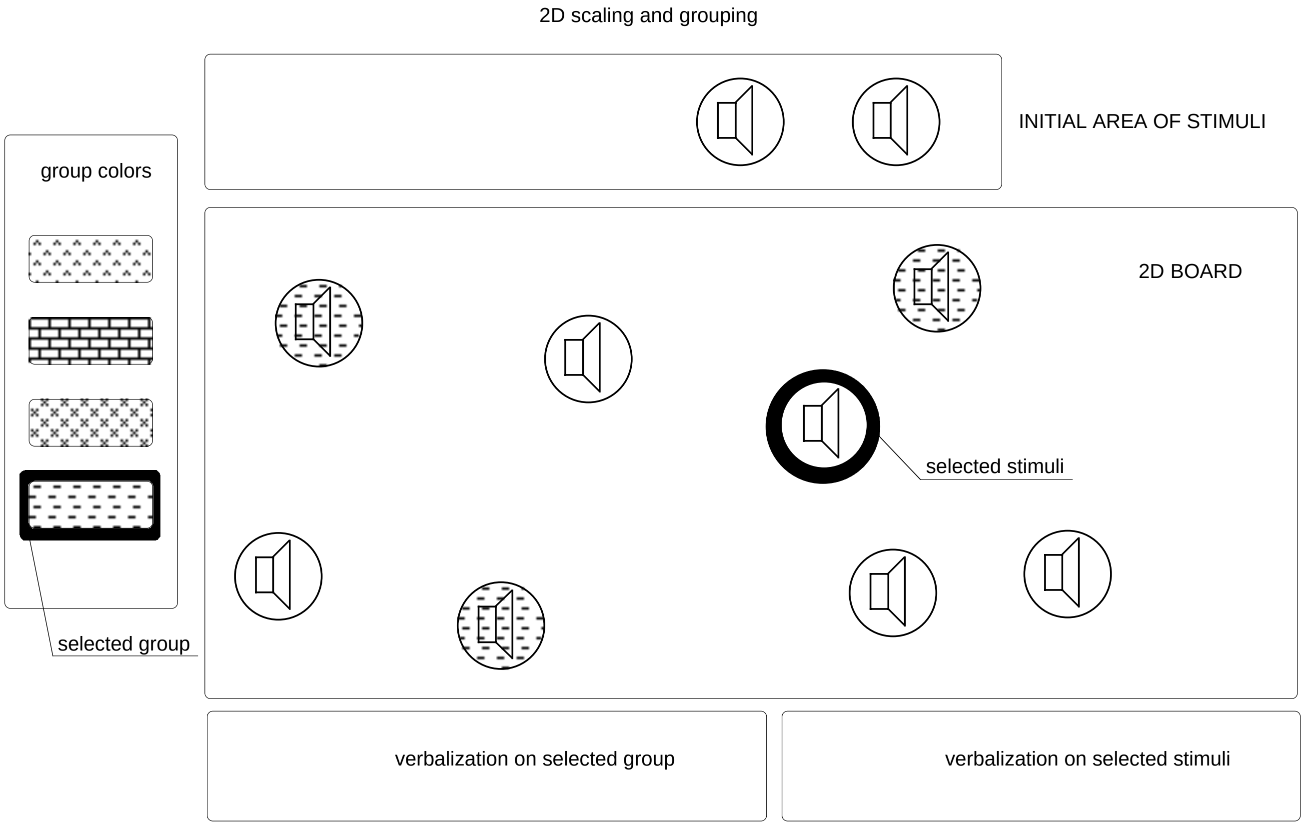


Figure 2. Layout for a 2D scaling and grouping test administered via computer. By selecting a color/pattern, subjects can define groups or categories of stimuli. Within the whole set of stimuli or for a particular category, distances between stimuli reflect stimuli inter-differences. Verbalizations on both categories and stimuli are easily and naturally added during the process of grouping or scaling.

This test has been designed so it can be used for ranking, rating (or scaling), grouping (or categorization) and the specification of verbalizations. Scaling can be made across the whole set of stimuli but also in a specific group.

A full listening-test environment was programmed in Matlab™, allowing the integration of different kind of tests in sequence (see Rioux, 2000a, 2000b).

A rather similar method has been used for scaling purposes by Bodden et al. (1998), but no analysis was shown to have been used or documented. Guyot (1996) and Maffiolo (1999) used a similar approach restricted to the study of categorization, applying tree analysis methods (Guénoche, 1998) to constructed dissimilarity matrices according to the model of Tversky (1977). The problem of analyzing 2D configurations given by a 2D scaling test, as shown in Figures 3a, 3b and 3c, is tackled in the next section.

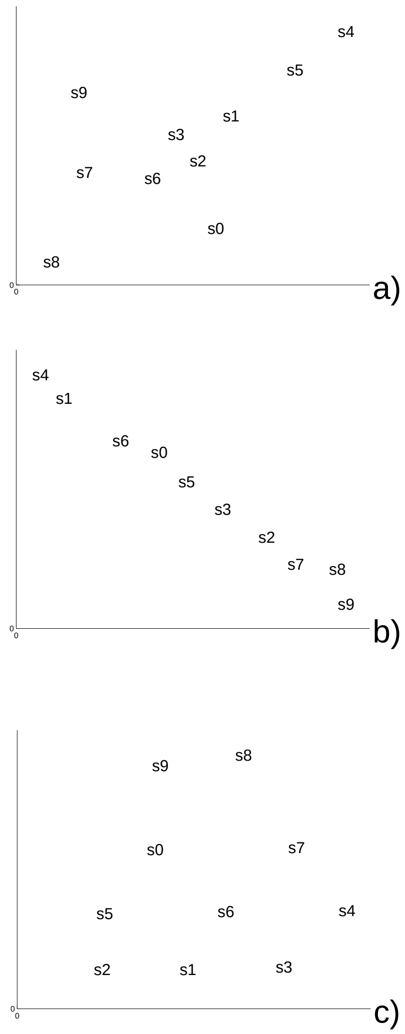


Figure 3. Examples of different output configurations of a 2D scaling test.

a) Configuration of the organ-builder
b) Configuration of a subject using mainly a diagonal
c) Configuration of a subject using the whole board

4. STATISTICAL ANALYSIS

Statistical analysis was certainly the most crucial point for the development of this method. As we said previously, leaving some freedom to the participants increases the burden on the analyst. A 2D scaling method is not standard and had to receive a special treatment.

Figures 3a, b and c demonstrate different tactics used by different subjects in order to scale stimuli on a 2D board. At first glance, these three persons seem to have had a very different perception of these stimuli. Obviously, a geometric comparison of the patterns formed by such different configurations should include mathematical transformations such as translations, rotations, and dilatations. The complexity of such a combination of transformations is reflected in the fact that we found no direct geometrical way to compare these patterns to one another or to exhibit a "mean" pattern. But under the assumption that subjects respected the assignment and used relative distances to organize their own stimuli pattern, we can transform the raw individual matrices of coordinates of 2D space into a dissimilarity matrix containing Euclidean distances without modifying the perceptual meaning of the patterns. Mathematical notations and procedure are shown in Figure 4.

$n,$ number of subjects (observations)

$p,$ number of stimuli (variables i and j)

$$\underset{(n\times p\times 2)}{\mathbf{X}} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_n \end{bmatrix}, \text{ raw matrix (2modes 3 way)}$$

$$\underset{(p\times 2)}{\mathbf{X_1}} = \begin{bmatrix} \mathbf{x}_{11} \\ \vdots \\ \mathbf{x}_{1p} \end{bmatrix} = \begin{bmatrix} x_{11} & y_{11} \\ \vdots & \vdots \\ x_{1p} & y_{1p} \end{bmatrix}, \text{ individual matrix of coordinates (subject \#1)}$$

$$d_{1\text{ij}} = \left\|\mathbf{x}_{1i} - \mathbf{x}_{1j}\right\| = \sqrt{(x_{1i} - x_{1j})^2 + (y_{1i} - y_{1j})^2}, \text{ distance between stimuli } i \text{ and } j \text{ (subject \#1)}$$

$$\underset{(p\times p)}{\mathbf{D_1}} = \begin{bmatrix} 0 & & & \\ d_{121} & \ddots & & \\ \vdots & \ddots & \ddots & \\ d_{1p1} & \cdots & d_{1p(p\text{-}1)} & 0 \end{bmatrix}, \text{ individual triangular matrix of Euclidean distances (subject \#1)}$$

$$\underset{(n\times p\times p)}{\mathbf{D}} = \begin{bmatrix} \mathbf{D}_1 \\ \vdots \\ \mathbf{D}_n \end{bmatrix}, \text{ matrix of Euclidean distances (dissimilarity, 3 ways 2 modes)}$$

$$\Delta = <\underset{(p\times p)}{\mathbf{D}}> = \frac{1}{n}\sum_{k=1}^{n}\mathbf{D_k}, \text{ mean matrix (2 ways 2 modes) of proximity}$$

Figure 4. Notations and procedure used to produce space distance matrices and mean distance matrix from space coordinates.

We are concerned here with the analysis of dissimilarity matrix 3 ways 2 modes (see Fig. 5).

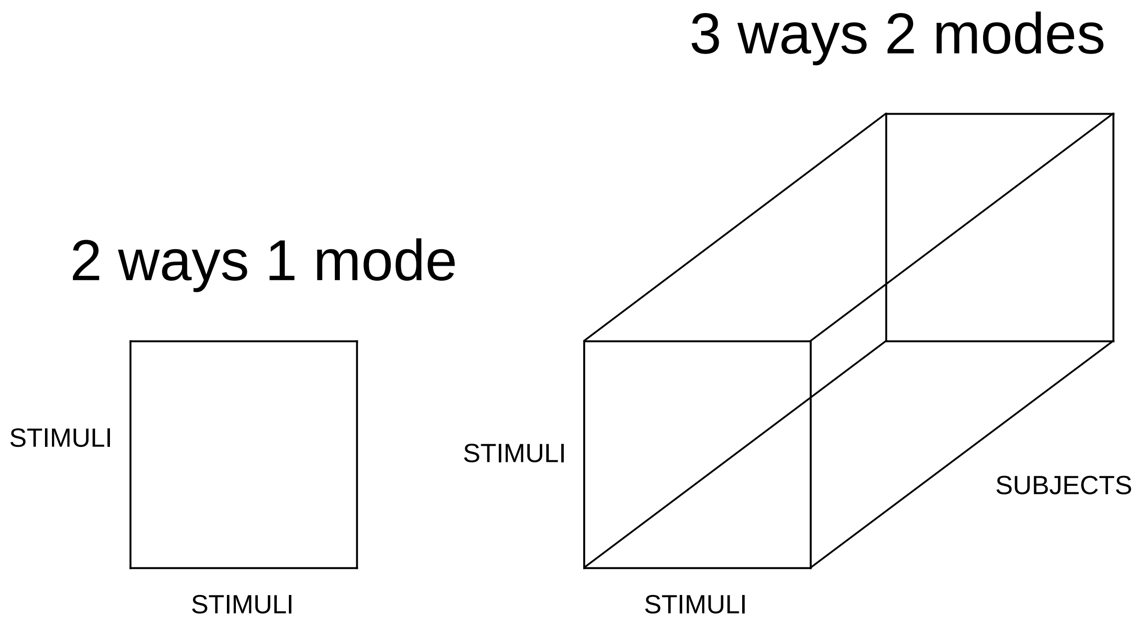


Figure 5. Organization of data matrices.

The analysis of 2 ways 1 mode matrices is straightforward using Torgerson's algorithm (cf. Appendix). 3 ways 2 modes can be performed using INDSCAL. An alternative is to obtain a mean matrix of the 3 ways 2 modes and apply the previous Torgerson algorithm. A problem posed in this article is to compare the data of a single subject (2 ways 1 mode) with the data of a group of subjects (3 ways 2 modes).

This analysis was handled by ALSCAL (see Gärling et al., 1989; Kruskal 1978; Coxon, 1982) a variant of INDSCAL, implemented in the SPSS™ package (for a review, see Gabrielsson 1974). For comparison, we also implemented a version of Torgerson's (1958) algorithm and performed it on the mean matrix <**D**> of distances (see Appendix for the code).

5. EXPERIMENT

5.1. Experiment description

Interdisciplinary research on human perception of musical sounds led us to prepare the following experiment. Ten sounds of flue organ pipes were recorded. Stimuli are short (about 1 second) and very similar, compared with a typical palette of timbre of, say, an orchestra. We were interested in studying the correlation between the perception of these stimuli, the description of this perception (through verbal comments) and the objective measurements (signal analysis and acoustics of the instrument). Moreover, the recorded pipes were voiced by an expert organ-builder, and it is definitely of major interest to be able to compare his perception with other groups of subjects. Twenty-two subjects and the organ-builder were then asked to perform a sequence of tests using a computer interface, including a 2D scaling/grouping test as described above and a pair comparison as well (see Table 1). This kind of test sequence is implemented in Matlab, and sketched in Appendix 2. The group of 22 subjects consisted of professional musicians and musicologists who were trained organ music listeners and were mostly susceptible to sharing their auditory impressions with an organ-builder.

Instructions for the 2D scaling test:

Subjects were asked to compare all sounds according to their global difference and to use the full width of the screen as much as possible. They were told that they could organize the stimuli in whatever pattern they wanted as long as the inter-stimuli distances reflected the amount of dissimilarity between stimuli.

5.2. Experimental results

It took between 30 minutes and one hour for each subject to perform these multiple comparisons. Some reported that it was demanding but not as boring as the pair comparison test, and very few non-computer-experienced persons had problems with the Graphical User Interface.

A sample of obtained patterns has already been shown in Figure 3. The matrices **D** and <**D**> (see Fig. 4) were constructed from this type of data. **D** was treated with ALSCAL and <**D**> with an algorithm from Torgerson detailed in Appendix 1.

For a comparison of these methods, see Figure 6.

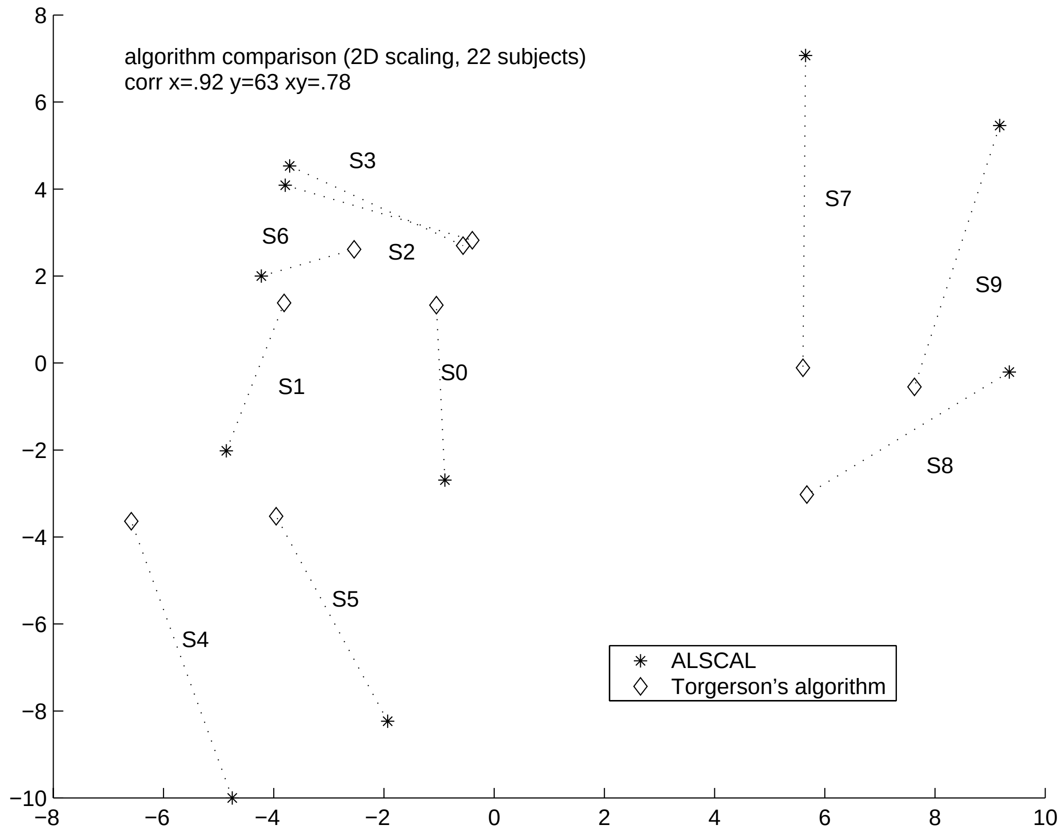


Figure 6. A comparison of two analytical methods applied to the same data. "Stars" represent an MDS analysis made with ALSCAL while "diamonds" symbolize the use of Torgerson's algorithm.

Multidimensional Scaling is projected on its first two dimensions (ALSCAL gives a squared correlation of .38). It should be stressed that in this case we are not particularly interested in the actual meaning/physical correlates of each dimension but more by finding a mean pattern of stimuli on the first two most important dimensions. Inspection of Figure 6 indicates that even when no inter-individual differences are taken into account, formed patterns are similar. The dotted lines of Figure 6 represent the difference in results with both methods for each stimulus. Even if some of these segments are rather large compared to the size of screen, they never cross and so confirm their similarity. Three groups of stimuli clearly emerge, (s0, s1, s2, s3, s6), (s4, s5) and (s7, s8, s9).

As subjects performed a pair comparison of the same stimuli, it is thus possible to check the consistency of their answers. An MDS analysis of the pair-comparison data was made and compared to the ALSCAL analysis of the 2D scaling test (see Fig. 7). Obviously, correlation coefficients are high and provide a check of subject responses of surprisingly good consistency. This result can furthermore be regarded as a validation of the new "2D scaling" method presented here compared with the classical pair-comparison method.

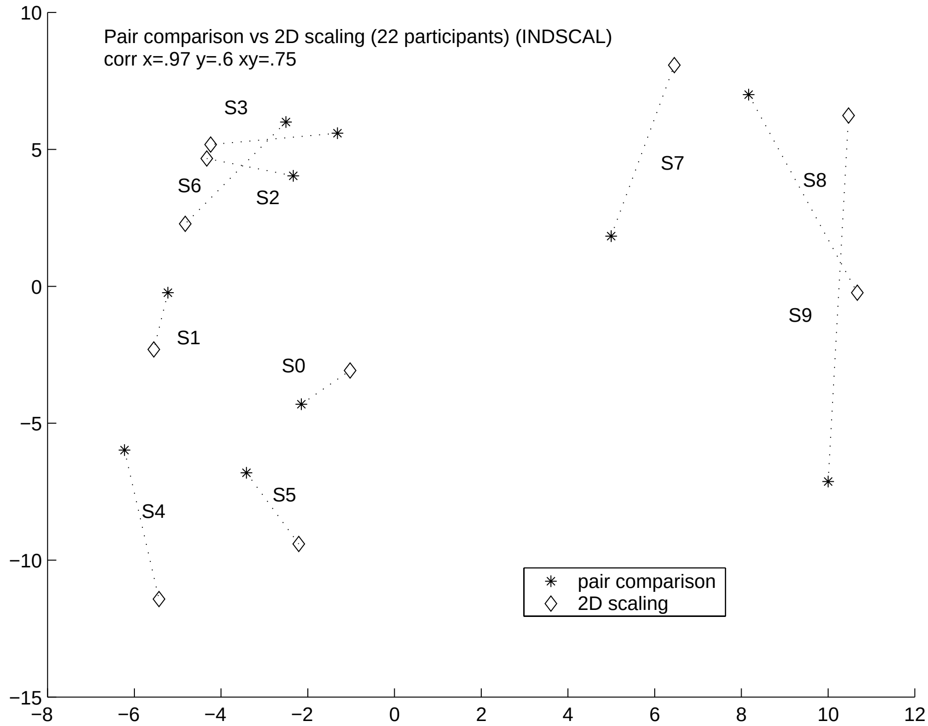


Figure 7. A comparison of two test methods (pair comparison, 2D scaling) of the same stimuli rated by the same subjects.

Moreover, having established that a "mean" 2D pattern of stimuli has been found, we can now examine the mean answer of a group of 22 listeners as compared with the answer of an expert.

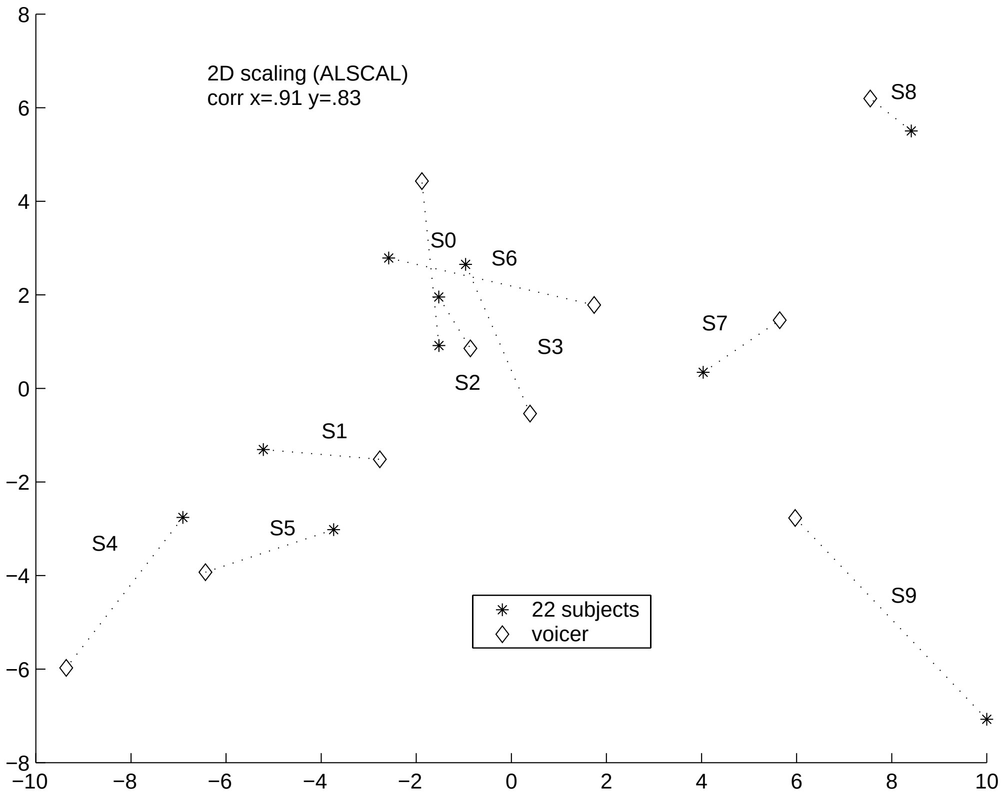


Figure 8. Comparison of a group versus expert response.

This is shown in Figure 8, in which the configuration of stimuli given by the voicer (organ-builder) is plotted against the ALSCAL analysis of the same 2D scaling performed by the group of 22 musicians and musicologists. Correlation coefficients of the two configurations along the two axes (.91 and .83) are very high. Previously suggested groups of stimuli are preserved. These groups were also confirmed by tree analysis (Sattath and Tversky, 1997; Guénoche, 1998) of the same data (see Fig. 9).

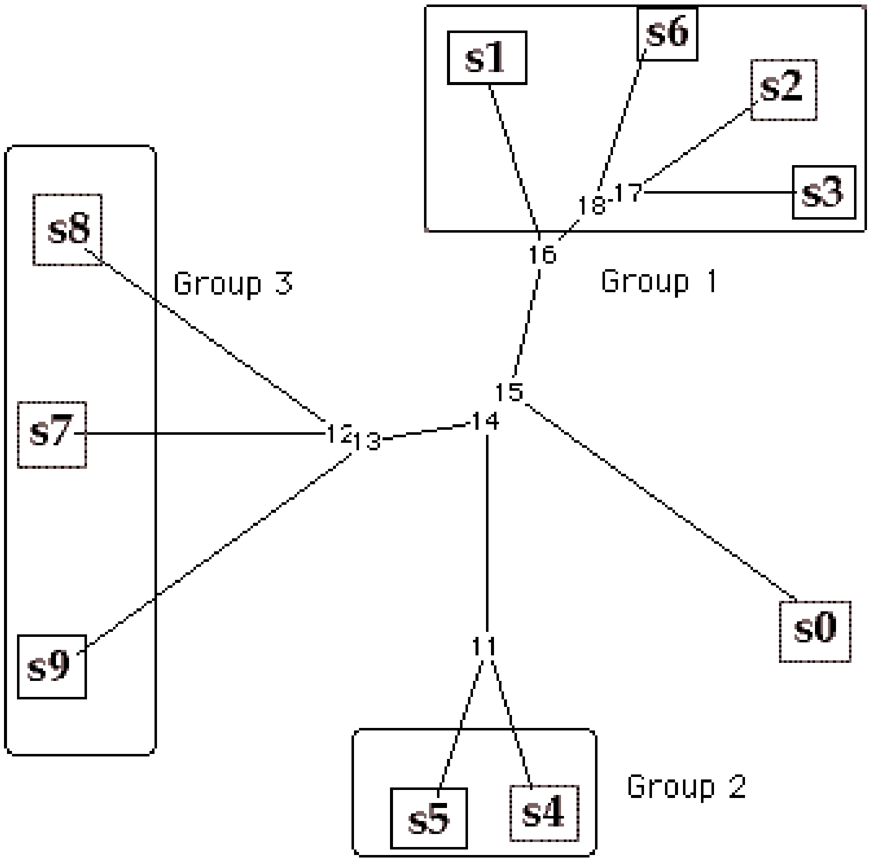


Figure 9. Tree analysis of 22 subjects' answers. The leaves represent stimuli, and branches symbolize their inter-distance.

This was a major step to be able to prove that even though they may use different kinds of wording (Rioux and Västfjäll, 2000c), this group of musicians basically perceived the small differences between these stimuli in the same way as the organ-builder.

5. CONCLUSIONS

We presented a new method for the study of perception of a set of auditory stimuli. This method is based on previous research made in the field of musical acoustics, psychoacoustics and sound quality. It is unique in that it offers the possibility of combining the advantages of methods based on multiple comparisons, categorization, and verbalization and avoids the drawbacks of pair comparisons and semantic differentials. In this method, stimuli are displayed on a computer screen and subjects can scale, group and comment on them in a very flexible way. In order to interpret the 2D scaling test, an analysis method was derived, based on Multidimensional Analysis techniques.

We also presented an experiment performed on 22 subjects listening to 10 stimuli. This type of test proved to be very useful for checking individual perception as well as group perception against the perception of an expert. We showed an example of 2D scaling without specifying any interpretation of the underlying dimensions of the two-dimensional space, asking participants to focus only on the organization of distances between objects. But of course, when stimuli differ in only two variables, dimensions could be prescribed in advance (e.g. for the study of an emotion circumplex, see Västfjäll 2000).

This method proved to be valuable for exploratory research on auditory stimuli. We used it mainly for multidimensional stimuli differing in small amounts, but it could be applied to a more extended set of stimuli, especially if used in conjunction with a categorization/grouping task. This type of test could fit into every exploratory auditory research study concerned with complex multidimensional stimuli, like the perception of musical sounds, for instance, or the sound quality of industrial products.

REFERENCES

Bodden, M., Heinrichs, R., and Linow, A. (1998). “Sound Quality Evaluation of Interior Noise Using an Efficient Psychoacoustic Method.” Proceedings of Euro Noise, München.

Coxon, A. P. M. (1982). *The User’s Guide to Multidimensional Scaling*. London: Heinemann Educational Books, Ltd.

David, H. A. (1969). *The Method of Paired Comparisons*. London: Charles Griffin & Company, Ltd.

Gabrielsson, A. (1974). “An Empirical Comparison Between Some Models for Multidimensional Scaling.” *Scandinavian Journal of Psychology*: 73-80.

Grey, J. M. (1977). “Perceptual Evaluation of Synthesized Musical Timbres.” *JASA*, **61**: 1270-1277.

Guénoche, A. (1998). “Ordinal Properties of Tree Distances,” *Special Issue of Discrete Mathematics on Discrete Metric Spaces*, **192**:103-117.

Guyot, Frédérique (1996). *Étude de la perception sonore en termes de reconnaissance et d'appréciation qualitative: une approche par la catégorisation*, Ph.D. Thesis, Université du Maine.

Gärling, T., and Arce, C. (1989). “Multidimensional Scaling.” *Anuario de psicología*: **43** (4).

Harley, E., and Loftus, G. (2000). “MATLAB and Graphical User Interfaces: Tools for Experimental Management.” *Behavior Research Methods, Instruments, & Computers*, **32** (2): 290-296.
Hartmann, W. M. (1997). *Signals, Sound, and Sensation*. Woodbury, NY: American Institute of Physics. 61-63.
Kruskal, J. (1978). *Multidimensional Scaling*. Murray Hill: Sage University.
Levin, D. N. (2000). “A Differential Geometric Description of the Relationships among Perceptions.” *Journal of Mathematical Psychology,* **44**: 241-284.
McAdams, S. (1999). “Perspectives on the Contribution of Timbre to Musical Structure.” *Computer Music Journal*, **23:3**: 85-102.
Maffiolo, V. (1999). *De la caractérisation sémantique et acoustique de la qualité sonore de l'environnement urbain*. Ph.D. Thesis, Université du Mans.
Parizet, E., and Nosulenko, V. (1999). “Multi-dimensional Listening Tests: Selection of Sound Descriptors and Design of the Experiment.” *Noise Control Engineering Journal*, **47** (6): 227-232.
Rioux, V. (2000a). “Methods for an Objective and Subjective Description of Starting Transients of Some Flue Organ Pipes—Integrating the View of an Organ Builder.” *Accepted for publication in Acustica, to be issued in Autumn 2000*.
Rioux, V. (2000b). Matlab routines and additional data used in this article are provided at:
http://www.ta.chalmers.se/homepages/vincent/index.html
Rioux, V. & Västfjäll, D. (2000c) “Verbal Description of Flue Organ Pipe Sounds (I) and (II).” *Proposed for publication.*
Samoylenko, Elena, McAdams, Stephen and Nosulenko Valery (1996). “Systematic Analysis of Verbalizations Produced in Comparing Musical Timbres.” *International Journal of Psychology*, **31** (6): 255-278.
Sattath, S. and Tversky, A. (1977). “Additive Similarity Trees.” *Psychometrika*, **42** (3): 319-345.
Torgerson, S. W. (1958) *Theory and Methods of Scaling*. Florida: Krieger R. Publishing Company. 247-298.
Tversky, A. (1977). “Features of Similarity.” *Psychological Review*, **84** (4): 327-352.
Von Bismarck, G. (1974). “Timbre of Steady Sounds: A Factorial Investigation of its Verbal Attributes.” *Acustica,* **30**: 146-158.
Västfjäll, D., Friman, M., Gärling, T., & Kleiner, M. (2000). The measurement of core affect: A Swedish self-report measure derived from the affect circumplex. In press. *Scandinavian Journal of Psychology*.

APPENDIX I: MATLAB CODE FOR TORGERSON ALGORITHM

```
% TRIG contains the triangular similarity matrix
[l,c]=size(TRIG);

% make O a full squared matrix from m
SQUA=TRIG;
for nl=1:l-1
        for nc=nl+1:c
                SQUA(nl,nc)=SQUA(nc,nl);
        end
end

% Calculate SCAL pseudo scalar products and center it (around the center
of gravity)

SQUA2=SQUA.^2;
[l,c]=size(SQUA2);
l2=''; c2='';
for lno=1:l
        l2(lno,:)=-mean(SQUA2,1);
end
for cno=1:l
        c2(:,cno)=-mean(SQUA2,2);
end

SCAL=.5*(SQUA2+l2+c2+mean(mean(SQUA2)));

% Calculate the coordinates X by singular value decomposition of SCAL
[U,V,Up]=svd(SCAL);
X=U*V^.5;
```

APPENDIX II: MODULAR CODE OF LISE SYSTEM

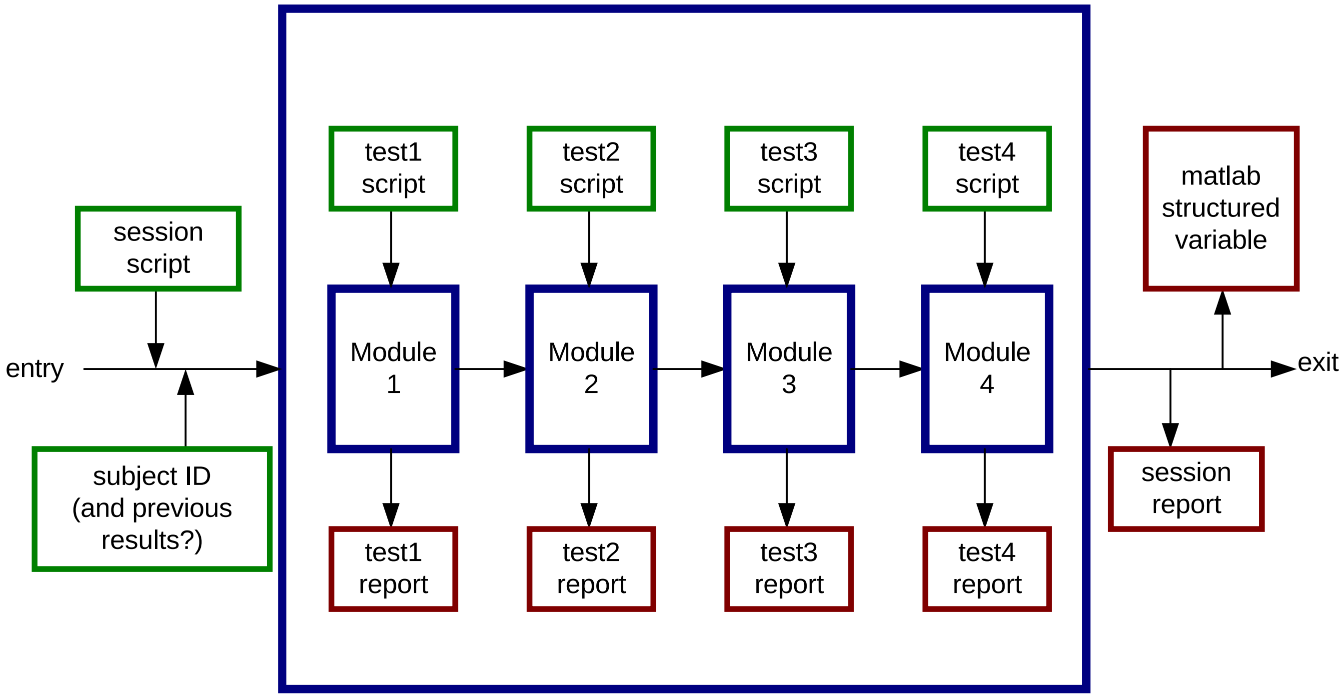


Figure 10. Flow chart of a test-session specified by the LISE (Listening Interface for Sound Experiments) package written in Matlab (see Rioux 2000b). Each test-session consists of a sequence of modules which can for example be a 2D scaling test followed by a pair comparison test.

APPENDICES

Appendix I : Construction of the experimental pipe

In order to study, in a controlled way, the geometrical modifications of the mouth that occurs during the voicing process, an experimental pipe with interchangeable parts was designed. This pipe was made of Plexiglas in order to allow visualization with Schlieren techniques. A model was first made in wood (see the two pictures below).

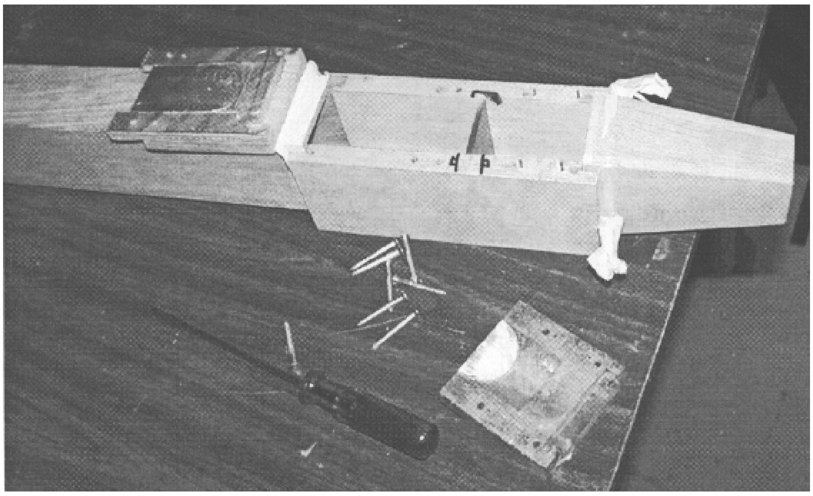


Figure 42. From right to left: the foot, languid, and upper-lip turned upside-down.

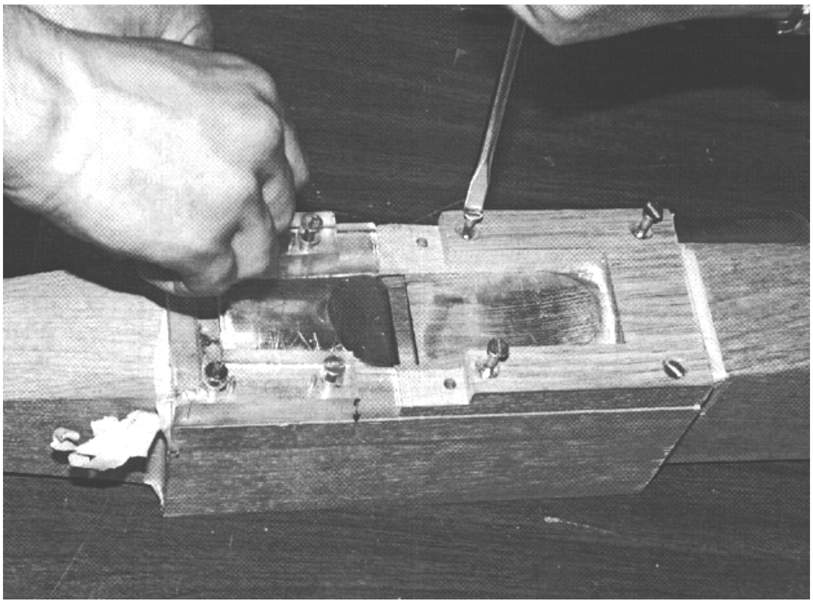


Figure 43. The mouth: here, interchangeable parts are screwed to make a complete mouth.

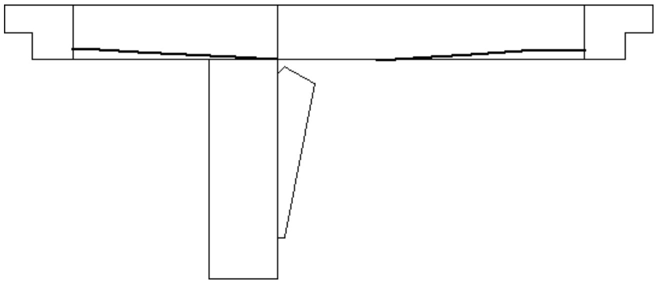


Figure 44. Schematic representation of the mouth shape. The lower-lip (left), upper-lip and languid are glued on Plexiglas frames which can be retrieved from the pipe structure.

Note that the geometry of the mouth is the one of a typical metallic and cylindrical pipe, whereas the geometry of the body is the one of a typical wooden pipe (squared). This type of pipe is indeed a unique piece and it was a not without a certain surprise that we could verify that this type of pipe could produce an acceptable sound (in terms of organ building). The construction of the experimental pipe was then acknowledged and started according to a reviewed design (see sketches below) in a workshop which could work with Plexiglas[13].

Moreover two holes for the positioning of the microphones were chosen. The positions were calculated with the equations presented in chapter III (Measurements in passive mode) and chosen to be x1=71mm x2=511mm, so that the pressures at the microphones would take most of the harmonics. This is shown on Figure 45.

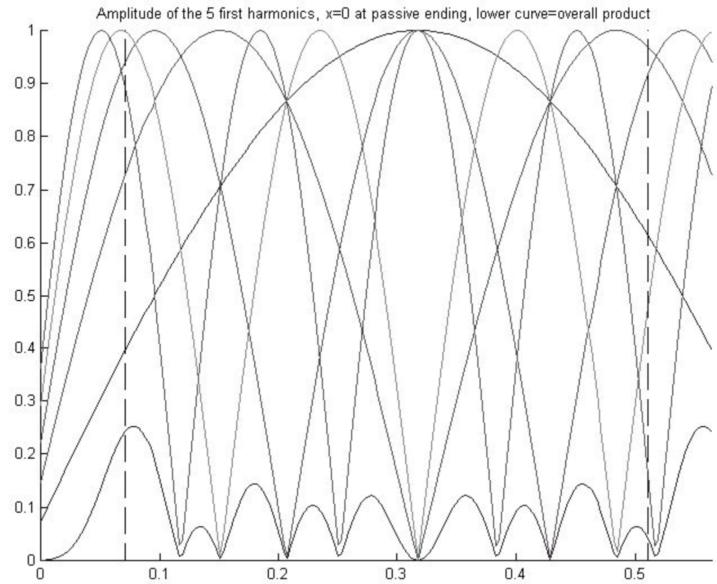


Figure 45. Amplitude of the 5 first harmonics with varying position in the body (x-axis in meters). The positions of the two microphones are represented by the dotted lines.

[13] Televerkstaden, Chalmers University Of Technology, Göteborg.

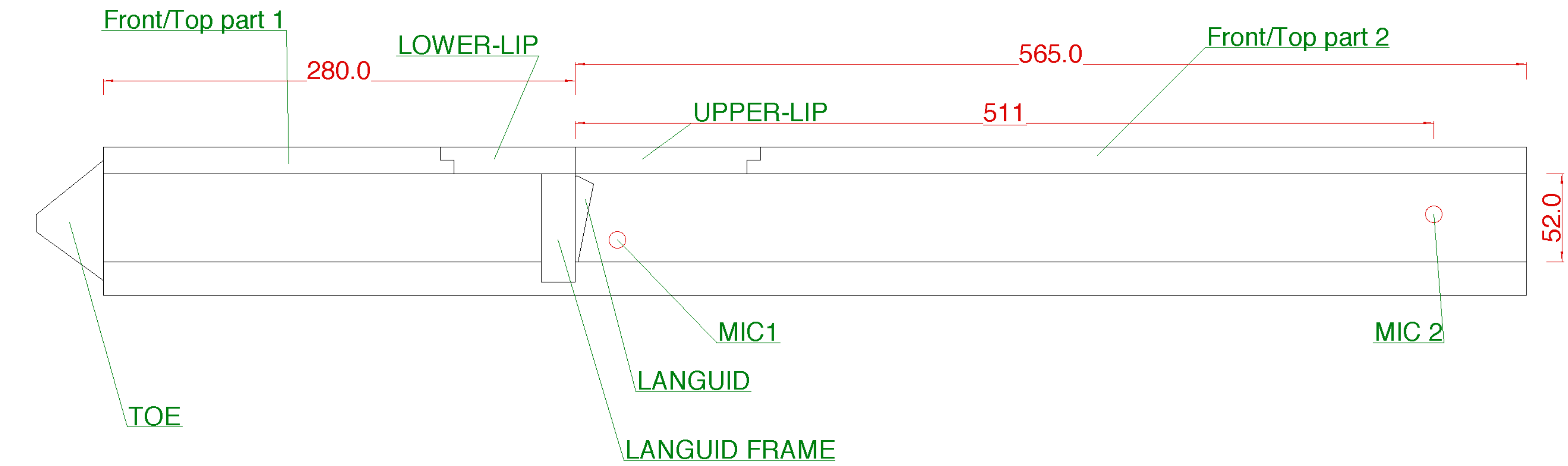


Figure 46. Sketch of the experimental pipe (in total, and horizontally placed) seen from its left side.

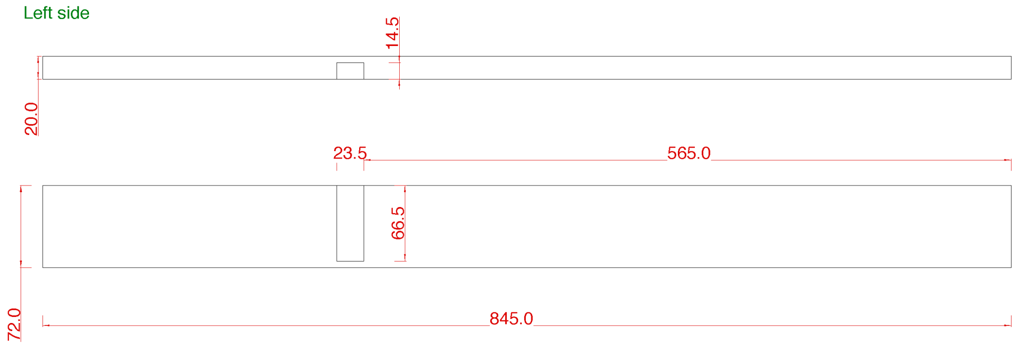


Figure 47. Left side plate.

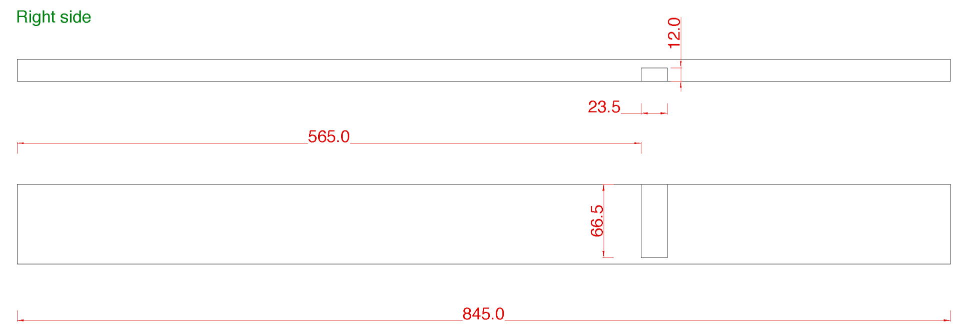


Figure 48. Right side plate.

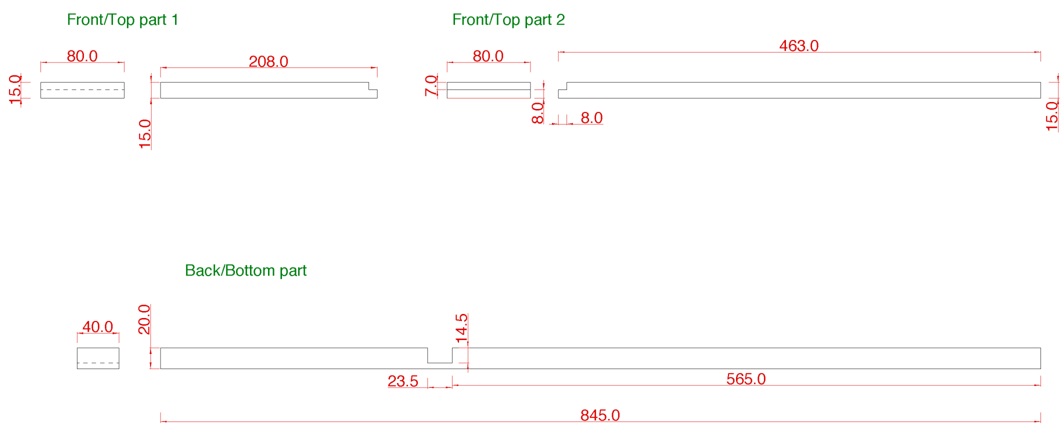


Figure 49. Top and bottom parts.

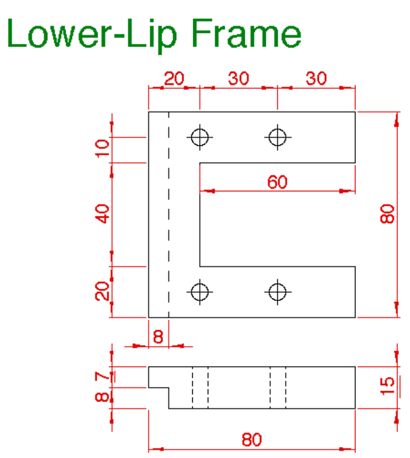


Figure 50. Lower-lip frame.

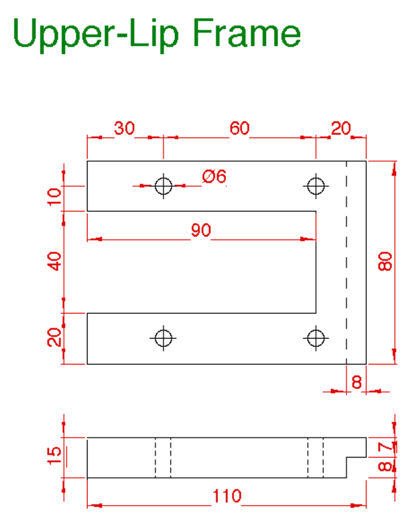


Figure 51. Upper-lip frame.

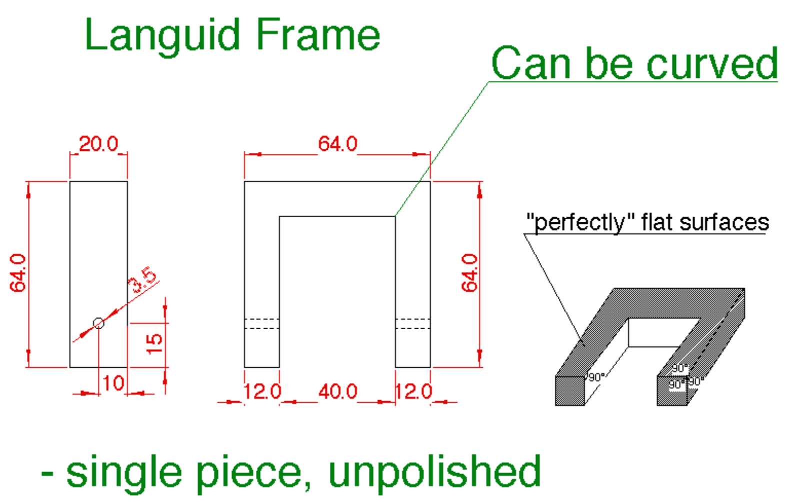


Figure 52. Languid frame

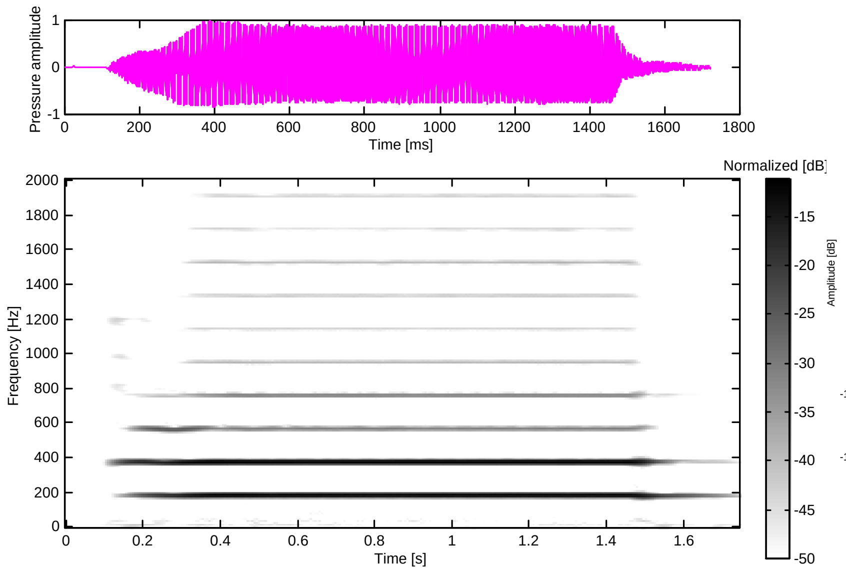
Appendix II : Pipe dimensions

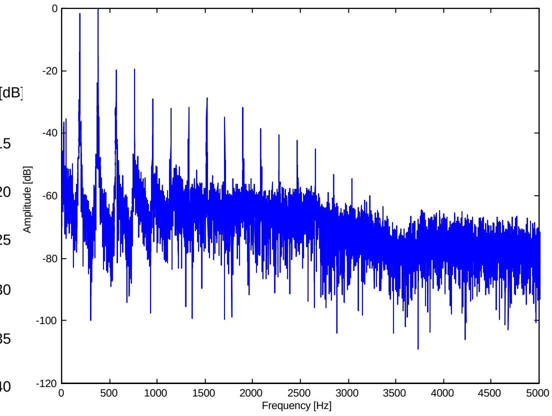
Pipe name	Type of pipe / stop	F0	Body L.	Body Circ. (O.D.)	Foot L.	Foot Ø (O.D.)	Measuring Point @	Toe-hole (I.D.)	Mouth W.	Cut-up	Windway W.	Wall Th. Foot	Wall Th. Body B	Wall Th. Body T
G1	**Oct 8' g**	**412**	369	105.5	195	18.8	185	9.0	24.3	8.2	0.40	1.40	0.90	0.65
G2	**Rauschpfeife 3'choir c°**	**413**	369	105.0	191	17.1	183	8.3	25.5	8.8	0.38	1.10	0.85	0.75
G3	**Oct 4' g°**	**412**	369	105.0	194	14.9	187	8.9	24.9	7.9	0.48	1.14	0.85	0.60
g1	**Oct 4' g**	**1629**	87	43.3	195	10.9	190	6.6	10.3	3.7	0.50	0.77	0.65	0.60
g2	**Rauschpfeife 3'choir c**	**1643**	88	37.2	189	9.8	183	4.9	8.6	3.3	0.35	0.95	0.74	0.60
#1 (s0)	Principal 4 Gs	**191**	731+slot= 1013	180.5	385	16.5	375	8.5	45.1	13.5	0.80	0.80	0.80	0.45
#3 (s7)	Pr 4' Gs	**191**	731+slot= 1016	183	383	17.8	375	9.0	45.3	13.5	0.78		0.80	0.6
#5 (s8	Pr 4' Gs	**192**	730+slot= 1013	184	384	20.2	371	9.7	44.7	13.2	0.77		1.15	1.25
#7 (s9)	Pr 4' Gs	**192**	734+slot= 1020	185	384	17.9	375	8.4	46.2	13.4	0.87		0.75	0.65

L=length, W=width, Circ. =circumference, ø=diameter, Th.=thickness, O.D.=Outside Dimensions, I.D.=Inside Dimensions; B=bottom; T=top. (Dimensions in mm)

Appendix III: Figures from test on separation methods.

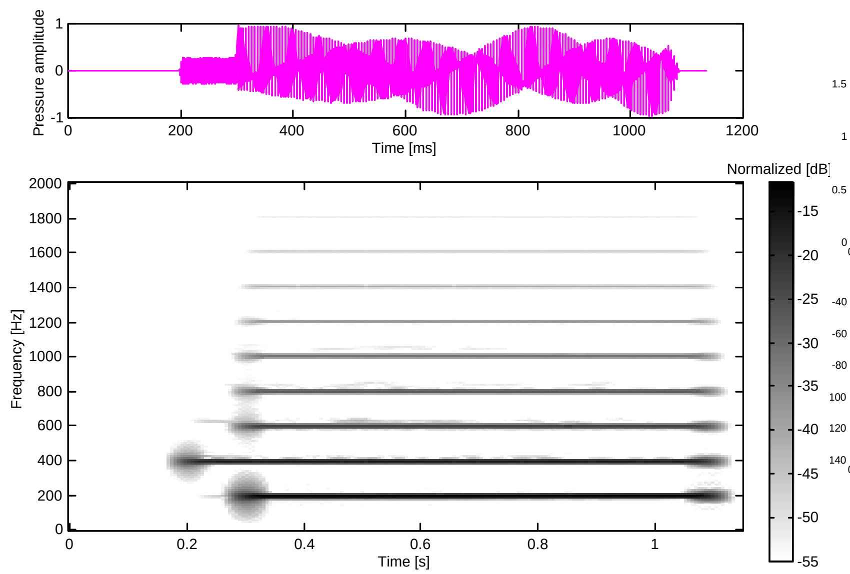
ORIGINAL 'S4'

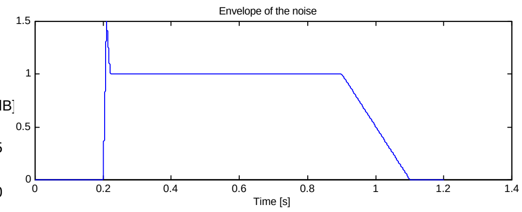


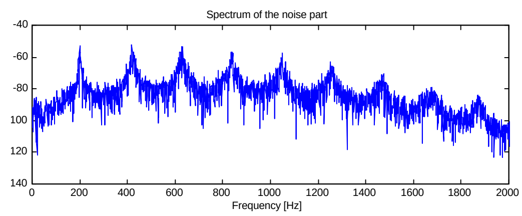


Window : Hamming   FFT size: 92.8798 ms = 4096 pts
DATE: 03-Nov-1998   Overlap: 93.75 % ⇒ fftMove: 256 pts = 5.805 ms

ORIGINAL TEST SIGNAL

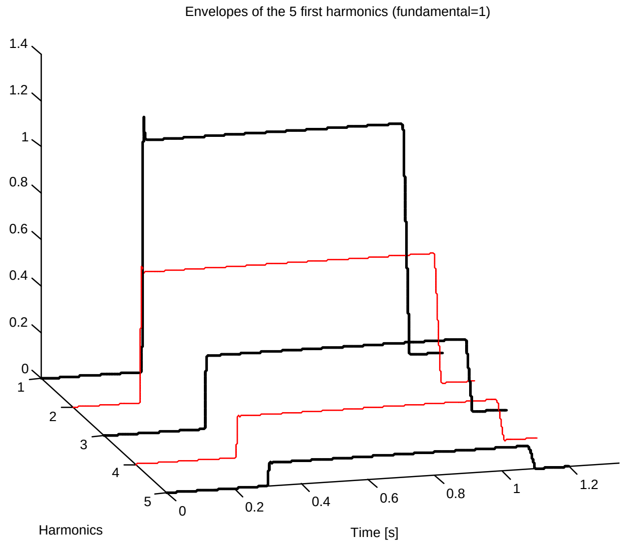




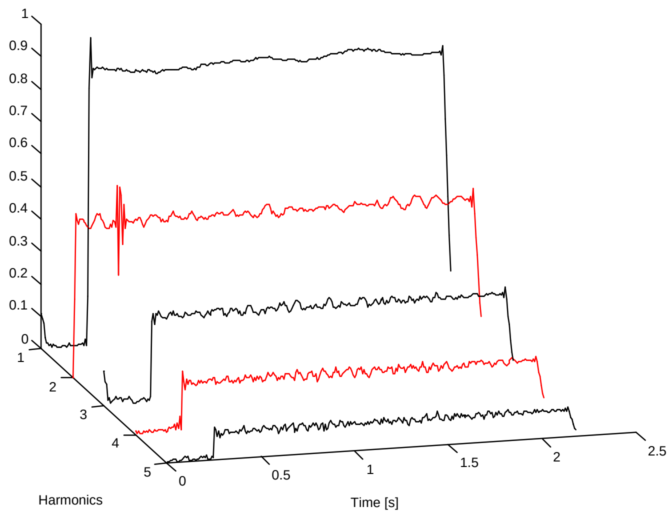


Window : Hamming   FFT size: 92.8798 ms = 4096 pts
DATE: 03-Nov-1998   Overlap: 93.75 % ⇒ fftMove: 256 pts = 5.805 ms

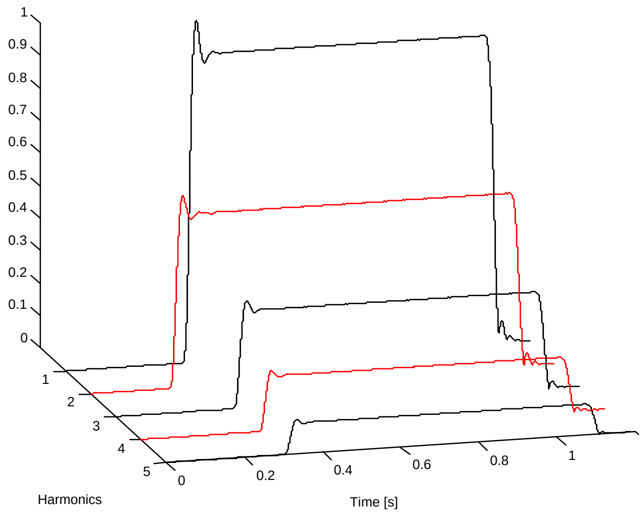
Decomposition of the test signal (I) : Harmonics



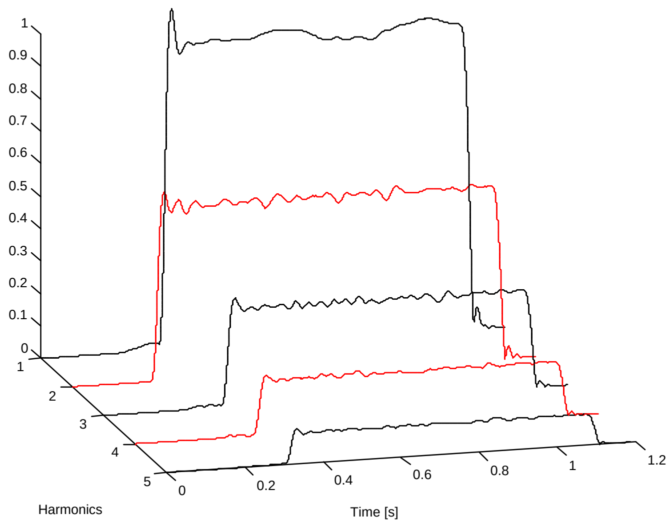
ORIGINAL ENVELOPES (without noise added)



QUASAR ENVELOPES

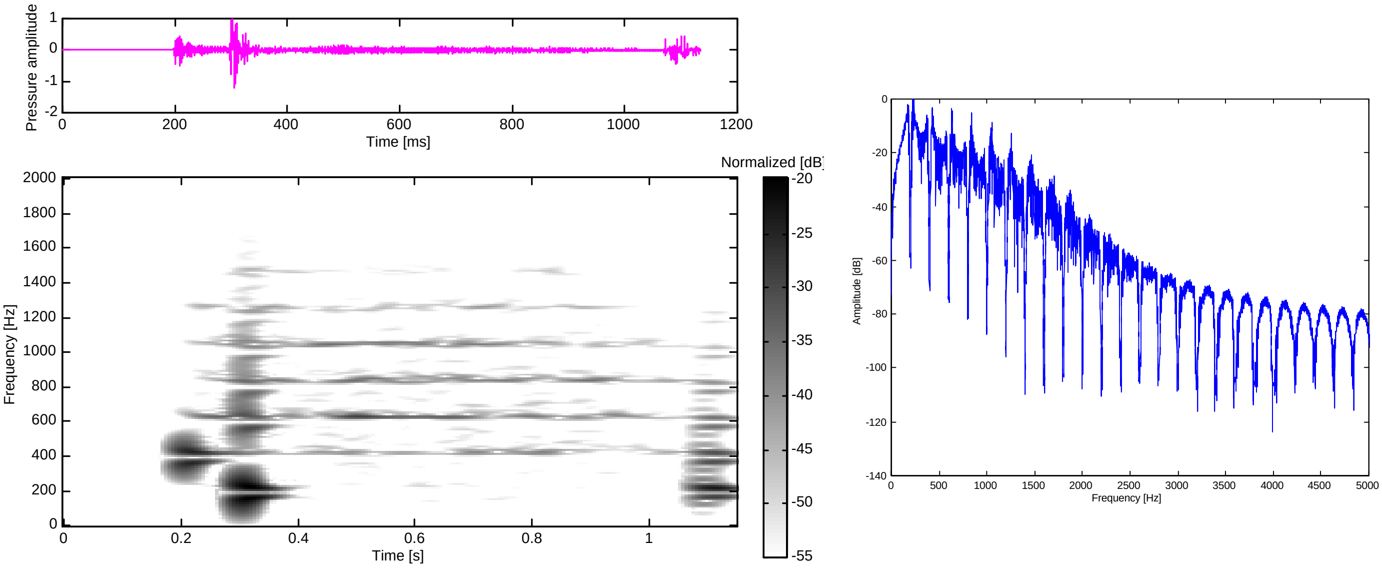


MYFILT ENVELOPES



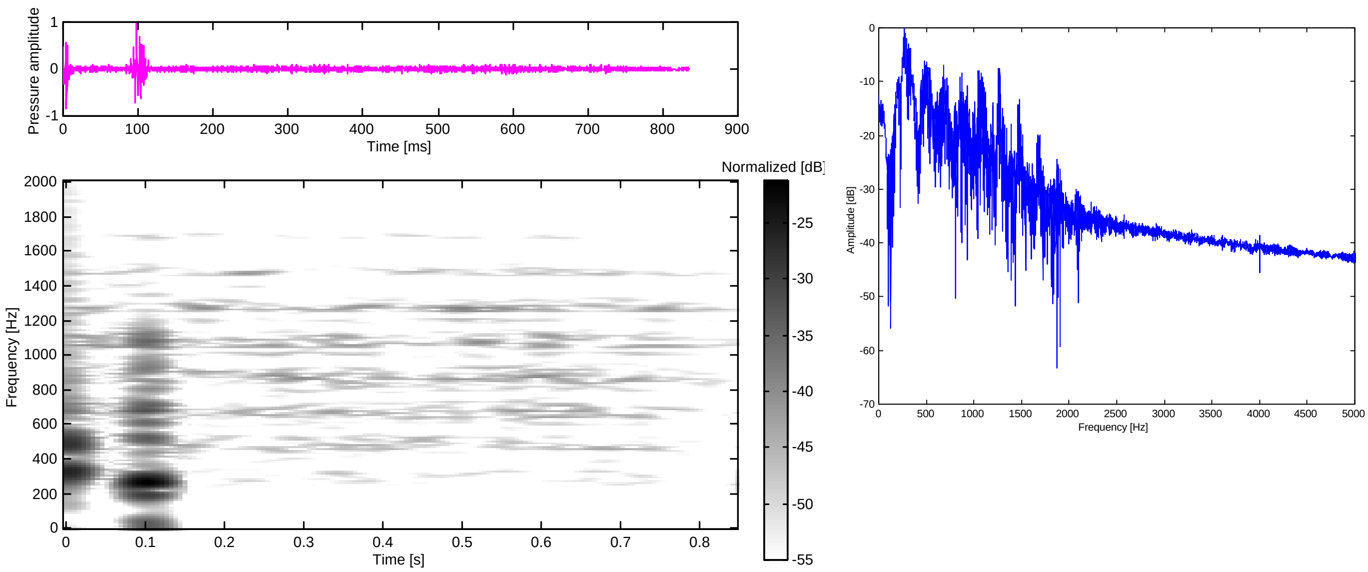
MYFILT ENVELOPES (without noise added)

Decomposition of the test signal (II) : Noise



Window : Hamming   FFT size: 92.8798 ms = 4096 pts
DATE: 03-Nov-1998   Overlap: 93.75 % ⇒ fftMove: 256 pts = 5.805 ms

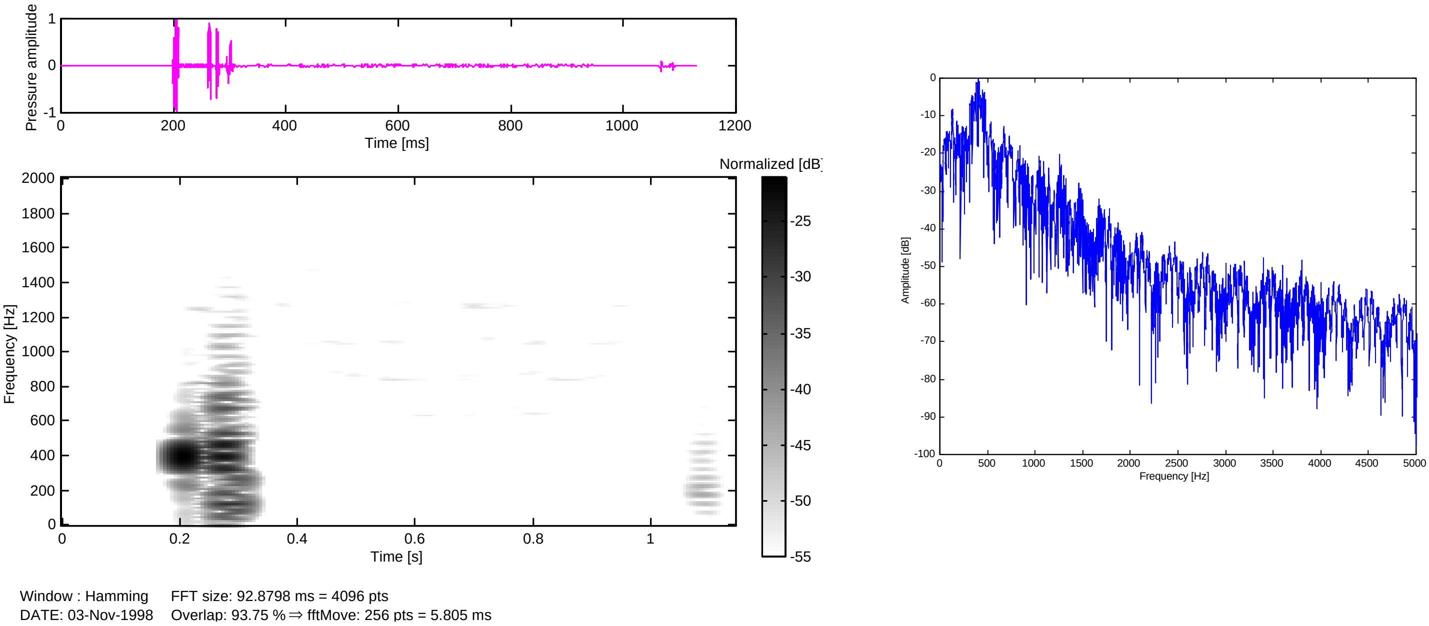
Spectrogram and spectrum of Noise with MYFILT



Window : Hamming   FFT size: 92.8798 ms = 4096 pts
DATE: 03-Nov-1998   Overlap: 93.75 % ⇒ fftMove: 256 pts = 5.805 ms

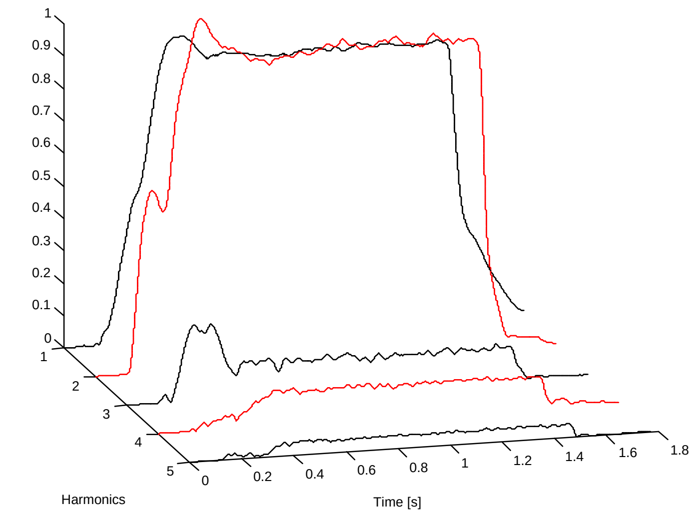
Spectrogram and spectrum of Noise with QUASAR

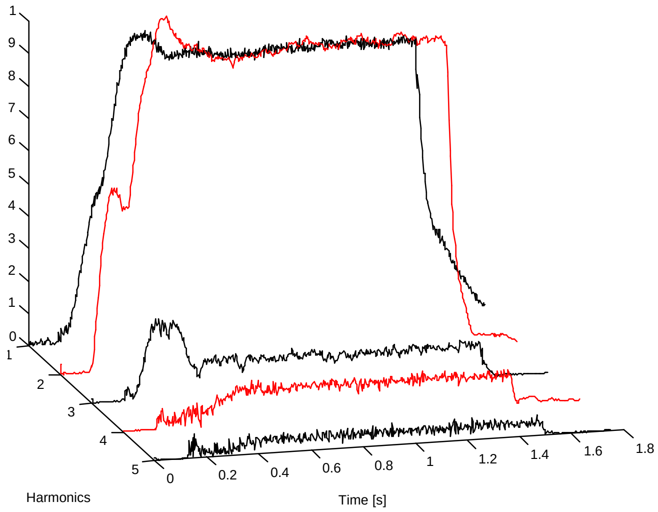
Decomposition of the test signal (II) : Noise



Spectrogram and spectrum of Noise with SMS

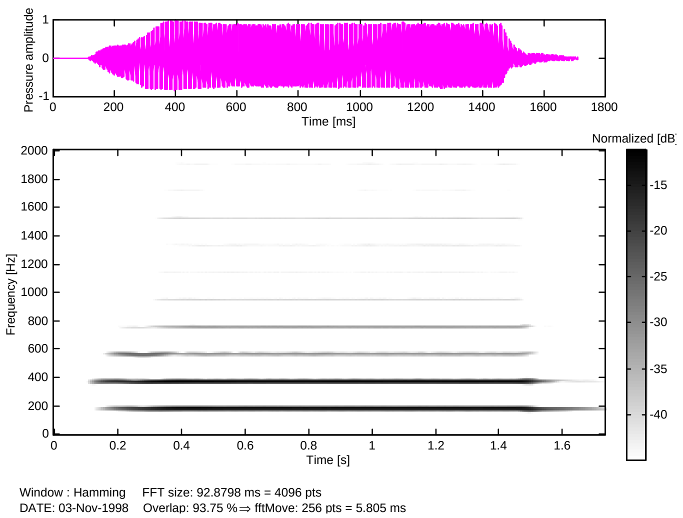
Decomposition of the sound 's4'(I) : HARMONICS

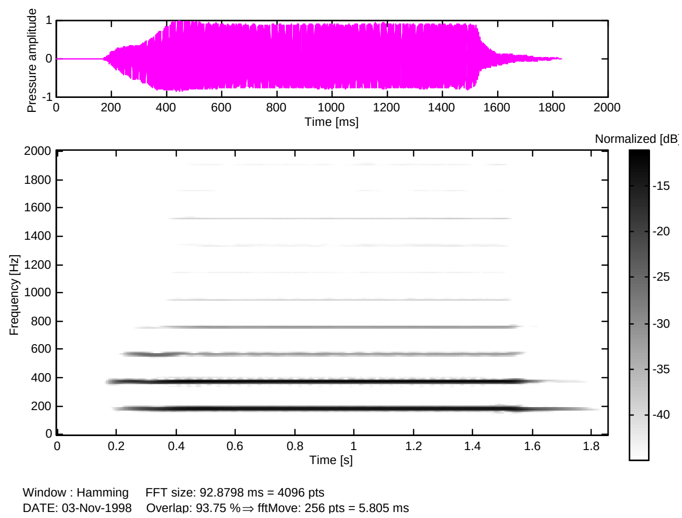




MYFILT ENVELOPES

QUASAR ENVELOPES

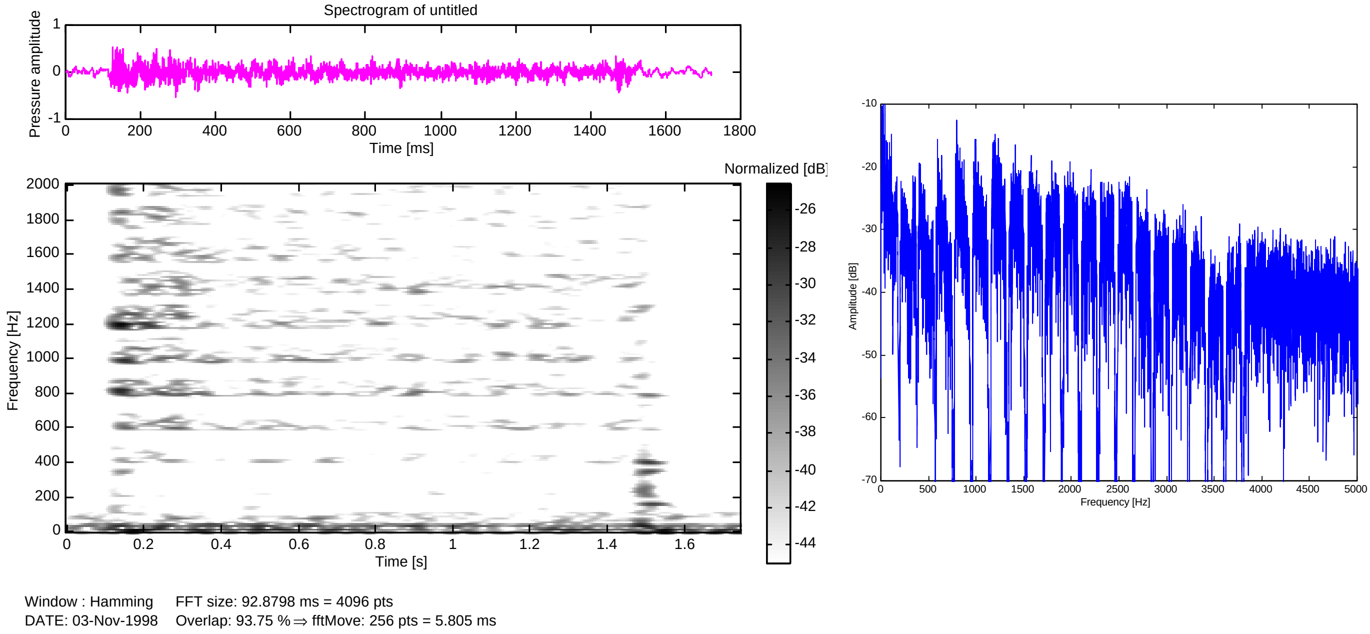




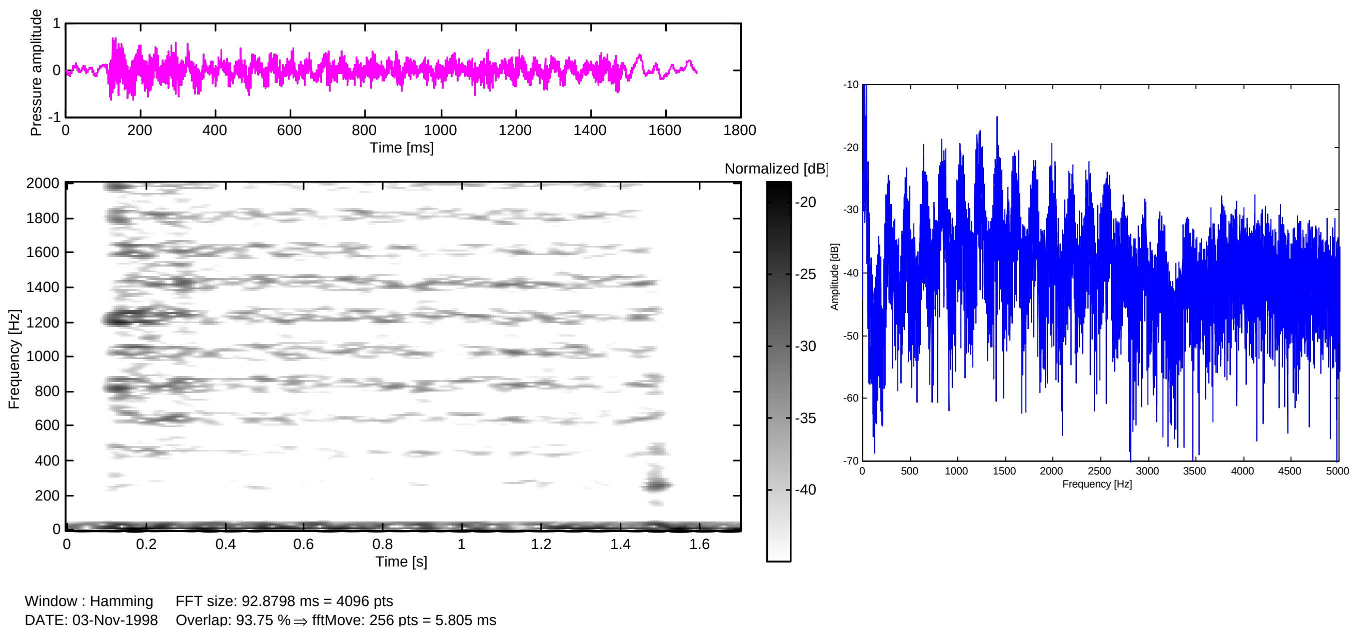
Spectrogram with SMS method

Spectrogram with K.G. method

Decomposition of the sound 's4'(II) : NOISE

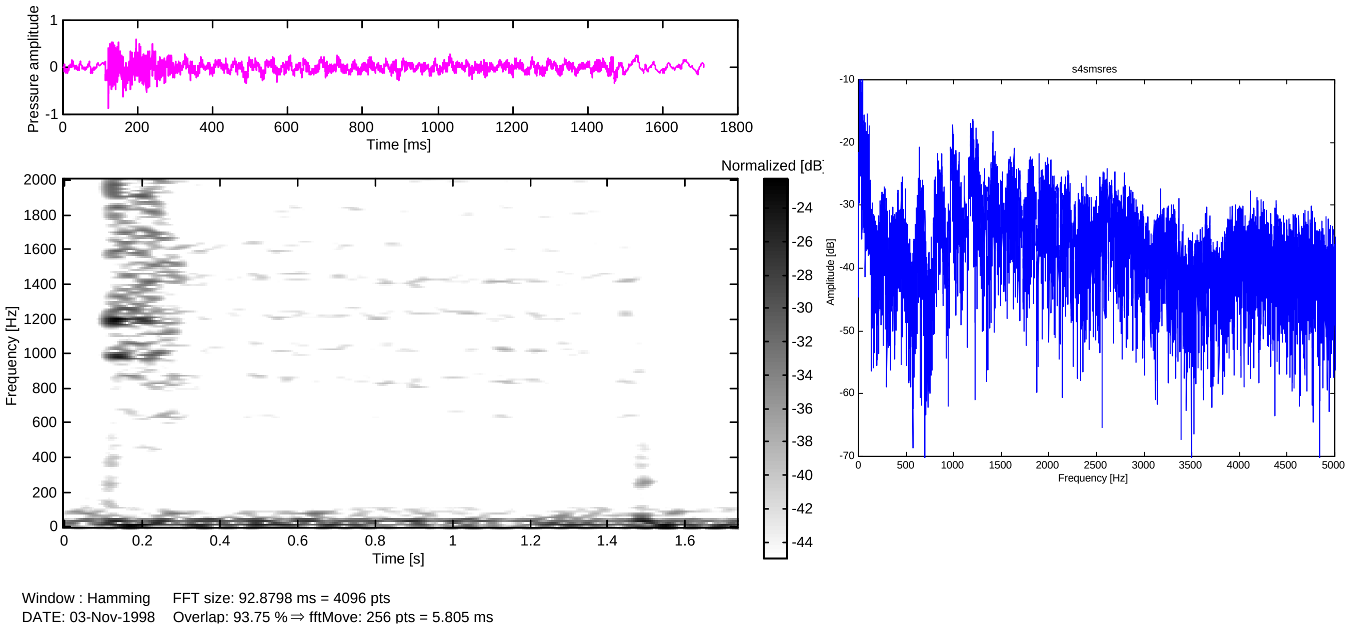


Spectrogram and spectrum of Noise with MYFILT

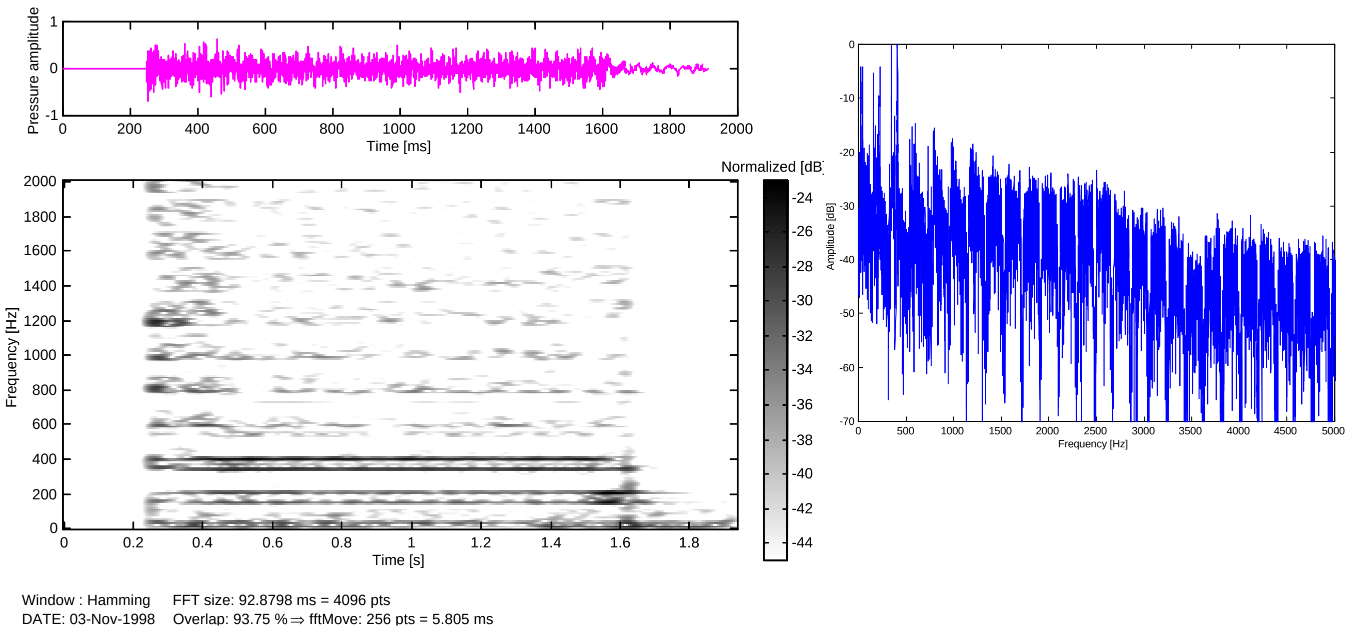


Spectrogram and spectrum of Noise with QUASAR

Decomposition of the sound 's4'(III) : NOISE



Spectrogram and spectrum of Noise with SMS



Spectrogram and spectrum of Noise with K.G. method

Appendix IV: Commented list of verbal descriptors.

Part of the sound	verbal descriptors	physical nature				verbal nature		
		noisy	tonal	fluctuation	source	RELATED	SYNONYMS	ANTONYM
Transient	**"chi-ff"**							
	"cou-gh"							
	"hi-ss"							
	"ki-ss"							
	spit (sounds like)							
	aggressive					strong		gentle, weak
	gentle					soft		strong, aggressive
	weak					gentle		strong, aggressive
	short					fast?		long
	long					slow?		short
	slow					long?		fast
	fast							slow
	soft					gentle		strong,aggressive
	strong					aggressive		weak
	pitch							

Part of the sound	verbal descriptors	physical nature				verbal nature		
		noisy	tonal	fluctuation	source	RELATED	SYNONYMS	ANTONYM
General Impression	aggressive					strong		gentle, weak
	bright					*light, clear*		*dull*
	clear					bright, light, clean		dark (fuzzy, dirty)
	cold							*warm*
	complex					dense, diffuse		simple, easy
	convincing					*pleasant (seduction)*		
	dark					dense	deep, (heavy)	clear, light
	deep					*strong*		*light, thin?*
	defined						focused, distinct	undefined
	dense					*diffuse, dark*	*heavy*	*diffuse, light?*
	diffuse					fuzzy		dense, defined, focused
	distinct					*sharpness, stability*		*diffuse, fuzzy*
	dull						unlively, matt, boring?	bright, convincing, pleasant, exciting, lively
	easy					*simple*		*complex*
	focused						defined, distinct	unfocused
	forced					*pressed, strong, oppressive, tensed*		
	gentle						soft	aggressive
	hollow							
	intense							
	light						*clear, bright*	*heavy, dense?*
	noisy					dirty?		clear,clean
	old							
	pleasant					convincing, nice?		unpleasant
	pressed					*forced, strong, oppressive, tensed*		*relaxed, easy*
	raw					undefined, unfinished?		finished, refined?
	relaxed					*pleasant, calm?*		*tensed*
	sandy							
	simple					*easy*		*complex*
	stable					defined, distinct, focused		unstable
	strong					*tensed, pressed, stable*		*weak, unstable*
	sweet							
	tensed					*pressed, strong*		*relaxed, easy, free?*
	thin					light, shallow		dense, deep, heavy
	undefined					*unfocused, fuzzy*		*defined, focused*
	unfocused					undefined, unfinished?		focused, defined
	unpleasant					*aggressive?*		*pleasant, convincing*
	unstable					unfocused, fuzzy, undefined		stable, relyable?
	warm					*round, deep*		*cold, sharp*
	weak					light, shallow, thin, unstable?		strong

Part of the sound	verbal descriptors	physical nature				verbal nature		
		noisy	tonal	fluctuation	source	RELATED	SYNONYMS	ANTONYM
Steady State	airy					openness, nature, wind, spacious?		clear, clean, closed
	breathy					*forced, pressed*		*clear, distinct, healthy*
	bright						sharp, light	
	clean						*clear, bright*	*dirty,rough*
	clear							
	cold					*light, sharp, bright, clear?*		*warm*
	dirty							clean
	dull						*unlively, matt, boring?*	*bright, convincing pleasant exciting*
	floppy			random				
	floating							
	flowy			laminar				
	fluffy			*gentle*			*wooly*	*hard, flat, regular,*
	fluty					strong fundamental		nasal? Bright?
	free					*open, spacious*		*opressive, closed, tight*
	full						round, thick?, generous?	empty,hollow,(tight, thin, sharp)
	harsh						*rough, coarse*	*smooth, soft, sweet*
	heavy							
	horn-like			*periodic*				
	leaky			poor				
	light							
	loose					open, free, disconnected?		firm, tight
	nasal					*closed*	*reedy*	
	oppressive					pressed, tight		free, open, spacious, floating,
	reedy						*nasal*	
	rough			thick, violent		harsh, raspy		smooth
	round					*full,*		*squared, sharp,hollow,*
	sandy			fine, small		grain		
	sharp					*reedy, edgy*		*dull, blunt, sweet? (positive), flat*
	singing						lively	quiet, dull, closed
	spatial							
	spitting						leaky	
	stringy					*nasal, bright, light*		*fluty*
	thin					hollow, weak		thick, full
	throaty					*opressed, noisy, rough*	*dirty, fuzzy*	*clear, clean*
	tight					closed, pressed		loose, open free
	warm					*balance, round, low pitch*		*cold*
	windy							
	wooly			*slow*		*soft*		*flat, rough, sharp, sticky*

Appendix V: LISE - Listening Interface for Sound Experiments

style conventions

italics	**scripts (ASCII .txt files)**
bold	modules names (.m files)
<u>underline</u>	modules parameters
%	introduce a comment line

Introduction
LISE is a package of routines aimed at listening, comparing and describing sounds. LISE runs under Matlab and has been tested under versions 5.3 for PC and UNIX-LINUX platforms and 5.2 for Macintosh.
At this time, LISE supports 5 types of tests.
Loudness equalization
Pair comparison
Multiple comparison (one-dimensional scales)
Grouping and scaling (on a two-dimensional board)
ABX

The concept used in LISE to specify a listening test session composed of several tests is described on figure below.

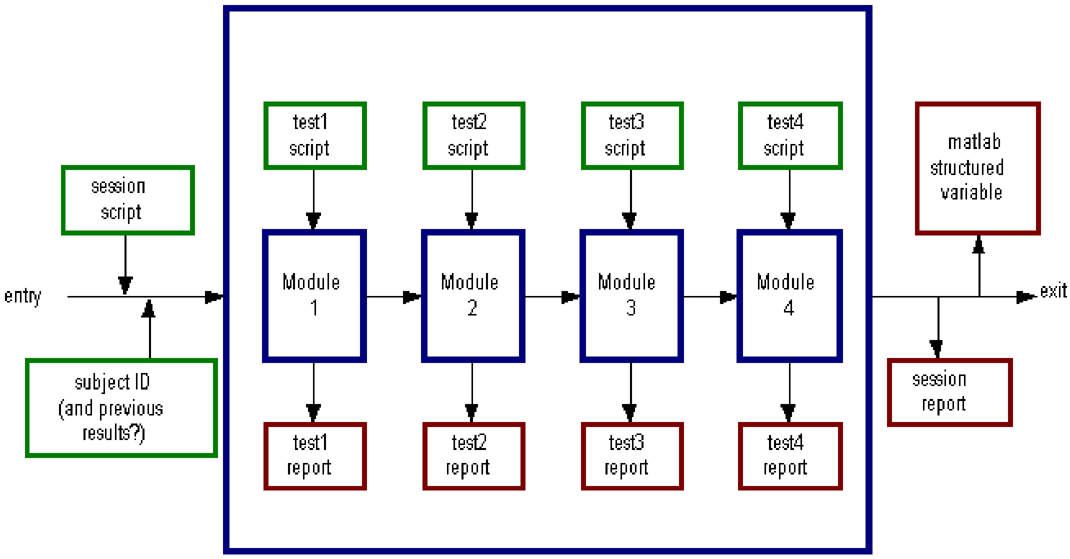


Figure 1. Description of a listening test session (containing here 4 modules).

All scripts are ASCII files of a predefined format. The test-leader uses them to define a session or to prescribe specific parameters of a module.

The session script

This text file specifies the progression of the whole session (the biggest rectangle in Figure 1).
A typical example is shown on Table 1.
After you had a look to Table 1, several preliminary remarks can be made:
The script file obeys the Matlab comments conventions. An empty line or a line preceded by the character . %. will be ignored.
Each non commented line (a command line), shares the format module name *test script*
Commands are sequenced in a top to bottom order in the session script which corresponds to the left to right direction of Figure 1.
The first command line always specifies a call to the module ***module name*** *test script* this module is actually not a test but is used to load a particular sound list (see below).
It is possible to change the sound list during the same session.
All modules will use the last sound list specified.
We will from now on make a distinction between session-scripts and module-scripts and between session-commands and module-parameters respectively.

% A typical session example
% Load the sound list described in *soundList1.txt*
sndList *sndList1.txt*
% Make a loudness equalization of these sounds
loudEq
% Make a scaling - grouping
scalGroup *scalGroup.txt*
%Make a ABX comparison with script: *abx.txt*
abx *abx.txt*
%Make a multiple comparison 1D with script: *comp1D.txt*
comp1D *comp1D.txt*
%Make a multiple comparison 1D with script: *comp2D.txt*
pairComp *comp2D.txt*
% Load the sound list described in *soundList2.txt*
sndList *sndList2.txt*
% and so on...
loudEq
comp2D *comp2D.txt*
comp1D *comp1D.txt*

Table 1. An example of a complex session constituted of several tests and of 2 different sound lists. Possible name: *sessionScript.txt*

Let's now describe each module.
The sound list module : *sndList*
This module basically loads all sounds specified in the associated script.
An example of such a script is given in Table 2.

```
DIRECTORY: ../sounds/
% the sounds to be put in the test
sound1
sound2
%sound99
sound17.wav
```

Table 2. An example of a sound file list. Possible name: *sndList1.txt*

The format of a sound file script is very basic. You may want to specify a directory (see below on how to handle directories) after the module-script command DIRECTORY:
IMPORTANT! All module-script parameters have a strict syntax. They usually appear in CAPITAL LETTERS.A single change from e.g. DIRECTORY: to DIRECTORY will probably make the session crash. The best way to avoid problems is to use the given examples as templates... The sounds to be loaded are then specified one after the other.
In this example, sound99 will here be ignored. An index is assigned to each sounds. sound17.wav will get index #3.

The *ABX* module

An ABX test is typically used when it is useful to know if listeners perceive a noticeable difference between e.g., a set of processed sounds and a set of the same sounds unprocessed. Sounds are presented by pairs (A, B). A sound X is also presented and is selected randomly from the pair (A, B). The subject must determine wether X is A or B.
A typical layout for this test is presented in Figure 2. Its typical module-script is displayed in Table 3.

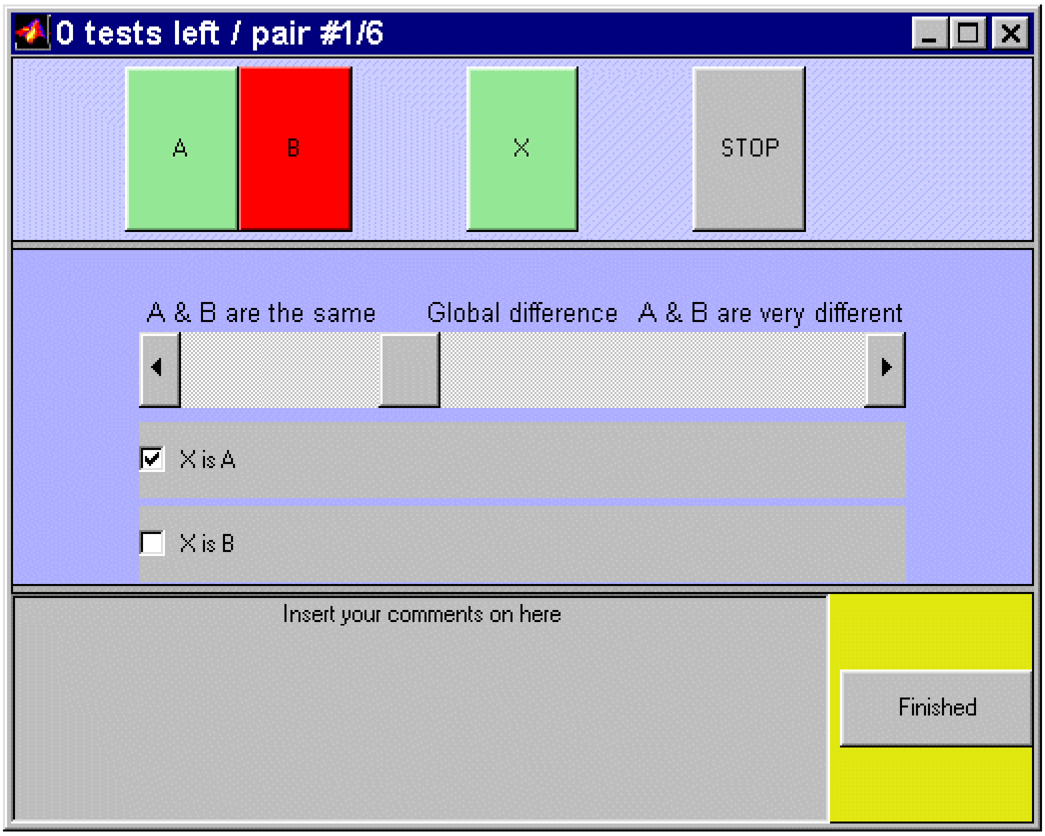


Figure 2. Example layout for an ABX test.

%abx test specifications
NAME
abx test #1
NBSOUNDS: 3
RANDOM?: 1
REPETITIONS?: 1
NBSCALES: 1
SCALES
SCALE:
Global difference
LEVEL
min 1 A & B are the same
max 7 A & B are very different
step 1
NBQUESTIONS: 2
QUESTIONS
2 X is A
1 X is B
COMMENTS?: 1

Table 3. A typical module-script of an ABX test. Possible name: *abxScript.txt*

The pair comparison module : *pairComp*

This module offers the possibility to compares sounds by pairs. A typical layout for such a test is presented on Figure 3.

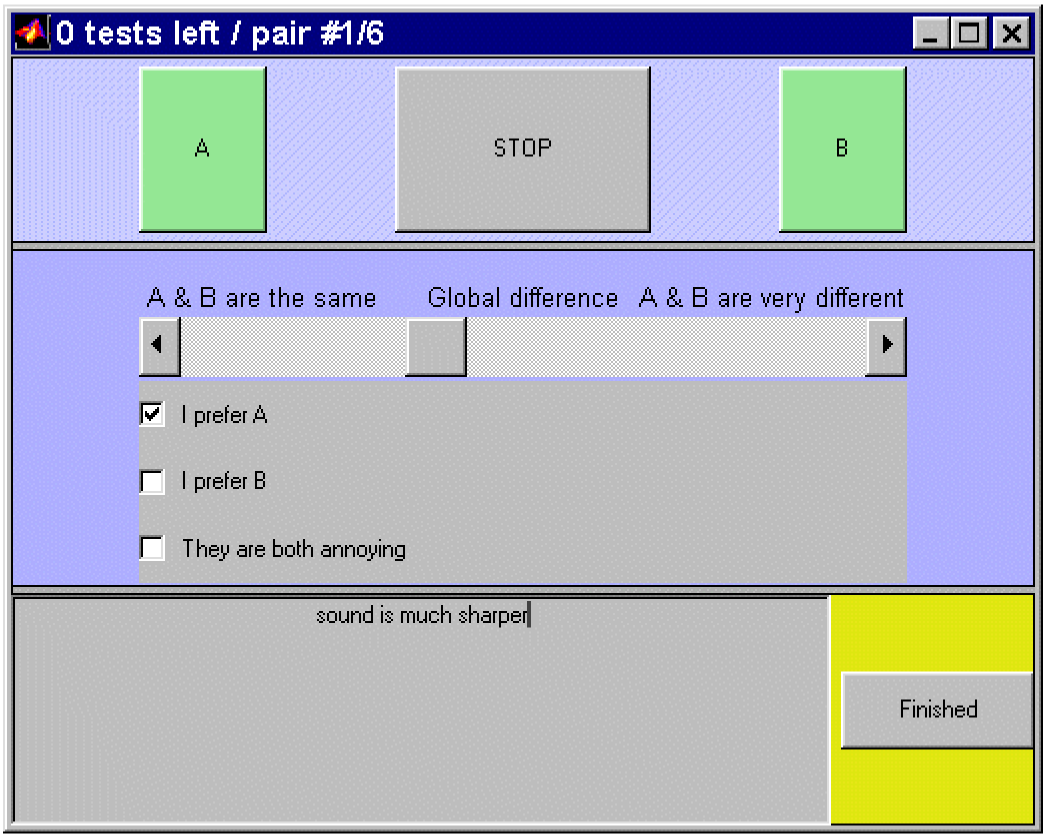


Figure 3. Example layout for a pair comparison test.

% pair comparison specifications % (combinations specified) NAME pair comparison test #1 NBCOMBINATIONS: 3 COMBINATIONS 2 3 2 1 3 1 NBSCALES: 1 SCALES SCALE: Global difference LEVEL min 1 A & B are the same max 7 A & B are very different step 1 SCALE: Sharpness LEVEL min 1 A & B are the same max 7 A & B are very different step 1 NBQUESTIONS: 3 QUESTIONS	% pair comparison specifications % (random combinations) NAME pair comparison test #1 NBSOUNDS: 3 RANDOM?: 1 REPETITIONS?: 1 NBSCALES: 1 SCALES SCALE: Global difference LEVEL min 1 A & B are the same max 7 A & B are very different step 1 SCALE: Sharpness LEVEL min 1 A & B are the same max 7 A & B are very different step 1 NBQUESTIONS: 3 QUESTIONS 2 I prefer A

2 I prefer A 1 I prefer B 0 They are both annoying COMMENTS?: 1	1 I prefer B 0 They are both annoying COMMENTS?: 1

Table 4. A typical module-script of a pair comparison test. Possible name: *pairCompScript.txt*

The multiple comparison module : *comp1D*
This module offers the possibility to compares sounds by pairs. A typical layout for such a test is presented on Figure 4.

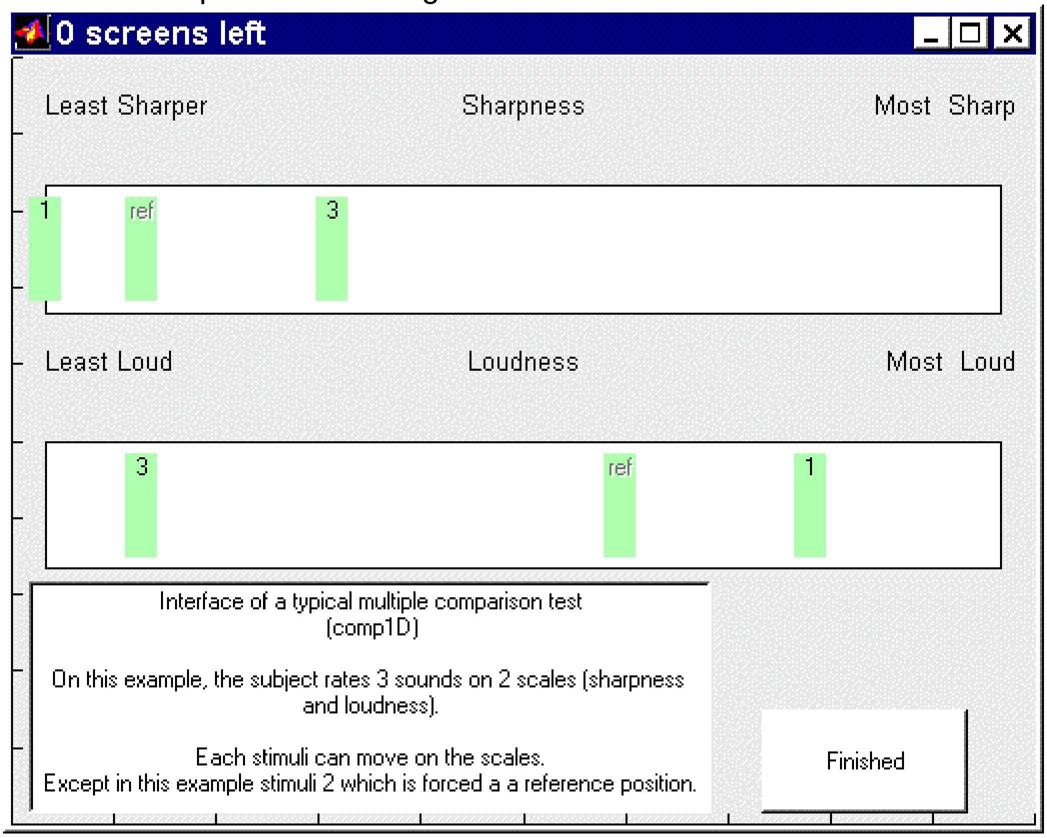


Figure 4. Example layout for a multiple comparison test.

% a multiple comparison test script NAME multiple comparison 1D test #1 NBSCALES: 2 NBSOUNDS: 3 COMBINATIONS 1 position 0 9 2 reference position 1 7 3 position 3 2 SCALES SCALE: Sharpness LEVEL min 0 Least Sharper max 10 Most Sharp

<u>step</u> 1 <u>SCALE:</u> Loudness <u>LEVEL</u> <u>min</u> 1 Least Loud <u>max</u> 11 Most Loud <u>step</u> 1

Table 5. A typical module-script of a multiple comparison test. Possible name: *comp1DScript.txt*

The grouping-scaling module : *groupScal*

This module offers the possibility to group sounds on 2D screen, to scale them (globally or according to a user-specified attribute) and to add comments. A typical layout for such a test is presented on Figure 5a and 5b.

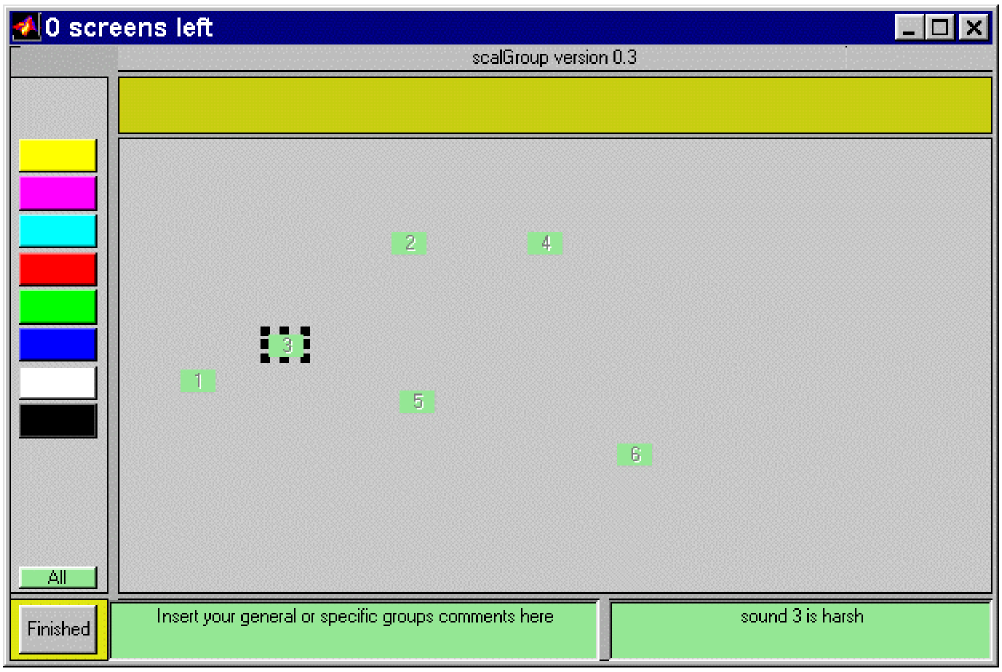


Figure 5a. Example layout for a grouping/scaling test.

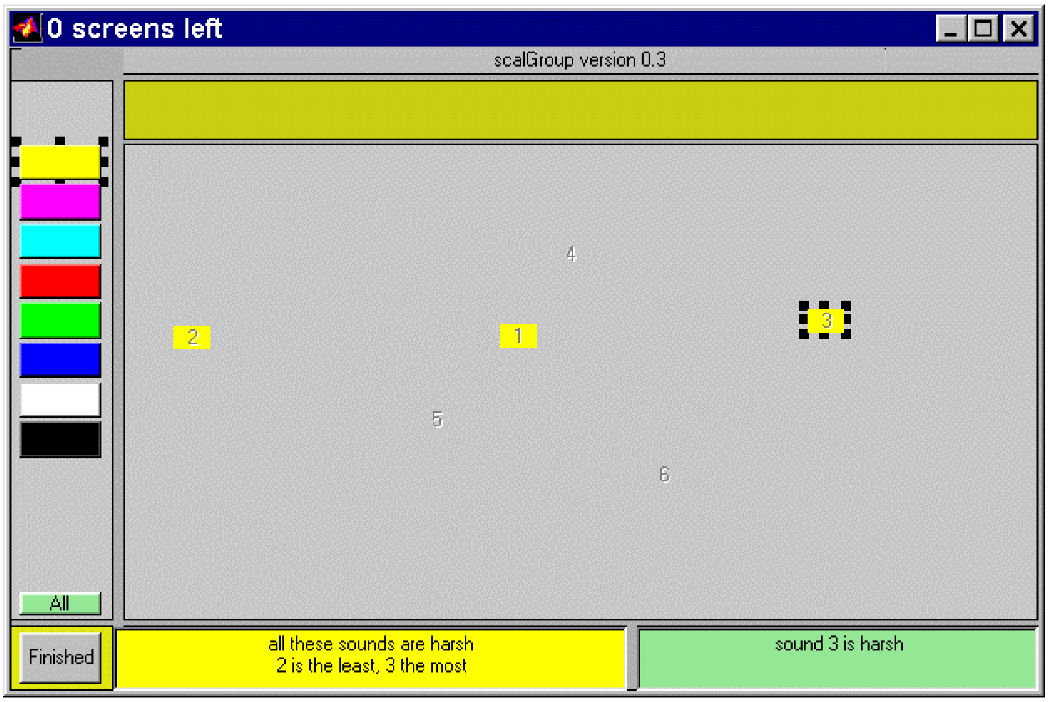


Figure 5b. Example layout for a grouping/scaling test. A group is being formed, in yellow, containing three stimuli {1,2,3}. Those are stimuli are scaled within the window board.

%grouping and scaling specification
NAME
scaling and grouping
GROUPS?: 1
COMMENTS?: 1

Table 6. A typical module-script of a scaling-grouping test. Possible name: *scalGroupScript.txt*

The loudness equalization module : *loudEq*

This module offers the possibility to equalize the loudness of all stimuli. It is particularly interesting to to so when a detailed description of the timbre is needed. A typical layout for such a test is presented on Figure 6.

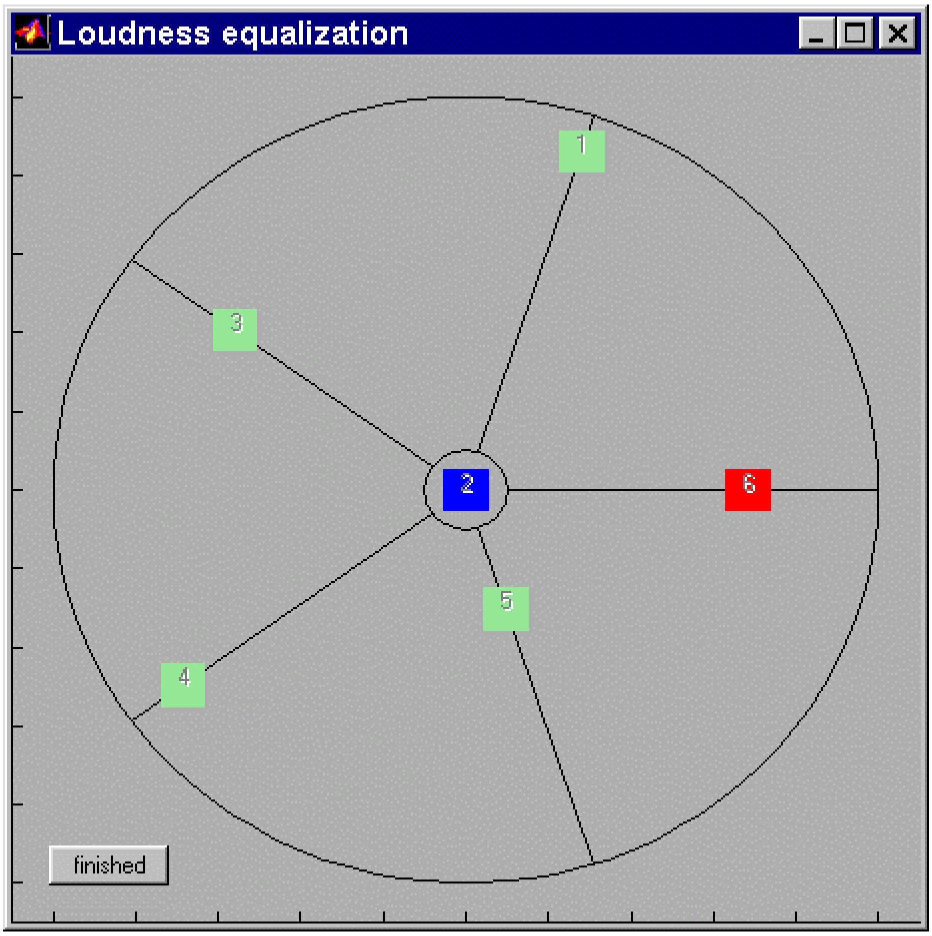


Figure 6. Example layout for a loudness equalization procedure.

Sounds are displayed on a circle. One sound is placed at the middle and serves as a reference.

Commands:
- left mouse button : select the current icon (displayed in red) and play the sound.
- right mouse button : press once to drag and once more to drop.
- key '>' : exchange the current sound with the reference.
- space key : play the reference.

Procedure:
In order to achieve a quick equalization, the following steps are recommended:

step 1:
- compare each sound with the reference
- if a sound is softer than the reference, make it the reference by pressing '>'

at this stage the softest sound should be at the middle and all other sounds should be placed at their original position (on the outer circle).

step 2:
- compare each sound with the reference
- adjust its loudness by varying its distance to the reference

Notes:
If LoudEq is integrated in a test session, all following modules will use the equalized sounds until a new sound list is loaded.
LoudEq can also be used as a stand-alone routine. In that case you can retrieve the scaling factors in a variable called:
loudeqData

Organization of a test session
Sessions' samples are given in the folder "demo".
In order to launch a particular session script make this folder the current directory and type **launch**
You are then asked to choose a particular session script among the following ones :
- sessionABX.txt
- sessionPairComp.txt
- sessionComp1D.txt
- sessionScalGroup.txt
- sessionGlobal.txt

You are then asked to enter your name (without accentuated characters...).
The test will then begin.
Once you are done with the test session, a script file and a matlab variable containing your results are saved in the folder "demo" with the following conventions:
myID_name-of-the-session.txt
myId.mat

The LISE environnement offers the possibility to launch a previously recorded session. If you are in a directory where an output variable like *myId.mat* is placed then typing *myId* in the identification window will automatically launch the previously saved session with the previous results.

Conventions for the structure of the output variable
If you load an output variable like e.g. myId.mat you will get:
» load myId
» myID
myID =
id: 'myID'
sesScript: 'sessionGlobal.txt'
tstDat: {1x10 cell}
The results associated to the module n are stored in variable **myId.tstDat{n}.rsl**

Notes
Test windows are locked so that you can't close them while passing the test.
In order to force one window to close, select it and type: **closereq**

GLOSSARY

This glossary contains some words or expressions chosen and indicated in *italics* in the chapters of this thesis (it is by no means intended to be exhaustive).

additive synthesis: A sound synthesis method based on the summation of basic components like sinusoids in which parameters like amplitude and frequency are controlled and may vary in time. This type of synthesis has a counterpart in subtractive synthesis, in which a simple “block” of sound (like white noise) is filtered at different frequencies. This duality is loosely analogous to sculpture in stone (shaping an already existing element, like subtractive synthesis) in contrast to painting (adding elements to a blank canvas, like additive synthesis). It should be added that additive synthesis, if used in a detailed and complex way, is a powerful tool for simulating organ pipe sounds (Comerford, 1993) but has some inherent problems like the storage of data or the choice of breakpoints on the amplitude of harmonics (Horner *et al.*, 1996). Furthermore, it certainly lacks physicality (Jaffe, 1995), in contrast to physical models such as those proposed by Verge (1997) or Kropp (1996), which provide control parameters that are much closer to real voicing parameters.

allophone: One of two or more variants of the same *phoneme* (the aspirated \p\ of *pin* and the unaspirated \p\ of *spin* are *allophones* of the phoneme \p\).

attribute: This term refers to the continuous dimensions or discrete features of auditory perception. Attributes are considered to be the equivalent of prototypes. They refer to the common dimension of a class of objects which can consequently be compared, grouped and scaled. We took care in this thesis to differentiate between *descriptors* and attributes. Verbal descriptors give access to the degree of existence of an attribute for a specific class of stimuli.

auditory grouping: from (McAdams *et al.*, 1993a). “Auditory grouping processes effect the fusion and segregation of concurrent sound elements into auditory events, as well as the temporal integration and segregation of successive sound events into auditory streams. These processes are brought into play daily in many different ways. For example, as you sit in front of the television you may suddenly hear the honking horn of a car in the street outside the window. Without the slightest hesitation, this newly arrived noise will be heard as separate from and superimposed upon that of the television and not as forming a single sound object with whatever happens to be sounding on the television at the moment. Bregman (1990) discusses principles according to which the array of time-varying activity issued from the tonotopic channels is then believed to be processed in order to constitute separate auditory representations of the various sound sources present in the environment.” In our study, concerned with restricted auditory streams (down to a single sound), auditory grouping needs to be revised. On this scale, only expert ears are used to segregate various sound events. Sources are identified only if the physical source (the flue pipe) is known in detail (this is the case for an organ-builder).

centroid / spectral centroid: Center of gravity of a frequency spectrum. This quantity is the mean of the frequencies of the harmonic components weighted by their respective amplitudes. It is used as a practical value to characterize the whole harmonic spectrum (of a stationary part of a sound) instead of all the frequencies and amplitudes of each component. A time-varying version of this quantity is proposed in Chapter III (see also Beauchamp et al., 1990).

compactness : Differentiates between complex sounds and noise (Pollard 88).

equalization: A sound was considered equalized relative to another sound when their (perceived) *loudness* was set identically (see routines Matlab® – LoudEq).

feature: Used by McAdams (1999a) for "boolean characteristics." In opposition to a dimension that may represent a continuous attribute of a sound, a feature of a sound is activated mainly by its identification. Its degree of presence (or intensity) is then perceptually of less interest than its associated threshold of detection. The term feature is also more generally used by Pollard for "perceptive characteristics" (H.F. Pollard, 1988). In this thesis we follow McAdams' definition as much as possible.

learning effect: In the context of a listening test, the learning effect occurs at the beginning of a test. In considering a particular set of sounds, knowledge of the whole set is required in order to perform some mutual comparisons on a scale. This knowledge improves progressively during the test, meaning that first answer should be considered as less significant. A check for this assumption is that the time spent on the answers will decrease as the test proceeds to its end. In order to reduce this drawback, a "passive" presentation of all the stimuli or an "active" grouping test can be done.

loudness: Perceived intensity of a sound (Zwicker, 1999).

micro-temporal / micro-scale components: Components of a single sound "object" the duration of which is of the order of a second (a "nominal" value for a musical note). Used in contrast to macro-temporal components, which are defined at a compositional level (rhythms, keys, etc.).

nicking: A pipe is nicked when some indentations are made on the edge of the languid. These indentations are designed to reduce the amount of noise. The use of the nicking technique, even though regulating some *micro*-components of the sound (e.g. "chiff" and the amount of noise during the stationary part), separates distinctively two "schools" of voicing.

noise: (1) Unwanted, undesired sound or part of a sound (etymological meaning). In that sense, this word is related to **annoyance** and thus is an **affective judgment**. In this sense, a noise can be any type of signals, *stochastic* (aleatory) or deterministic (periodic). For example, sound from an exhaust pipe (stochastic) and sound from a mobile telephone (deterministic) can both be considered to be noises.

(2) Stochastic part of a signal (see *white noise*). Because in *information theory*, a stochastic signal does not carry information, it is by extension classified as undesirable. For musical sounds, the stochastic part is an important cue for *identification* and *recognition* and cannot be reduced to an undesired element.

(3) Residual part of a signal. See *residual*.

In order to avoid ambiguity, we try to use only the first definition of the word "noise" and refer to "stochastic part" when needed.

onomatopoeia: (1) The naming of a thing or action by a vocal imitation of the sound associated with it (as buzz, hiss).

(2) The use of words whose sound suggests the sense.

partial: Tonal component. Often used when one wants to emphasize its isolated or inharmonic character.

phoneme: Any of the abstract units of the phonetic system of a language that correspond to a set of similar speech sounds (as the velar \k\ of *cool* and the palatal \k\ of *keel*) which are perceived to be a single distinctive sound in the language

pitch : Each pipe has a particular pitch, mainly linked to the **perception** of the fundamental frequency of the stationary part.

rank : See *stop*.

residual: The part of a sound that is left when periodic (harmonic) components are retrieved. The residual thus contains bursts of stochastic and tonal components (during transients) and the stochastic part of a signal. (See Chapter III).

sharpness: von Bismarck (1974a) defined it in a way quite similar to the spectral centroid, but instead used loudness and a psychoacoustics frequency scale. von Aures (1985) revised this first definition.

sonic icon: In a computer environment, an icon that represents a sound. It can be dragged and dropped like usual file icons. When clicked, a sound is heard, for example through headphones connected to a sound card.

sonic object: Following the paradigm issuing from the electro-acoustical experiments of Schaeffer (1966), a sonic object is a sound that is appreciated for itself without prior identification. This describes rather well the sound files that were presented in our listening tests. The durations of these sound files were all on the order of a second.

sound quality: Blauert (1997) gives the following definition of product sound quality: "Sound quality is the adequacy of a sound in the context of a technical goal and/or context." Note that the notion of **context** appears to be essential.

stochastic: Aleatory. Preferred to the term "noisy" because "stochastic" is a signal-based (objective) description, whereas "noisy" is an affective (subjective) description.

stop : Term denoting a rank of pipes sharing the same global shape (or proportion) through several octaves. A particular stop thus has a particular global "sound." Several stops can be assigned to the same keyboard so that a single note pressed will produce a complex sound composed of different pipes whose sounds blend with a particular timbre (for an introduction, see e.g. Rochas, 1997).

tonal: Refers to the subjective evaluation of a periodic signal. A tone or tonal component is closely related to the notion of pitch. There is a point where narrow-band noise and pure tones (sinusoids) meet perceptively.

transient: Perceptive character. Opposed to established, stationary, sustained. A transient sound is a short sound during which the notion of pitch, vibrato, duration is hardly perceived. The transient of a sound often refers to its onset. The starting transient constitutes the onset of a sound. Also referred to as "speech" (organ-builder vocabulary), attack or attack transient. Acoustically, it corresponds to the period of time during which steady oscillations are not yet reached. It involves complex phenomena and constitutes a signature of the instrument both perceptually and acoustically. The onset of most musical instruments is brief, ranging from a few milliseconds for plucked strings to as much as 300 ms for a slowly bowed string (H. F.

Pollard, 1988). The ending transient constitutes the **offset** or **decay** of a sound. For pipe organ sounds, it is often disregarded, as the reverberant field of large halls like churches tends to drown it.

white noise: A signal that statistically contains all the audible frequency. The frequency spectrum of such a stochastic signal is then flat. The concept was borrowed from the theory of light in physics (Feynman et al., 1964), where a white signal contains all visible frequencies (perceived as “colors”). Pink and blue noises are constructed on the same analogy and are similar to white noise after subtraction of some parts of the spectrum.

REFERENCES

Abelin, Å. (1988). “Patterns of Synaesthesia in the Swedish Vocabulary.” *Studies in Com puter-Aided Lexicology*.17-34. Stockholm: Almqvist & Wiksell International.

__________. (1999). *Studies in Sound Symbolism.* Unpublished Ph.D. thesis, Department of Linguistics, Göteborg University, Göteborg, Sweden.

Allvar, G. (2000). Private communication.

Angster, J. (1997). “Effect of Voicing Steps on the Stationary Spectrum and Attack Transient of a Flue Organ Pipe.” Paper presented at the International Symposium on Musical Acoustics (ISMA), Edinburgh.

Aures (von), W. (1985). “Berchnungverfahren für den sensorischen Wohlklang beliebiger Schallsignale.” *Acustica, 59*, 130-141.

Baird, J. C., and Noma, E. (1978). *Fundamentals of Scaling and Psychophysics*. John Wiley & Sons Inc, New York.

Ballas, J. A. (1993). ”Common Factors in the Identification of an Assortment of Brief Everyday Sounds.” *Journal of Experimental Psychology: Human Perception and Performance, 19*(2), 250-267.

Beauchamp, J. (1982). “Synthesis by Spectral Amplitude and 'brightness' Matching of Analysed Musical Instrument Tones.” *Journal of the Audio Engineering Society, 30*(6), 396-406.

__________ and Freed, D. (1990). “Auditory Correlates of Perceived Mallet Hardness for a Set of Recorded Percussive Sound Events.” *Journal of the Acoustical Society of America, 87*, 311-322.

Bech, S. (1999). “Methods for Subjective Evaluation of Spatial Characteristics of Sound.” Paper presented at the Audio Engineering Society 16th International Conference, April 10-12, 1999.

Bédos de Celles (Dom), F. (1766-78). *L'art du facteur d'orgues - 4 vols*.; facsimile edition, Paris: Bärenreiter-Kassel, 1958.

Bekesy (von), G. (1959). “Similarities Between Hearing and Skin Sensations.” *Psychological Review, 66*, 1-22.

Bel, B. (1992). “Symbolic and Sonic Representations of Sound-Object Structures. In *Understanding Music with Ai Perspectives on Music Cognition*, ed. M. Balabau, K. Ebcioglu, and O. Lasse. 64-109. Cambridge: MIT Press.

Beranek, L. (1993). *Acoustics*. 2nd ed. New York: Acoustical Society of America through the American Institute of Physics.

__________ and Takayuki, H. (1998). “Sound Absorption in Concert Halls by Seats, Occupied and Unoccupied, and by Halls’ Interior Surfaces.” *Journal of the Acoustical Society of America, 101*(6), 3169-3177.

Berg, J., and Rumsey, F. (1999). “Spatial Attribute Identification and Scaling by Repertory Grid Technique and Other Methods.” Paper presented at the Audio Engineering Society 16th International Conference, April 10-12, 1999.

__________ and Rumsey, F. (2000). “In Search of the Spatial Dimensions of Reproduced Sound: Verbal Protocol Analysis and Cluster Analysis of Scaled Verbal Descriptors.” Paper presented at the Audio Engineering Society 108th , Paris.

Bernston, A. (1987). *Preferred Acoustical Conditions for Talkers: Judgements of Simulated Multiple Early Reflection Patterns*. Göteborg: Chalmers Tekniska Högskola.

Bismarck (von), G. (1974). ”Sharpness as an Attribute of the Timbre of Steady Sounds.” *Acustica, 30*, 159-172.

__________. (1974). “Timbre of Steady Sounds: A Factorial Investigation of Its Verbal Attributes.” *Acustica, 30*, 146-158.

Björk, E. A. (1985). “The Perceived Quality of Natural Sounds” (research notes). *Acustica, 57*, 185-188.

Blauert, J., and Jekosch, U. (1997). “Sound Quality Evaluation—A Multi-layered Problem.” *Acustica, 83*, 747-753.

Bodden, M. (1997). “Instrumentation for Sound Quality Evaluation.” *Acustica, 83*, 775-783.

__________. (1999). “Psychoacoustics and Industry: Instrumentation versus Experience?” *Acustica, 85*, 604-607.

__________, Heinrichs, R., and Linow, A. (1998). Sound Quality Evaluation of Interior Vehicle Noise Using an Efficient Psychoacoustic Method.” Paper presented at Euronoise 98, Munich.

Bower, G. H., and Holyoak, K. (1973). “Encoding and Recognition Memory for Naturalistic Sounds.” *Journal of Experimental Psychology, 101*(2), 320-366.

Bregman, A. (1990). *Auditory Scene Analysis*. Cambridge: MIT Press.

Caddy, S., and Pollard, H. F. (1957). “Transient Sounds in Organ Pipes.” *Acustica, 7*, 277-280.

Cadoz, C., Luciani, A., and Florens, J. L. (1981). “Synthèse musicale par simulation des mécanismes instrumentaux. Transducteurs Gestuels Rétroactifs pour l'étude du jeu instrumental.” *Revue d'Acoustique, 59*, 279-292.

Carlsson, T. (1999). *On Wind System Dynamics of Pipe Organs*. Thesis for the degree of Licentiate in Engineering, Chalmers University of Technology, Göteborg, Sweden.

Carterette, E. C., and Kendall, R. A. (1982). “Comparative Music Perception and Cognition.” In *The Psychology of Music.* 2nd ed. ed. D. Deutsch. 766-768. New York: Academic Press.

Castellengo, M. (1969). *Paramètres sensibles d'un tuyau d'orgue à embouchure de flute* (42). Paris: Groupe d'Acoustique Musicale. Faculté des sciences.

__________. (1976). *Contribution à l'étude expérimentale des tuyaux à bouche*, Thèse doctorale de l'université Pierre et Marie Curie, Paris.

__________. (1994). "La perception auditive des sons musicaux." In *Psychologie de la musique*, ed. Zenatti A. 55-86. Paris: Presses Universitaires de France.

__________. (1999). "Acoustical Analysis of Initial Transients in Flute-like Instruments." *Acta acustica, 85* (3), 387-400.

Cavaillé-Coll, A. (1979). *Études Expérimentales sur les Tuyaux d'Orgues.* In *Complete Theoretical Works of A. Cavaillé-Coll*. Buren (GLD.), The Netherlands: Frits Knuf.

Civille, G. V. (1998). "Understanding the Flavors of Foods and Beverages in Terms of Subjective and Objective Sensory Evaluation Responses." Paper presented at the Sound Quality Symposium in Ypsilanti, Michigan, April 9, 1998.

Coltman, J. (1976). "Jet Drive Mechanisms in Edge Tones and Organ Pipes." *Journal of the Acoustical Society of America, 60*(3), 725-733.

Comerford, P. (1993). "Simulating an Organ with Additive Synthesis." *Computer Music Journal, 17*(2), 55-65.

Coxon, A. P. M. (1982). *The User's Guide to MULDIMENSIONAL SCALING with Special Reference to the MDS(X) Library Computer Program*. : Heinmann Educational Books.

d'Alessandro, C., Yegnanarayana, B., and Darsinos, V. (1995). "Decomposition of Speech Signals into Deterministic and Stochastic Components." *IEEE, 0-7803-2431-5/95*, 760-763.

David, S. (1997). "Représentations sensorielles et marques de la personne: contrastes entre olfaction et audition." In *Catégorisation et Cognition: de la perception au discours*, ed. D. Dubois. Paris: Kimé.

Davidsson, H. (2000). *Tracing the organ master's secret – The Vision The Process The Goal – The Reconstruction of the North German Baroque Organ for Örgyte Nya Kyrka*. GOArt Publications, no. 2. Göteborg: Göteborg University.

Deddene, L., Ormeire, M. V., Guillaume, P., and Valgaeren, R. (1999). "Objectivization of The Disturbing Sound Elements in Automotive Exhaust Noise." Paper presented at the 6th International Congress on Sound and Vibration, Copenhagen, Denmark.

Denzin, N. K., and Lincoln, Y. S. (1994). "The Landscape of Qualitative Research Theories and Issues." In *The Handbook of Qualitative Research Theories and Issues*. 1-17. Thousand Oaks/London/New Delhi: SAGE Publications.

Ding, Y., and Qian, X. (1998). “Processing of Musical Tones Using a Combined Quadratic Polynomial-Phase Sinusoid and Residual (QUASAR) Signal Model.” *Journal of the Audio Engineering Society, 45*(7/8), 571-584.

Dolédec, S., and Chessel, D. (1994). *Introduction to Linear Data Analysis*. Available: http://biomserv.univ-lyon1.fr/DocThemF.html.

Dossing, O. (1998). “Uncertainty in Time/Frequency Domain Representations.” *Sound and Vibration*, 14-24.

Dubois, D., and Grinevald, C. (1999). “Pratiques de la couleur et denominations.” In *Expériences individuelles et savoirs collectifs – Questions de méthodes*. Cahiers du LCPE, no. 3 (mai 1999). Paris: LCPE (languages, Cognitions, Pratiques).

Ducrot, O., and Schaeffer, J.-M. (1995). *Nouveau dictionnaire encyclopédique des sciences du language*. new edition. Paris: Seuil.

Eaton, W. C. C., G.V. (1998). “Sensory Evaluation Techniques for Sound Quality.” *Proceedings of the Sound Quality Symposium, Ypsilanti, Michigan, April 9,1998.*

Elder, S., A. (1973). ”On the Mechanism of Sound Production in Organ Pipes.” *Journal of the Acoustical Society of America, 54*(6), 1554-1564.

__________. (1978). ”Edgetones versus Pipetones.” *Journal of the Acoustical Society of America, 64*(6), 1721-1723.

__________. (1992). ”The Mechanism of Sound Production in Organ Pipes and Cavity Resonators.” *Journal of the Acoustical Society of Japan, 13*(1), 11-23.

Fabre, B. (1992). *La production du son dans les instruments de musique à embouchure de flûte: modèle aéro-acoustique pour la simulation temporelle.* Unpublished monograph, Pierre et Marie Curie, Paris.

__________, Hirshberg, A. and Wijnands, A.P.J. (1996). “Vortex Shedding in Steady Oscillation of a Flue Organ Pipe.” *Acustica, 82*, 863-877.

__________. (2000). “Physical Modeling of Flue Instruments: A Review of Lumped Models.” *Acustica, 86*(4), 599-610.

Fabrigar, L., Wegener, D., MacCallum, R., and Strahan, E. (1999). “Evaluating the Use of Exploratory Factor Analysis in Psychological Research.” *Psychological Methods, 4:3*, 272-299.

Faure, A., and McAdams, S. (2000). “Verbal Descriptions of Musical Sounds and Their Recognition from Verbal Portraits.” (Submitted to *Science*.)

Feynman, R., Leighton, R. B., and Sands, M. L. (1964). *Lectures on Physics*. Addison-Wesley Publishing Company.

Fitz, K., and Haken, L. (1996). “Sinusoidal Modeling and Manipulation Using Lemur.” *Computer Music Journal, 20:4*, 44-59.

Fletcher, H. (1934). “Loudness, Pitch and the Timbre of Musical Tones and Their Relation to the Intensity, the Frequency and the Overtone Structure.” *Journal of the Acoustical Society of America, 6*(2), 59-69.

Fletcher, N. H. (1976). “Sound Production by Organ Flue Pipes.” *Journal of the Acoustical Society of America, 60*(4), 926-936.

__________ and Rossing, T. D. (1991). *The Physics of Musical Instruments*. New York: Springer-Verlag.

Gabrielsson, A., and Lindström, B. (1985). “Perceived Sound Quality of High-Fidelity Loudspeakers.” *Journal of the Audio Engineering Society, 33*, 33-53.

Gaillard, P. (2000). *Étude de la perception des transitoires d'attaque des sons de steeldrums : particularités acoustiques, transformation par synthèse et catégorisation.*, Université de Toulouse II - Le Mirail, Toulouse, France.

Gaver, W. W. (1993). “How Do We Hear in the World? Explorations of Ecological Acoustics.” *Ecological Psychology, 5*(4), 285-313.

__________. (1993).” What in the World Do We Hear? An Ecological Approach to Auditory Source Perception.” *Ecological Psychology, 5*(1), 1-29.

Gibson, J. (1986). *The Ecological Approach to Visual Perception*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Goad, P., and Castellengo, M. (1996). “Comparison of Reed Stops of Four French Classical Organs.” Paper presented at the International Conference for Music Perception and Cognition, Montréal.

GOArt. (1998). *Göteborg Organ Art Center (web site)*. Available: www.hum.gu.se/goart/.

Goebel, J. (1975). *Theorie und Praxis des Orgelpfeiffenklanges*. Frankfurt am Main.

Grey, J. (1977). “Multidimensional Perceptual Scaling of Musical Timbres.” *Journal of the Acoustical Society of America, 61*(5), 1270-1277.

Guénoche, A. G., H. (1999). “Quelle confiance accorder à une représentation arborée?” Paper presented at the Journées de la Société Francophone de Classification, September 14-17, 1999, Nancy, France.

Guenther, F., Husain, F., Cohen, M., and Shinn-Cunningham, B. (1999). “Effects of Categorization and Discrimination Training on Auditory Perceptual Space.” *Journal of the Acoustical Society of America, 106*(5), 2900-2912.

Guettler, K. (1999). “Musical Acoustics, the Music Student and the Music Teacher.” Paper presented at the Forum Acusticum, Berlin, Germany.

Guilford, J. (1954). *Psychometric Methods*. New York: McGraw, Hill. 154-176.

Guski, R. (1997). “Psychological Methods for Evaluating Sound Quality and Assessing Acoustic Information.” *Acustica, 83*, 765-774.

Guyot, F. (1996). *Étude de la perception sonore en termes de reconnaissance et d'appréciation qualitative: une approche par la catégorisation.* Université du Maine, Le Mans.

Hammarberg, B., Fritzell, B., Gauffin, J., Sundberg, J., and Wedin, L. (1980). “Perceptual and Acoustic Correlates of Abnormal Voice Qualities.” *Acta Otolaryngol, 90*, 441-451.

Handel, S. (1991). *Listening: An Introduction to the Perception of Auditory Events.* 2nd edition. Cambridge: MIT Press.

Hartmann, W. M. (1997). *Signals, Sound, and Sensation*. Woodbury, NY: American Institute of Physics.

Helmholtz, H. (1885). *On The Sensations of Tone*. New York: Dover.

Hempel, T., and Chouard, N. (1999). “Evaluation of Interior Car Sound with a New Specific Semantic Differential Design.” Paper presented at the Forum Acusticum, ASA-EAA-DAGA Joint Meeting, Berlin, Germany.

Henkel, L., Francklin, N., and Johnson, M. (2000). “Cross-Modal Source Monitoring Confusions Between Perceived and Imagined Events.” *Journal of Experimental Psychology: Learning, Memory and Cognition, 26*(2), 321-335.

Hinton, L., Nichols, J., and Ohala, J. (1994). “Sound-Symbolic Processes.” In *Sound Symbolism*. New York: Press Syndicate of the University of Cambridge.

Horner, A., and Beauchamp, J. (1996). “Piecewise-Linear Approximation of Additive Synthesis: A Comparison of Various Methods.” *Computer Music Journal, 20*(2), 72-95.

Howard, J., and Silverman, E. (1976). “A Multidimensional Scaling Analysis of 16 Complex Sounds.” *Perception and Psychophysics, 19*(2), 193-200.

Ising, H. (1971). “Erforschung und Planung des Orgelklanges.” *Walcker Hausmitteilung, 42*, 38-57.

Jaffe, D. A. (1995). “Ten Criteria for Evaluating Synthesis Techniques.” *Computer Music Journal, 19*(1), 76-87.

Jekosch, U. (1999). “Meaning in the Context of Sound Quality Assessment.” *Acustica, 85*, 681-684.

Johansson, T., and Kleiner, M. (submit. 2000). “Theory for Coupled Simple Acoustic Resonators, with Application to Flue Organ Pipes.” (Submitted to *Journal of the Acoustical Society of America*.)

Keeler, J. S. (1972). “The Attack Transients of Organ Pipes.” *IEEE Trans. AU, 20*, 378.

Keple, L. A. (1997). *The University of Alberta's Cognitive Science Dictionary*. Available: http://web.psych.ualberta.ca/%7emike/Pearl_Street/Dictionary/dictionary.html.

Kob, M. (1998). “On the Influence of Wall Vibrations on the Transient Sounds of Flue Organ Pipes.” Paper presented at the Nordic Acoustical Meeting Proceedings, Stockholm.

__________. (2000). “Influence of Wall Vibrations on the Transient Sound of a Flue Organ Pipe.” *Acta acustica, 86*(4), 642-648.

Kropp, W., Svensson, P., and Kleiner, M. (1996). “Time Domain Formulation of the Sound Generation Mechanism of Flue Organ Pipes, with Auralization.” Paper presented at the Nordic Acoustical Meeting, Helsinki, Finland.

Lachman, R., Lachman, J. L., and Butterfield, E. C. (1979). *Cognitive Psychology and Information Processing*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Larsson, P., Västfjäll, D., and Kleiner, M. (2000). ”Auditory-Visual Interaction in Virtual Reality: Auditory Cues Improve Memory and Presence in Virtual Environments.” (Submitted for publication.)

Leipp, É. (1989). *Acoustique et Musique*. Paris: Masson.

Liljencrants, J. (1999). *Ising's formula*. Available: www.foxtail.com.

Loughlin, P., J. (1998). “Moments and Maximum Entropy Densities in Time-Frequency.” Paper presented at the SPIE conference on Advanced Signal Processing Algorithms, Architectures and Implementations VIII, San Diego, California.

Lyon, R. F., and Mead, C. (1988). “An Analog Electronic Cochlea.” *IEEE Transactions on Acoustics, Speech, and Signal Processing, 36*(7), 1119-1134.

Maffiolo, V. (1999). *De la caractérisation sémantique et acoustique de la qualité sonore de l'environnement urbain.* Ph.D. thesis, l'Université du Mans.

__________, Castellengo, M., and Dubois, D. (1999). “Is Pleasantness for Soundscapes Dimensional or Categorical?” Paper presented at the Forum Acusticum, Berlin, Germany.

McAdams, S. (1993). “Recognition of Auditory Sound Sources and Events,” *Thinking in Sound: The Cognitive Psychology of Human Audition*. Oxford: Oxford University Press.

__________. (1996).” Audition: Cognitive Psychology of Music.” In *The Mind-Brain Continuum*, ed. R. Llinas and P. Churchland. 251-279. Cambridge: MIT Press.

__________, Susini, P., Misdariis, N. and Winsberg, S. (1998). “Multidimensional Characterisation of Perceptual and Preference Judgements of Vehicle and Environmental Noises. *Proceedings of Euro Noise 98, Munich*, 561-566.

__________. (1999). “Perspectives on the Contribution of Timbre to Musical Structure.” *Computer Music Journal, 23:3*, 85-102.

__________, and Bigand, E. (1993). *Thinking in Sound: The Cognitive Psychology of Human Audition*. Oxford: Oxford University Press.

__________, and Bigand, E. (1993). “Introduction to Auditory Cognition,” *Thinking in Sound: The Cognitive Psychology of Human Audition*. Oxford: Oxford University Press.

__________, and Winsberg, S. (1999). “Psychophysical quantification of individual differences in timbre perception.” Paper presented at the 8th Oldenburg Symposium on Psychological Acoustics, Oldenburg, Germany.

McAulay, R. J., and Quatieri, T. F. (1986). “Speech Analysis/Synthesis Based on a SInusoidal Representation.” *IEEE Transactions on Acoustics, Speech, and Signal Processing, 34*(4), 744-754.

McGurk, H., and MacDonald, J. J. (1976). “Hearing Lips and Seeing Voices.” *Nature, 264*, 746-748.

McNeil, M. (1983). “A Theory of Voicing and Scaling Flue Pipes.” *ISO Journal (International Society of Organ-builders), 23*, 3-22.

Mellody, M., and Wakefield, G. (1988). “Modal Distribution Analysis of Vibrato in Musical Signals.” Paper presented at the SPIE Conference, San Diego, California.

Mercer, D. M. A. (1951). “The Voicing of Organ Flue Pipes.” *Journal of the Acoustical Society of America, 23*(1), 45-54.

__________. (1954). “The Effect of Voicing Adjustments on the Tone Quality of Organ Flue Pipes.” *Acustica, 4*, 237-239.

Miklós, A., and Angster, J. (1998) “Sound Radiation of Open Labial Organ Pipes: The Effect of the Size of the Openings on the Formant Structure.” Paper presented at the International Symposium on Musical Acoustics (ISMA), Leavenworth, Washington, USA.

__________, and Angster, J. (2000). “Properties of the Sound of Flue Organ Pipes.” *Acustica, 86*(4), 611-622.

Miller, D. G. (2000). *Registers in Singing—Empirical and Systematic Studies in the Theory of the Singing Voice.* Rijkuniversiteit, Groningen.

Monette, L. G. (1992). *The Art of Organ Voicing*. Kalamazoo, MI: Western Michigan University.

Moore, B. C. (1997). *An Introduction to the Psychology of Hearing.* 4th edition. San Diego, CA: Academic Press.

Muckel, P., Ensel, L., Chouard, N., and Schulte-Fortkamp, B. (1999). “Sound Descriptions as a Matter of Evaluation in Sound Design.” Paper presented at the 8th Oldenburg Symposium on Psychological Acoustics, Oldenburg, Germany.

Namba, S. (1994). “Noise—Quantity and Quality.” Paper presented at Internoise, Yokohama, Japan.

__________, and Kuwano, S. (1999). “Noise—Quantity and Quality.” Paper presented at the 8th Oldenburg Symposium on Psychological Acoustics, Oldenburg, Germany.

Nederveen, C. (1998.). *Acoustical Aspects of Woodwind Instruments*. DeKalb, IL: Northern Illinois University Press.

Noble, A. C., Arnold, R.A., Buechsenstein, J., Leach, E.J., Schmidt, J.O. and Stern, P.M. (1987). “Modification of a Standardized System of Wine Aroma Terminology.” *American Journal of Enol. Vitic., 38*(2), 143-146.

Nolle, A. W. (1979). “Some Voicing Adjustments of Flue Organ Pipes.” *Journal of the Acoustical Society of America, 66*(6), 1612-1626.

___________. (1983). ”Flue Organ Pipes: Adjustments Affecting Steady Waveform.” *Journal of the Acoustical Society of America, 73*(5), 1821-1832.

Osgood, C. E., and Suci, G. J. (1955). *Experimental Psychology, 50*(325).

Pagano, R. (1994). *Understanding Statistics in the Behavioral Sciences.* 4th edition. West Publishing Company, 447-464.

Parizet, E., and Nosulenko, V. (1999). “Multi-dimensional Listening Tests: Selection of Sound Descriptors and Design of the Experiment.” *Noise Control Engineering Journal, 47(6)*, 227-232.

Pelto, P. (1995). “Four Voicing Techniques,” *The Organ Yearbook* (Vol. XXV).

Pickles, J. O. (1988). *An Introduction to the Physiology of Hearing*. 2nd edition. San Diego, CA: Academic Press.

Pielemeier, W., and Wakefield, G. (1996). “A High-Resolution Time-Frequency Representation for Musical Instrument Signals.” *Journal of the Acoustical Society of America, 99*(4), 2382-2397.

__________, Wakefield, G., and Simoni, M. (1996). “Time-Frequency Analysis of Musical Signal.” *Proceedings of the IEEE, 84*(9), 1216-1230.

Pollard, H. F. (1988). “Feature Analysis of Musical Sounds.” *Acustica, 65*, 232-244.

__________. (1988). “Feature Analysis and Musical Timbre.” *Journal of the Catgut Acoustical Society, 1*(Series II), 16-24.

__________, and Jansson, E. V. (1982). “Analysis and Assessment of Musical Starting Transients.” *Acustica, 51*(5), 249-262.

__________, and Jansson, E. V. (1982). “A Tristimulus Method for the Specification of Musical Timbre.” *Acustica, 51*, 162-171.

Powell, A. (1961). “On the Edgetone.” *Journal of the Acoustical Society of America, 33*(4), 395-409.

Pranter, H. U. (1999). ”Estimation of Sound Quality Measures Using FIR Neural Networks.” *Acustica, 85*, 674-677.

Pruzansky, S., Tversky, A., and Caroll, J.D. (1982). “Spatial versus Tree Representations of Prroximity Data.” *Psychometrika, 47*(1), 3-24.

Rabiner, L.R. and Schaffer, R.W.. (1978). *Digital Processing of Speech Signals*. 396-452. Prentice Hall, Englewood Cliffs,.

Ramig, L. O. (1994). ”Introduction to Communication Sciences and Disorders.” In *Vocal Fold Physiology—Voice Quality Sound*, ed. F. D. Minifie. 481-518. San Diego, CA: Singular Publishing Group.

Repp, B. H. (1987). ”The Sound of Two Hands Clapping.” *Journal of the Acoustical Society of America, 81*, 1100-9.

Rioux, V. (2000). “Methods for an Objective and Subjective Description of Starting Transients of Some Flue Organ Pipes—Integrating the View of an Organ-Builder.” *Acta acustica, 86*(4), 634-641.

__________, Kleiner, M., and Yokota, M. (1998). “Harmonics + Noise Extraction: A Tool for the Description of Flue Organ Pipe Sounds.” Paper presented at the Nordic Acoustical Meeting Proceedings, Stockholm.

__________, Västfjäll, D., Yokota, M., and Kleiner, M. (1999). ”Noise Quality of Transient Sounds: Perception of ‘hiss’ in a Flue Organ Pipe.” Paper presented at the Forum Acusticum, Berlin, Germany.

__________, Yokota, M., Västfjäll, D., Scholz, M., and Kleiner, M. (1998). ”Preliminary Study of an Organ Builder's Perception of a Flue Pipe Sound.” Paper presented at the International Symposium on Musical Acoustics, Leavenworth, Washington, USA.

Risset, J.-C., and Mathews, M. (1969). “Analysis of Musical-Instrument Tones.” *Physics Today, 22*(2), 23-30.

__________, and Wessel, D. (1982). “Exploration of Timbre by Analysis and Synthesis.” In *The Psychology of Music*, ed. D. Deutsch. 25-58. New York: Academic Press.

Rochas, P. (1997). *Le Petit Dictionnaire de L'orgue illustré (livré avec 2 CD)*. Arles: harmonia mundi.

Roederer, J. (1995). *The Physics and Psychophysics of Music—An Introduction*. New York: Springer-Verlag.

Rosch, E. (1978). “Principles of Categorization.” In *Cognition and Categorization*, ed. E. Rosch and B. Lloyd. Hillsdale, NJ: Laurence Erlaum Associates.

Samoylenko, E., McAdams, S., and Nosulenko, V. (1996). “Systematic Analysis of Verbalizations Produced in Comparing Musical Timbres.” *International Journal of Psychology, 31*(6), 255-278.

Sarris, V. (1999). “Perception and Judgement in Psychophysics: An Introduction into the Frame-of-Reference Theories.” Paper presented at the 8th Oldenburg Symposium on Psychological Acoustics, Oldenburg, Germany.

Schaeffer, P. (1966). *Traité des objets musicaux. Essai Interdisciplines.* Nouvelle édition. Paris: Seuil.

Scholz, M., Rioux, V., Västfjäll, D., and Kleiner, M. (1999). “Psychat—A Program for Psychoacoustical Tests.” Paper presented at Forum Acusticum, Berlin, Germany.

Scott, S. *PC-MDS Multidimensional Statistics Package*. Salt Lake City, Utah: Institute of Business Management, Brigham Young University.

Ségoufin, C., Fabre, B., Verge, M. P., Hanssen, E. M. S., Wijnands, A. P. J., and Hirschberg, A. (1998). “Recorder Windway Profile: Influence On Sound Production.” *Proceedings of the International Symposium on Musical Acoustics*, 203-208.

__________, Fabre, B., Verge, M. P., Hischberg, A., and Wijnands, A. P. J. (2000). “Experimental Study of the Influence of the Mouth Geometry on Sound Production in a Recorder-like Instrument: Windway Length and Chamfers.” *Acustica, 86*(4), 649-661.

Serra, X. (1997). “Musical Sound Modeling with Sinusoids plus Noise.” In *Musical Sound Processing,* ed. C. Roads, S. Pope, A. Picialli, and G. De Poli. (Also available On-Line at http://www.iua.upf.es/~sms/docs/msm): Swets & Zeitlinger Publisher.

__________, and Smith III, J. (1990). “Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic plus Stochastic Decomposition.” *Computer Music Journal, 14*(4), 12-24.

Sheridan, C. (1971). *Fundamentals of Experimental Psychology*. Kansas City, Missouri: Holt, Rinehart and Winston, Inc.

Solomon, L. N. (1958). “Semantic Approach to the Perception of Complex Sounds.” *Journal of the Acoustical Society of America, 30*(5), 421-425.

Sundberg, J. (1966). *Mensurens betydelse i öppna labialpipor. The Significance of the Scaling in Open Flue Organ Pipe*, Acta universitatis Upsaliensis Studia musicologica Upsaliensia Nova series 3, Uppsala, Sweden.

Susini, P., Misdariis, N., Winsberg, S. and McAdams, S. (1998). “Caractérisation perceptive de bruits.” *Acoustique & Technique, 13*, 11-15.

__________, McAdams, S., Winsberg, S. (1999). “A Multidimensional Technique for Sound Quality Assessment.” *Acustica, 85:5*, 650-656.

Svensson, P., and Friesel, M. (1999). *Influence of alloy composition, mechanical treatment and casting technique on loss factor and Young's modulus of lead-tin alloys* (F99-04). Göteborg: Chalmers University of Technology.

Titze, I. R. (1995). “Definitions and Nomenclature Related to Voice Quality.” In *Vocal Fold Physiology—Voice Quality Sound*, ed. O. Fujimura and M. Hirano. San Diego, CA: Singular Publishing Group.

Toole, F. E. (1985). “Subjective Measurements of Loudspeaker Sound Quality and Listener Performance.” *Journal of the Audio Engineering Society, 33*, 2-32.

Verge, M.-P. (1995). *Aeroacoustics of Confined Jets with Applications to the Physical Modeling of Recorder-Like Instruments*. Techische Universiteit, Eindhoven.

__________, Fabre, B., Hirschberg, A., and Wijnands, A. P. J. (1997). “Sound Production in Recorderlike Instruments I: Dimensionless Amplitude of the Internal Acoustics Field.” *Journal of the Acoustical Society of America, 101*(5), 2914-2924.

__________, Hirschberg, A., and Caussé, R. (1997). “Sound Production in Recorderlike Instruments II: A Simulation Model.” *Journal of the Acoustical Society of America, 101*(5), 2925-2939.

Västfjäll, D. (2000). Personal communication.

__________. (2001). *Affective Reactions to Interior Aircraft Sound.* Unpublished licentiate thesis, Chalmers University of Technology, Göteborg.

__________. (2001). *Methods for Assessment of Perceived Sound Quality*. Göteborg.

__________, Gärling, T., and Kleiner, M. (1999). ”Preference for Mood, Emotional Reaction, and Anticipated Emotional Reaction.” *(Submitted to Acta Psychologica)*.

Wells, J., and House, J. (1995). *The Sounds of the International Phonetic Alphabet*. London: Department of Phonetics and Linguistics, University College of London.

Williams, J. (1976). "Synaesthetic Adjectives: A Possible Law of Semantic Change". *Language* 52.2:461-78.

Williams, P., and Owen, B. (1988). *The Organ—The New Groove Musical Instrument Series*. New York: W. W. Norton & Company.

Zwicker, E., and Fastl, H. (1999). *Psychoacoustics—Facts and Models.* 2nd edition. Berlin, Germany: Springer Verlag.