Bibliographical reference

Modelling Perceptual Dimensions of Saxophone Sounds, Arne Nykänen, Örjan Johansson, Jan Lundberg and Jan Berg, Acta Acustica **vol. 95** (Number 3), 2009, pp. 539-549

DOI

https://doi.org/10.3813/AAA.918179

ACTA ACUSTICA UNITED WITH ACUSTICA
Vol. 95 (2009) 539 – 549

DOI 10.3813/AAA.918179

Modelling Perceptual Dimensions of Saxophone Sounds

Arne Nykänen, Örjan Johansson
Division of Sound and Vibration, Luleå University of Technology, 971 87 Luleå, Sweden. arne.nykanen@ltu.se, orjan.johansson@ltu.se

Jan Lundberg
Division of Industrial Design, Luleå University of Technology, 971 87 Luleå, Sweden. jan.lundberg@ltu.se

Jan Berg
School of Music, Luleå University of Technology, Box 744, 941 28 Piteå, Sweden. jan.berg@ltu.se

Summary
In the past, musical instruments were developed over long periods of time by skilled craftsmen. Today, most instruments are mass-produced. Design of musical instruments as mass-produced products requires using strategies which make it easier to identify customer needs and develop exact specifications. To develop useful specifications it is necessary to convert general descriptions into something which can be commonly understood and also be interpretable in terms of acoustic metrics. In this study, methods for analysis and specification of steady state parts of alto saxophone sounds were developed. Saxophonists' use of verbal descriptions of saxophone sounds was investigated. Sound stimuli were binaurally recorded. Judgements upon perceived qualities were made by saxophonists and non-saxophonists using the method of verbal attribute magnitude estimation. Perceptual dimensions were identified using principal component analysis of listening test data. Three prominent dimensions were found and described using the verbal attributes: 1) warm/soft, 2) back vowel analogues and 3) sharp/rough. The perceptual dimensions were modelled as linear functions of acoustic metrics. The results were validated through listening tests with new subjects and new stimuli. Based on the findings, the method was seen as an approach which can enhance the musical instrument design process.

PACS no. 43.66.Jh, 43.66.Lj, 43.75.Pq

1. Introduction

In the past, musical instruments evolved over long periods of time as craftspeople used more of a trial and error approach to find what was most satisfying. Today, some musical instruments are still handcrafted while others are more or less mass-produced on an industrial basis. Brass instruments and saxophones have been produced by industrial processes for more than a century. This development of musical instruments as industrial products deserves special attention. The modern product development process commonly starts with the identification of customer needs and then the writing of specifications [1]. For a musical instrument, it is of great importance to include properties of the sounds produced in specifications. Blauert and Jekosch [2] stated that "to be able to actually engineer product sounds, product sound engineer need potent tools, i.e. tools which enable them to specify design goals and, consequently, shape the product sounds in accordance with these design goals". Specifications are usually in writing. Even though sounds may be difficult to describe, written specifications are powerful tools for defining basic requirements. A prerequisite for useful written specifications is that they can be commonly understood. Suitable verbal attributes facilitate the subjective evaluation of sound and are a good base for written specifications. When the engineering process starts, descriptions need to be interpreted in terms of acoustic quantities. Guski [3] has proposed that the assessment of acoustic information should be made in two steps: 1) Content analysis of free descriptions of sounds and 2) Uni or multi-dimensional psychophysics. This approach was used in the present study through multidimensional analysis. The objectives were to elicit and select suitable verbal attributes for the description of steady state parts of saxophone sounds, identify prominent perceptual dimensions, label these dimensions and develop models for description of identified perceptual dimensions in terms of acoustic metrics. The procedure used consisted of four steps as shown in Figure 1.

Received 29 January 2007,
accepted 27 January 2009.

2. Background

2.1. Elicitation and selection of verbal attributes

As human communication is often very general, more specific descriptions are needed during the early stages of product sound design where discussions about requirements are often too abstract. By determining which words are commonly used by a particular user group (saxophonists in this paper) to describe product sounds it is possible to learn about which sound aspects are important. The elicitation of verbal attributes is a qualitative process. By applying structured procedures to this process the chances of finding representative attributes increase. Various procedures have been devised. Repertory Grid Technique [4] has been successfully used by Berg and Rumsey [5] for the elicitation of verbal attributes which describe spatial qualities of sound. Stone et al. have described a procedure called Quantitative Descriptive Analysis [6]. The procedure involves the selection of 12-15 panel members from a pool of some 25 experts. Those selected were considered to have the discriminatory ability and expertise most closely related to the product category in question. A descriptive language was then developed under the guidance of a panel leader. Gabrielsson [7] conducted extensive experiments on the perception of audio reproductions. He elicited verbal attributes by giving questionnaires to sound engineers, audiologists and peoples suffering from hearing loss. The questionnaires had about 200 adjectives collected from various sources and the subjects were asked to rate them according to perceived suitability as descriptions of sound reproductions. In addition, subjects could add their own descriptions. In the study reported here, a procedure similar to Gabrielsson's was used. Verbal elicitation procedures tend to generate a large number of attributes. To be able to evaluate attributes in listening tests, the number of attributes need to be limited. In this paper, frequency of use served as basis for selection.

2.2. Contribution of tone components to perception of sound

Tone components have a great influence on the perceived timbre of a sound. However, it is difficult to describe perceived differences arising from differences in relative distances in frequency and level between tonal components. Perception of sound character of sounds with harmonic or near harmonic spectra has been thoroughly investigated within musical acoustics [8] and voice acoustics [9]. The shape of the envelope of harmonic spectra has been shown to be a dominating factor in the identification of differences in sound character between vowels. The envelope of a vowel spectrum typically contains some separated humps known as formants. The differences in timbre between vowels depend to a large extent on the frequencies for the two formants with the lowest frequencies [9]. Some differences in sound character between different musical instruments may also be explained by a formant-like character of the envelope of their harmonic spectra [10, 11, 12].

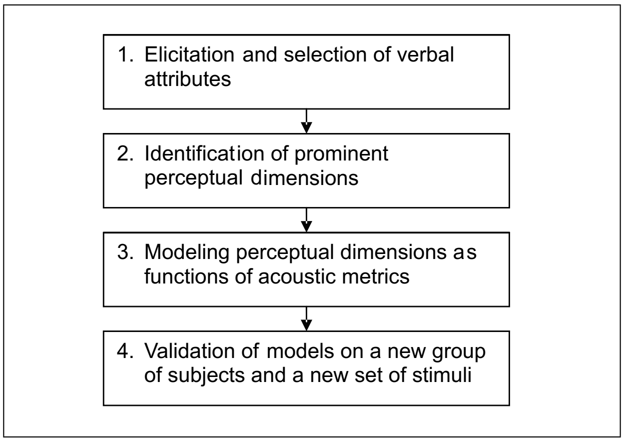

Figure 1. Procedure used for modelling perceptual dimensions of saxophone sounds as functions of acoustic metrics.

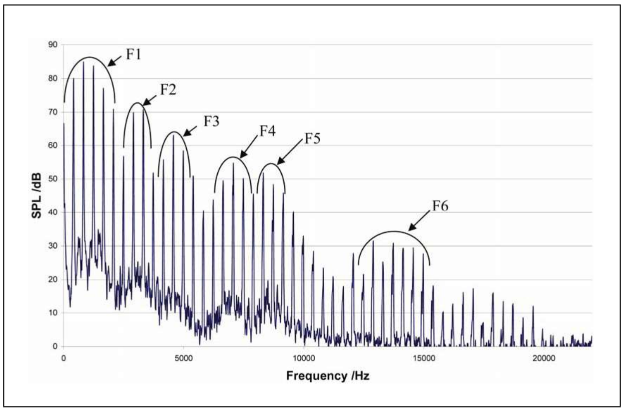

Figure 2. A representative frequency spectrum of a saxophone. F1 to F6 denote the formant-like humps found in the saxophone spectrum.

A typical frequency spectrum of the saxophone shows some similarities to the frequency spectrum of the human voice. Like the human voice, the saxophone spectrum has a strong fundamental and the amplitudes of the partials fall with the frequency [12]. The saxophone spectrum also has strong formant-like characteristics [10, 12] as can be seen in Figure 2. Therefore, saxophone sounds could be expected to resemble vowel sounds. Based on this, vowel analogues were tested as attributes for description of saxophone sounds in addition to the attributes selected from interviews.

2.3. Identification of prominent perceptual dimensions

A perceptual space is a model for how perceptions are structured and organised. Objects, e.g. sounds, will be located in the perceptual space at positions close to similar sounds. Each dimension in the perceptual space corresponds to one perceptual quality. Objects located close to each other in one dimension are similar with respect to the quality associated with that dimension. The most prominent perceptual dimensions are commonly determined by either Multidimensional Scaling (MDS) [13] or multivariate analysis of verbal attribute magnitude estimations [14]. MDS relies on ratings of differences or similarities between stimuli. A primary advantage of this method is that judgements do not depend on previously described scales. Differences in semantic meanings will not influence the results. MDS is a powerful tool for revealing hidden dimensions, i.e. dimensions not described by verbal attributes. The drawback is that it is difficult to interpret all the dimensions. In this paper, multivariate analysis of verbal attribute magnitude estimations was applied. The method was chosen due to the simpler interpretation of most dimensions and a less laborious procedure for the subjects involved in listening tests.

2.4. Modelling perceptual dimensions as functions of acoustic metrics

To prepare a technical specification for a product sound, the perceptual space has to be interpreted in terms of acoustic or psychoacoustic metrics. A psychoacoustic metric is a model of a perceptual dimension which is based on sound pressure variations. In the present study, the following psychoacoustic metrics have been used for development of models for prediction of identified prominent perceptual dimensions of saxophone sounds: 1) Zwicker's model for Loudness [15] (later standardised in ISO 532B), 2) von Bismarck's model for Sharpness [16] and 3) Aures' models for Roughness [17] and Tonalness [18]. However, an experienced perceptual dimension may not necessarily coincide with any model. Combined metrics are required. Such metrics may be developed as regression models predicting perceptual dimensions as functions of a number of acoustic and psychoacoustic metrics. Bodden has discussed the usability and generalisability of combined metrics [19]. He stated that a number of different attributes exist which describe sound quality but that humans usually select only four of them in a sound evaluation process. Aures made an attempt to form a global sound quality metric called "sensory pleasantness" which consisted of a combination of loudness, sharpness and tonalness [18]. Widman developed a combined metric named "psychoacoustic annoyance" based on 5th percentile loudness, sharpness, roughness and fluctuation strength [20]. Annoyance metrics have been developed for some more limited applications. AVL developed an annoyance index for engine noise based on loudness, sharpness and impulsiveness [21]. Kahn et al. developed an annoyance index for diesel engine noise based on loudness, sharpness and harmonic ratio [22]. Bodden wrote, "In contrast to those global indices specific sound quality indices for specific applications seem to be more reasonable" [19] – here "indicies" can be taken to mean the same as "metrics". In this paper, combined metrics for prediction of salient perceptual dimensions of saxophone sounds were developed. Such metrics are not widely generaliseable but give information about which acoustic quantities are most influential on the most prominent perceptual qualities of a certain product sound.

3. Elicitation and selection of verbal attributes

The objective of the first part was to elicit and select suitable verbal attributes for the description of the steady state part of saxophone sounds.

3.1. Method

Ten Swedish saxophone players were individually interviewed. Two were professional saxophonists, 4 studied saxophone performance at the university level and 4 were amateur saxophonists with each having more than 30 years experience. Since elicitation of attributes may be influenced by how material is presented to subjects, elicitation was carried out within a variety of contexts. The following procedure was followed:

1. Without listening to any examples, the subjects were asked which adjectives they use to describe the sound of the product (saxophones).
2. Without listening to any example, each subject was asked to pick from a list the adjectives they most preferred when describing saxophone timbre. The list was compiled from words used in literature dealing with the timbre of musical instruments [23, 24, 25, 26, 27, 28].
3. Various sound examples of recordings of saxophone sounds were played to the subjects who were then asked to describe the sounds in their own words.
4. Different semi-synthetic (digitally modified) saxophone sound examples were played to the subjects who were then asked to describe the sounds in their own words.
5. Each subject was asked to describe in their own words the sounds they strive for when playing.
6. The subjects were asked to play different saxophones and describe the sound characteristics of each.

Since examples of attributes and sounds may influence subjects, the order of introduction of examples is important. The elicitation procedure started with free descriptions and gradually more and more information and examples were introduced. Since the elicitation process is assumed to yield a large number of attributes, the most suitable must be selected. It was hypothesised that the most commonly used verbal attributes would be more distinct and universally understood. Therefore, attributes most commonly used in the different contexts were selected for further analysis. In addition, vowel analogues were tested since they were assumed to be commonly understood and easy to explain to subjects.

3.2. Results and discussion

From the interviews a list of verbal attributes used in the examined contexts was compiled. Totally, 138 attributes were identified. These are shown in Table I. Since the interviews were conducted in Swedish, the verbal attributes have been translated into English. It is important to keep in mind that many words may not be translated without a slight alteration in meaning. Therefore, the Swedish words are listed in the table in addition to their English translations.

Table I. The results of the verbal attribute elicitation process.

English	Swedish	English	Swedish	English	Swedish
airy	luftig	floating	flytande	raucous	skränigt
angular	kantig	fluffy	fluffig	relaxed	avspänd
attackful	attackfull	fluttery	flaxig	rich	rik
backbone	ryggrad	focused	fokuserad	rich of overtones	övertonsrik
balanced	balanserad	forced	forcerad	richly coloured	färgrik
beautiful	vacker	full-toned	fyllig	rocky	rockig
bellow	brölig	hard	hård	rough	rå
bleating	bräkig	harmonic	harmonisk	rough	sträv
blurred	suddig	harsh	kärv	round	rund
bottom	botten	hazy	luddig	screaming	skrikig
bright	klar	heavy	tung	sharp	skarp
bright	ljus	hissing	fräsig	sharp/keen	vass
brilliant	briljant	hoarse	hes	shimmering	skimrande
brittle	spröd	hollow	ihålig	shrill	gäll
brutal	brutal	i-like	i-ig	slim	smal
calm	lugn	intensive	intensiv	small	liten
centred	centrerad	juicy	mustig	smooth	slät
chilled	sval	large	stor	soft	mjuk
classical	klassisk	lifeless	livlös	solid	massiv
clean	ren	light	lätt	song like	sånglik
close	nära	little treble	lite diskant	sonorous	klangfull
closed	stängd	lively	livlig	sonorous	klangrik
coarse	grov	living	levande	sounds like a	låter som
cold	kall	luxuriant	frodig	clarinet	klarinett
compact	kompakt	mawkish	mjäkig	sour	sur
confined	instängd	metallic	metallisk	steady	stadig
convincing	övertygande	metallic	plåtig	straight	rak
cool	cool	mild	mild	strong	stark
cracking	skrällig	more tone	mer ton	supportive	bärig
crunchy	krispig	much bottom	mycket botten	thick	tjock
cutting	skärande	much overtones	mycket	thin	tunn
dark	mörk		övertoner	tidy	snygg
dead	död	nasal	nasal	tinny	burkig
deep	djup	no bottom	ingen botten	toneless	klanglös
dense	tät	open	öppen	ugly	ful
dirty	smutsig	peppery	ettrig	uncentred	ocentrerad
distinct	distinkt	piercing	genomträngande	uninteresting	ointressant
distinct attack	distinkt attack	plastic	plastig	unpleasant	obehaglig
dry	torr	pleasant	behaglig	unpolished	opolerad
dull	dov	pointed	spetsig	warm	varm
egal	egal	poorly coloured	färgfattig	weak	vek
emotional	känslofull	powerful	kraftfull	wet	blöt
empty	kärnlös	presence	pondus	wide	bred
enervating	enerverande	prominent	framträdande	whining	gnällig
explosive	krutig	pronounced	tydlig	with energy	med energi
fat	fet	punchy	kärnfull	woody	träig
filthy	skitig	rasping	raspig		

The attributes used by the highest number of subjects in each of the contexts listed in section 3.1 were selected. The following 10 attributes emerged: large, full-toned, rough, warm, soft, nasal, punchy, sharp/keen, bottom and sharp. In addition, the three established psychoacoustic descriptors: sharpness, roughness and tonalness were selected since these attributes are commonly used by acousticians for describing aspects of timbre. One of the subjects participating in the interviews used vowel analogues. Based on theoretical discussions on similarities between saxophone sound and vowel sound, 9 vowel analogues were added to the list of attributes. The list of attributes selected for further analysis is shown in Table II.

4. Identification of prominent perceptual dimensions

The objective of the second part of the experiment was to identify prominent perceptual dimensions and label these dimensions with verbal attributes.

4.1. Method

In a listening test 16 subjects were asked to judge how well binaurally recorded saxophone sounds were described by the previously selected attributes using a verbal attribute magnitude estimation method [23]. Of these subjects, 10 were saxophonists. The other 6 subjects were experienced listeners, either acousticians or professional musicians who play other instruments. All subjects had self-reported normal hearing. The judgements were made on an 11 point scale ranging from "not at all" (= 0) to "extremely" (= 10). In addition, the subjects were asked to judge their overall impression of the sound on an 11 point scale ranging from "extremely bad" (= -5) to "extremely good" (= +5).

4.1.1. Stimuli

Examples of saxophone sounds were binaurally recorded with an artificial head (Head Acoustics HMS III) and played back to subjects using equalised headphones (Head Acoustics HPS IV) [29]. The sound character was altered by recording two different saxophonists, both studying saxophone performance at the university level. Each played the same two alto saxophones of different brands. This gave four possible combinations. The saxophonists were asked to play a jazz standard ("I Remember Clifford" by Benny Golson). Parts of the recordings were used. By having them play an entire piece it was assumed the playing conditions would feel normal. Recording was carried out in a small concert hall designed for jazz and rock music. Each musician stood in the centre of the stage while they performed. The artificial head was placed 5 m in front of the saxophonists and faced them. The saxophonists tuned their instruments before the recording and they were given a tempo by a metronome. Both players were instructed to play a mezzo forte. Since the objective was to study differences between steady state parts of saxophone sounds, short sound excerpts with attack and decay removed were used to ensure that the same section of sound was judged in the listening tests and measured in acoustic measurements. While it may be argued that this leads to unrealistic listening conditions, there is a trade-off between realism and control of the experimental situation. It is likely that attack and decay are perceptually important [30] and thus have an unwanted influence. Even if subjects are instructed to only judge the steady state parts of sound, their judgments would be influenced by differences in attack and decay. This could lead to inaccurate results regarding steady state parts of sounds. To get valid results for complete saxophone tones, further studies have to be made with a focus on perception of attack and decay and possible interactions between attack, decay and steady state parts of tone. Stimuli were created by cutting out 0.5 second long excerpts from the same single notes in the recordings (concert Ab4, f = 415 Hz). The cuts were made so that both the attack and the decay were removed. The sounds were ramped-up over 0.075 s and rampeddown for the same length of time. Sections at nearly the same loudness (68±8 sones(GF), 90.5±1.8 dBA) were selected to avoid differences in loudness which might influence the results. A total of 12 stimuli were created; 3 of each possible combination of saxophone and musician cut from the recordings.

Table II. The result of the verbal attribute elicitation and selection process. *: To give a hint of how the vowels should be pronounced, an example of a word was given in the following way: like a, as for example in "car". In the English translation, the International Phonetic Alphabet (IPA) has been used.

Swedish	English translation
Stor	Large
Fyllig	Full-toned
Rå	Rough
Varm	Warm
Mjuk	Soft
Nasal	Nasal
Kärnfull	Punchy
Vass	Sharp/keen
Botten	Bottom
Skarp	Sharp
Tonal	Tonal
like a, as e.g. in the word "mat"*	[a]-like
like e, as e.g. in the word "smet"*	[e]-like
like i, as e.g. in the word "lin"*	[i]-like
like o, as e.g. in the word "ko"*	[u]-like
like u, as e.g. in the word "lut"*	[ʉ]-like
like y, as e.g. in the word "fyr"*	[y]-like
like å, as e.g. in the word "båt"*	[o]-like
like ä, as e.g. in the word "färg"*	[æ]-like
like ö, as e.g. in the word "kör"*	[œ]-like

4.1.2. Identification of verbal attributes useful for separation of musician and saxophone

The ability of the judged verbal attributes to separate the two saxophones and the two musicians from each other was examined by ANOVA. The judged verbal attributes were used as dependent variables. The judgements of the 12 samples (3 repetitions of each of the 4 combinations of saxophone and musician) were used as observations which resulted in 192 observations (16 subjects × 12 samples). The influence of two factors was analysed: 1) musician playing (2 levels) and 2) saxophone played (2 levels).

4.1.3. Identification of prominent perceptual dimensions

The data was arranged in a matrix whose columns represented the variables (the judged verbal attributes) and rows represented the observations (all subject ratings - 16 subjects × 12 samples). Principal Component Analysis (PCA)

Table III. ANOVA tables for the judged verbal attributes. Only variables significantly affected by the factors at the 95% confidence level are shown. p-values below 0.05 are in bold.

Attribute	Source		Sum of Squares	Df	Mean Square	F-ratio	p
Full-toned	Main effects	A: saxophone	0.188	1	0.188	0.05	0.82
		B: musician	21.3	1	21.3	5.95	**0.02**
	Interactions	AB	18.8	1	18.8	5.23	**0.02**
	Residual		673.5	188	3.58		
	Total (corrected)		713.8	191			
Rough	Main effects	A: saxophone	0.333	1	0.333	0.10	0.75
		B: musician	0.021	1	0.021	0.01	0.94
	Interactions	AB	80.1	1	80.1	25.2	**<0.00005**
	Residual		598.5	188	3.18		
	Total (corrected)		679.0	191			
Warm	Main effects	A: saxophone	1.51	1	1.51	0.39	0.53
		B: musician	7.92	1	7.92	2.04	0.15
	Interactions	AB	55.3	1	55.3	14.26	**0.0002**
	Residual		728.3	188	3.87		
	Total (corrected)		793.0	191			
Soft	Main effects	A: saxophone	1.51	1	1.51	0.39	0.53
		B: musician	7.92	1	7.92	2.04	0.15
	Interactions	AB	55.3	1	55.3	14.26	**0.0002**
	Residual		728.3	188	3.87		
	Total (corrected)		793.0	191			
Nasal	Main effects	A: saxophone	4.38	1	4.38	1.51	0.22
		B: musician	16.9	1	16.9	5.85	**0.017**
	Interactions	AB	4.38	1	4.38	1.51	0.22
	Residual		543.8	188	2.89		
	Total (corrected)		569.5	191			
Sharp/keen	Main effects	A: saxophone	13.0	1	13.0	3.79	0.053
		B: musician	77.5	1	77.5	22.6	**<0.00005**
	Interactions	AB	80.1	1	80.1	23.3	**<0.00005**
	Residual		645.3	188	3.43		
	Total (corrected)		815.9	191			
Sharp	Main effects	A: saxophone	15.2	1	15.2	3.88	**0.050**
		B: musician	154.1	1	154.1	39.4	**<0.00005**
	Interactions	AB	54.2	1	54.2	13.85	**0.0003**
	Residual		735.5	188	3.91		
	Total (corrected)		959.0	191			
[a]-like	Main effects	A: saxophone	6.38	1	6.38	1.02	0.31
		B: musician	0.13	1	0.13	0.02	0.89
	Interactions	AB	27.8	1	27.8	4.42	**0.037**
	Residual		1180	188	6.28		
	Total (corrected)		1214	191			
[o]-like	Main effects	A: saxophone	13.0	1	13.0	2.04	0.15
		B: musician	1.69	1	1.69	0.27	0.61
	Interactions	AB	30.1	1	30.1	4.72	**0.031**
	Residual		1197	188	6.37		
	Total (corrected)		1242	191			
Overall imp.	Main effects	A: saxophone	1.33	1	1.33	0.38	0.54
		B: musician	1.02	1	1.02	0.29	0.59
	Interactions	AB	60.8	1	60.8	17.22	**0.0001**
	Residual		663.2	188	3.53		
	Total (corrected)		726.3	191			

[31] on standardised data was used to find the interrelationships among the variables (judged verbal attributes). Correlating verbal attributes were assumed to represent prominent perceptual dimensions.

Table IV. r^2-values of the four VARIMAX rotated components found by PCA.

Dim	r^2X	Eigenvalue	Salient variables
1	19%	5.2	Positive: full-toned, warm, soft
2	15%	3.3	Positive: [a]-like, [o]-like
3	14%	2.2	Positive: sharp/keen, rough, sharp, nasal Negative: soft, warm
4	10%	1.3	Positive: [i]-like, [y]-like
Sum	58%		

4.2. Results and discussion

4.2.1. Identification of verbal attributes useful for separation of musician and saxophone

The results of the ANOVA of the 192 observations are presented in Table III. Verbal attributes with p-values below 0.05 were deemed suitable for describing common differences between the sounds arising from differences between the saxophone-musician combinations. This resulted in 7 adjectives and 2 vowel analogues: full-toned, rough, warm, soft, nasal, sharp/keen, sharp, [a]-like and [o]-like.

4.2.2. Identification of prominent perceptual dimensions

Scatter plots of the loadings on the first four VARIMAX rotated components of the PCA are found in Figure 3. The number of components was limited to four based on the scree test criterion [31]. The components were assumed to represent prominent perceptual dimensions. The VARIMAX rotation was applied to simplify interpretation of the components in terms of salient original variables (the verbal attributes). The r^2-values and eigenvalues of each component are presented in Table IV together with the variables with highest loadings on each component.

Four pronounced and describable perceptual dimensions were identified: 1) warm/soft, 2) back vowel analogues ([u], [o], [a]), 3) sharp/keen/rough and 4) front vowel analogues ([i], [y], [e], [œ], [æ]). For each component of the PCA, attributes varying significantly between musicians, saxophones or combinations between musician and saxophone were selected to represent the component. Soft and warm were chosen to represent component 1, [o]-like component 2 and sharp and rough component 3. None of the front vowel analogues showed significant variations between the two saxophones or the two musicians at the 95% confidence level.

4.2.3. Differences in judgements between saxophonists and non-saxophonists

Differences in judgements by the saxophonist and non-saxophonist groups were examined using ANOVA. Judgements of the verbal attributes were used as dependent variables. The type of listener (saxophonist or non-saxophonist) was used as the factor. Judgements of all

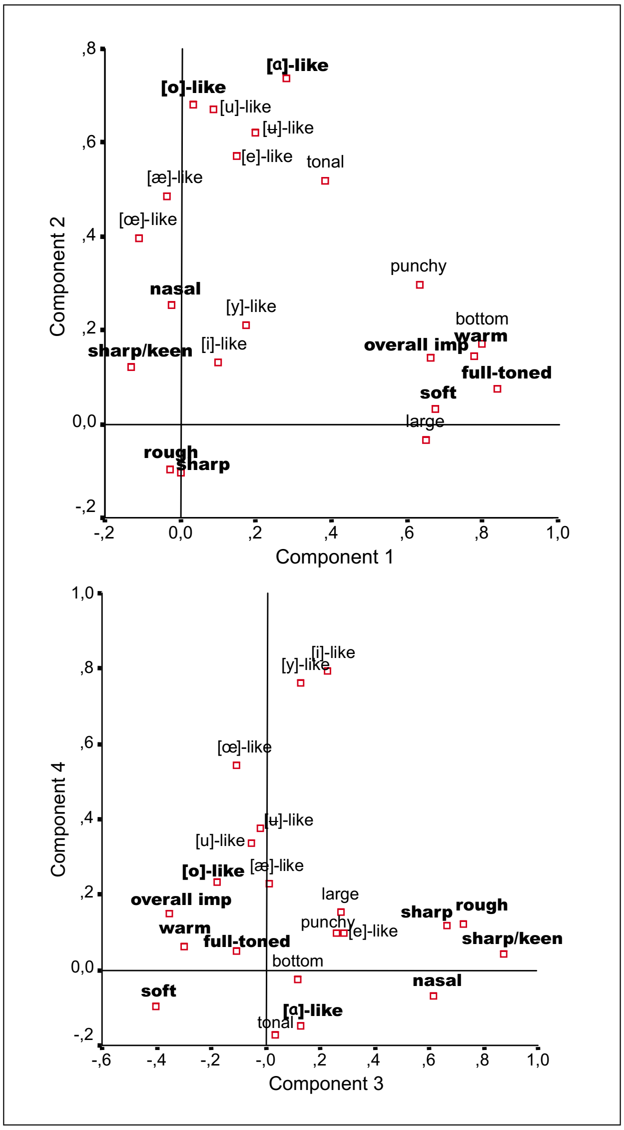

Figure 3. Four VARIMAX rotated components found by PCA. Descriptors in bold were significantly (p=0.05) affected by musician and/or saxophone in the ANOVA.

the 12 test sounds were treated individually. No significant differences (95% confidence level) between the groups were found for the vowel analogues and the attributes "full-toned", "rough", "warm", "soft", "nasal", "punchy", "sharp/keen", "bottom", "sharp". For "large" and "sharp/keen" one sound example was judged significantly different by the two groups. For "tonal" four sound examples were judged significantly different by the groups. The conclusion was that non-saxophonists may be used as subjects in listening tests as long as the attributes "large", "sharp/keen" and "tonal" are avoided.

5. Modeling perceptual dimensions as functions of acoustic metrics

The objective of the third part was to develop models for description of identified perceptual dimensions in terms of acoustic or psychoacoustic metrics.

Table V. Acoustic quantities used as independent variables in the linear regression models.

Index	Description
Fund	Fundamental frequency /Hz
Loudness	Loudness according to ISO532B /sone
Sharpness	Sharpness /acum [16]
Roughness	Roughness /asper [17]
Tonalness	Tonalness [18]
Width	Spectral width: The order number of the highest partial with a SPL above 20 dB
$N'_{4.5}$ to $N'_{22.5}$	Specific loudness per critical band [32] (4.5 Bark-22.5 Bark) /sone/Bark
$R'_{4.5}$ to $R'_{22.5}$	Specific roughness per critical band [32] (4.5 Bark-22.5 Bark) /asper/Bark
F1 to F6	Frequency of formant 1 to 6 (see Figure 2)

5.1. Method

Models for prediction of verbal descriptions based on acoustic metrics were developed by linear regression. As dependent variables, the means of all subjects' judgements of each verbal attribute were used. As independent variables, the acoustic quantities listed in Table V were tested. Suitable independent variables were selected by stepwise estimation [31]. The limits used for probability of F for entry was 0.05 and for removal 0.1.

For those critical bands where specific loudness or roughness [32] made a significant contribution to a model, the correlation with specific loudness and roughness in neighbouring bands was examined. If neighbouring critical bands correlated it was likely that a wider frequency band contributed to the modelled quality. With stepwise estimation one of the critical bands will always result in the highest correlation for model using a specific data set. For another data set another critical band may give the highest correlation. This makes it important to base models on wider frequency bands in cases where neighbouring critical bands correlate. Therefore, models based on sums of neighbouring critical bands were constructed and evaluated to avoid models relying on information in single critical bands. Even though the intention was to keep fundamental frequency and loudness at the same level for all stimuli, small variations did exist. The average fundamental frequency was 419 Hz with a standard deviation of 2.9 Hz. Average loudness was 67 sones (standard deviation 6 sones). To ensure that these variations did not influence judgements, they were included in the analysis.

5.2. Results

The developed linear regression models for prediction of verbal descriptions based on acoustic quantities are presented in Table VI.

6. Validation of developed models

Two requirements are fundamental for a psychoacoustic model to be considered valid. The model should be able to predict judgements by another group of subjects and

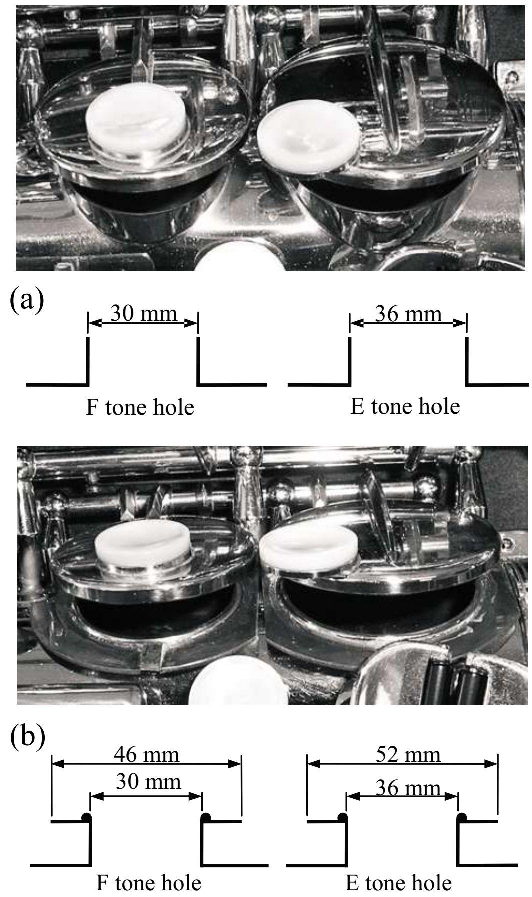

Figure 4. a) Mass produced instrument and b) prototype instrument used for recording of stimuli for validation of models.

the model should be able to predict judgements of other sound samples. Therefore, validation was carried out using listening tests based on a new group of subjects and a new set of stimuli.

6.1. Method

A total of 20 subjects participated in the validation listening tests. The subjects had self-reported normal hearing. As the results from the first listening test showed no significant differences between the judgements of saxophonists and non-saxophonists, for convenience non-saxophonists were used as subjects. The stimuli were loudness equalised (ISO 532B) before use to avoid differences in loudness which might have affected judgements. The adjustments were in the range of −1 dB to +2 dB. The number of judged attributes were fewer than in the other tests described in this paper. Five attributes selected to represent the perceptual dimensions identified in the first listening test were used as variables. These were: sharp, rough, soft, warm and [o]-like. Judgements were made on an 11-point uni-polar scale which ranged from "not at all" (=0) to "extremely" (=10). Two sets of stimuli were judged. Stimuli Set 1 consisted of eight stimuli arbitrary selected from the twelve stimuli used for the development of the models -

Table VI. Linear regression models for prediction of verbal descriptions of saxophone sounds. SS: Sum of squares, MS: Mean square, SE: Standard error.

a) Variables correlating with component 1

Model		ANOVA						Coefficients					
	r^2_{adj}		SS	Df	MS	F	p		B	SE	β	t	p
warm	0.76	Regr.	7.45	1	7.45	35.0	0.000	Constant	19.6	2.6		7.4	0.000
		Resd.	2.13	10	0.21			Sharpness	-12.0	2.0	-0.88	-5.9	0.000
		Totl.	9.58	11									
soft	0.68	Regr.	8.20	1	7.45	24.8	0.001	Constant	20.3	3.3		6.2	0.000
		Resd.	3.31	10	0.331			Sharpness	-12.5	2.5	-0.84	-5.0	0.000
		Totl.	11.5	11									

b) Variable correlating with component 2

Model		ANOVA						Coefficients					
	r^2_{adj}		SS	Df	MS	F	p		B	SE	β	t	p
[o]-like	0.74	Regr.	6.58	2	3.29	17.0	0.001	Constant	18.4	2.5		7.3	0.000
		Resd.	1.74	9	0.19			Sharpness	-9.45	2.0	-0.75	-4.8	0.001
		Totl.	8.32	11				$R'_{9,5} + R'_{10,5}$	-11.1	4.5	-0.38	-2.5	0.036

c) Variables correlating with component 3

Model		ANOVA						Coefficients					
	r^2_{adj}		SS	Df	MS	F	p		B	SE	β	t	p
sharp	0.92	Regr.	23.7	1	23.7	131	0.000	Constant	-3.37	0.78		-4.4	0.001
		Resd.	1.81	10	0.18			$\sum_{i=13.5}^{18.5} N'_i$	0.39	0.034	0.96	11.4	0.000
		Totl.	25.5	11									
rough	0.83	Regr.	8.17	2	4.09	28.4	0.000	Constant	-9.83	2.2		-4.5	0.001
		Resd.	1.30	9	0.14			Sharpness	10.5	1.7	0.78	6.2	0.000
		Totl.	9.47	11				Roughness	2.44	0.79	0.39	3.1	0.013

Table VII. Correlation (r^2) between predictions of the developed models and average results from validation listening tests.
*: Correlation is significant at the 0.05 level (2-tailed, 7 degrees of freedom). **: Correlation is significant at the 0.05 level (2-tailed, 5 degrees of freedom).

	Set 1	Set 2
Variables correlating with component 1:		
warm = $19.6 - 12.0$sharpness	0.78*	0.43
soft = $20.3 - 12.5$sharpness	0.81*	0.54**
Variable correlating with component 2:		
[o]-like = 18.4-9.45sharpness $-11.1(R'_{9.5} + R'_{10.5})$	0.77*	0.67**
Variables correlating with component 3:		
sharp = $-3.37 + 0.39 \sum_{i=13.5}^{18.5} N'_i$	0.91*	0.16
rough = $-9.83 + 10.5$sharpness +2.44roughness	0.77*	0.81**

two of each combination of musician and saxophone. For Stimuli Set 2, new sound samples were created by artificial head recordings of two saxophones - one mass-produced and one prototype instrument. The only controlled difference between the two instruments was the shape of the tone hole rims as illustrated in Figure 4. Other dimensions and materials were the same. The same reed, mouthpiece and neck were used for both saxophones.

The recordings were made under similar conditions as described in 4.1.1. Since the prototype instrument was only playable on two notes, E and F (concert G4 and Ab4) a simple melody switching between these two tones was played. This was closest to normal playing conditions achievable with the prototype instrument. The instruments were played by one musician. Stimuli were created in the same way as described in 4.1.1, by cutting out 0.5 s long excerpts from the steady-state part of a note (concert Ab4, f=415 Hz). Six stimuli were created (3 from each saxophone) by cutting from different parts of the recordings. The developed models were used to predict the perceived timbre and were based on acoustic measurements. The predictions were compared to judgements from the listening tests.

6.2. Results and discussion

The correlations between predictions by the models and results from listening tests based on the two stimuli sets are found in Table VII.

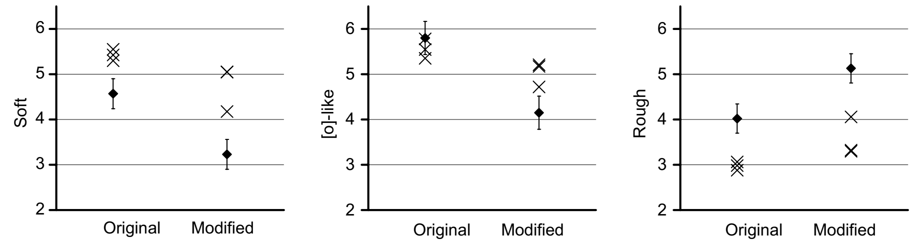

Figure 5. (◆) Averages and 95% Bonferroni intervals of the results from the validation listening test comparing a mass produced and a prototype instrument. (x) Predictions by the developed models at three points in time.

From Table VII it can be seen that the models for soft, [o]-like and rough meet the validation requirement, i.e. predict judgements by another group of subjects and for a new set of stimuli. The results confirm the strength of the psychoacoustic metrics sharpness and roughness, but they also show the ambiguity in people's interpretations of the corresponding verbal attributes. When applied to saxophone sound, rough is interpreted as something that can be described as a mix of the psychoacoustic metrics roughness and sharpness. The attribute soft is a better label for the psychoacoustic metric sharpness (although correlating negatively).

To assess the dynamic character of the steady state parts of saxophone sounds the models were used to measure perceptually relevant variations over time. In Figure 5 predictions at three points in time by the model of three attributes which describe the most prominent perceptual dimensions are shown. The same trend is seen in the predictions as in the results from the listening tests. In all three cases the predictions shifted in magnitude. This is not unexpected since another set of sounds was used for the development of the models. It is well known that perceptual properties of other stimuli presented recently or at the same time may influence judgements. In the literature this is referred to as differential context effects [33]. The spread of predictions based on measurements from different time segments of the recordings was indicative of the normal range of the sound as measured by the developed prediction models. Comparison of the two instruments through listening tests or through the use of the developed models gave the same results. The modified instrument was less soft, less [o]-like and rougher. Since the results were based on a single recording the conclusions should be made with care. Differences may result from the design change but may also originate from other uncontrolled differences between the two instruments and the two recordings. Further studies are required to determine the cause of the differences. For the purpose of validation of the models it is enough to make certain there were audible differences between the recordings. This is obvious from judgements and predictions as displayed in Figure 5.

7. Conclusions

The sequence of interviews and tests reported in this paper form a systematic approach for the identification of salient perceptual dimensions of product sound. In addition, the approach enables development of models for prediction of how steady-state parts of musical instrument sounds are perceived when based on psychoacoustic measurements. This is a good foundation for musical instrument sound specifications. The results presented are specific for the alto saxophone, but the suggested method should make it possible to identify salient perceptual dimensions and relevant psychoacoustic models for other instruments and products.

For the examined alto saxophones the attributes; sharp, rough, soft, warm, and [o]-like described differences in steady state parts of the sounds and were judged significantly different depending on which saxophone and which saxophonist. Four perceptual dimensions were identified: 1. warm/soft, 2. back vowel analogues, 3. sharp/rough and 4. front vowel analogues. The first three could be modelled as functions of psychoacoustic metrics. The psychoacoustic metrics of sharpness and roughness described much of the perceptually relevant variations in saxophone sound. It was also possible to identify critical bands where specific roughness was of particular importance. The best prediction model for the attribute "rough" was a mix of Aures' roughness model [17] and Von Bismarck's sharpness model [16]. This suggests that the Swedish word for rough ("rå") is interpreted as a mix between the established psychoacoustic models for roughness and sharpness. The models were shown to be useful for identification of the physical properties of sound which influence perceived character. The applied procedure resulted in a manageable number of commonly understood verbal attributes, each interpreted in terms of acoustic metrics. Inclusion of such verbal and physical descriptions in a specification should be of great help in the design process of musical instruments and other sound-generating products.

References

[1] K. T. Ulrich, S. D. Eppinger: Product design and development. 4th ed. McGraw-Hill, Boston, Massachusetts, 2008.

[2] J. Blauert, U. Jekosch: Sound-quality evaluation – A multi-layered problem. Acustica united with Acta Acustica **83** (1997) 747–753.

[3] R. Guski: Psychological methods for evaluating sound quality and assessing acoustic information. Acustica united with Acta Acustica **83** (1997) 765–774.

[4] G. Kelly: The psychology of personal constructs. Norton, New York, 1955.

[5] J. Berg, F. Rumsey: Identification of quality attributes of spatial audio by repertory grid technique. J. Audio Eng. Soc. **54** (2006) 365–379.

[6] H. Stone, J. Sidel, S. Oliver, A. Woolsey, R. C. Singleton: Sensory evaluation by quantitative descriptive analysis. Food Technology (1974) 24–34.

[7] A. Gabrielsson: Dimension analyses of perceived sound quality of sound-reproducing systems. Scand. J. Psychol. **20** (1979) 159–169.

[8] E. Bigand: Contributions of music to research on human auditory cognition. – In: Thinking in sound – The cognitive psychology of human audition. S. McAdams, E. Bigand (eds.). Oxford University Press, Oxford, 1993, 231–277.

[9] B. S. Rosner, J. B. Pickering: Vowel perception and production. Oxford University Press, Oxford, 1994.

[10] R. A. Smith, D. M. A. Mercer: Possible causes of woodwind tone colour. J. Sound and Vibration **32** (1974) 347–358.

[11] L. Wedin, G. Goude: Dimension analysis of the perception of instrumental timbre. Scand. J. Psychol. **13** (1972) 228–240.

[12] A. H. Benade, S. J. Lutgen: The saxophone spectrum. J. Acoust. Soc. Am. **83** (1988) 1900–1907.

[13] J. Kruskal, M. Wish: Multidimensional scaling. Sage Publications, Beverly Hills, 1978.

[14] C. E. Osgood, G. J. Suci, P. H. Tannenbaum: The measurement of meaning. University of Illinois Press, 1957.

[15] E. Zwicker: A method for calculation of loudness. Acustica **10** (1960) 304–308.

[16] G. von Bismarck: Sharpness as an attribute of the timbre of steady sounds. Acustica **30** (1974) 159–172.

[17] W. Aures: Ein Berechnungsverfahren der Rauhigkeit. Acustica **58** (1985) 268–281.

[18] W. Aures: Berechnungsverfahren für den sensorischen Wohlklang beliebiger Schallsignale. Acustica **59** (1985) 130–141.

[19] M. Bodden: Instrumentation for sound quality evaluation. Acustica united with Acta Acustica **83** (1997) 775–783.

[20] U. Widman: Aurally adequate evaluation of sounds. Proc. Euro-Noise 98, München, 1998.

[21] C. V. Beidl, W. Stücklschwaiger: Application of the AVL-annoyance index for engine noise quality development. Acustica united with Acta Acustica **83** (1997) 789–795.

[22] M. S. Kahn, O. Johansson, U. Sundbäck: Development of an annoyance index for heavy-duty diesel engine noise using multivariate analysis. Noise Control Eng. J. **45** (1997) 157–167.

[23] R. A. Kendall, E. C. Carterette: Verbal attributes of simultaneous wind instrument timbres. 1. Von Bismarck adjectives. Music Perception **10** (1993) 445–468.

[24] G. von Bismarck: Timbre of steady sound: a factorial investigation of its verbal attributes. Acustica **30** (1974) 146–158.

[25] M. C. Gridley: Trends in description of saxophone timbre. Perceptual and Motor Skills **65** (1987) 303–311.

[26] V. Rioux: Methods for an objective and subjective description of starting transients of some flue organ pipes – Integrating the view of an organ-builder. Acustica united with Acta Acustica **86** (2000) 634–641.

[27] J. Stephanek, Z. Otcenasek: Psychoacoustic aspects of violin sound quality and its spectral relations. Proc. 17th International Congress on Acoustics, Rome, Italy, 2001.

[28] R. L. Pratt, P. E. Doak: A subjective rating scale for timbre. J. Sound and Vibration **45** (1976) 317–328.

[29] J. Blauert, K. Genuit: Evaluating sound environments with binaural technology – some basic considerations. J. Acoust. Soc. Jpn. (E) **14** (1993) 139–145.

[30] P. Iverson, C. L. Krumhansl: Isolating the dynamic attributes of musical timbre. J. Acoust. Soc. Am. **94** (1993) 2595–2603.

[31] J. F. Hair, R. E. Anderson, R. L. Tatham, W. C. Black: Multivariate data analysis. 5ed. Prentice Hall International, London, 1998.

[32] E. Zwicker, H. Fastl: Psychoacoustics – Facts and models. 2nd ed. Springer-Verlag, 1999.

[33] L. E. Marks, Y. Arieh: Differential effects of stimulus context in sensory processing. Revue européenne de psychologie appliquée **56** (2006) 213–221.