

Szepannek, Gero; Ligges, Uwe; Luebke, Karsten; Raabe, Nils; Weihs, Claus

Working Paper

Local Models in Register Classification by Timbre

Technical Report, No. 2005,47

Provided in Cooperation with:

Collaborative Research Center 'Reduction of Complexity in Multivariate Data Structures' (SFB 475),
University of Dortmund

Suggested Citation: Szepannek, Gero; Ligges, Uwe; Luebke, Karsten; Raabe, Nils; Weihs, Claus (2005) : Local Models in Register Classification by Timbre, Technical Report, No. 2005,47, Universität Dortmund, Sonderforschungsbereich 475 - Komplexitätsreduktion in Multivariaten Datenstrukturen, Dortmund

This Version is available at:

<https://hdl.handle.net/10419/22640>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Local Models in Register Classification by Timbre

G. Szepannek U. Ligges K. Luebke N. Raabe
C. Weihs

Lehrstuhl für Computergestützte Statistik
Fachbereich Statistik
Universität Dortmund

Abstract Investigating a data set containing different sounds of several instruments suggests that local modelling may be a promising approach to take into account different timbre characteristics of different instruments. For this reason, some basic ideas towards a local modelling are realized in this report yielding a framework for further studies.

1 Introduction

Sound characteristics of orchestra instruments derived from spectra are currently a very important research topic (see, e.g., Reuter (1996, 2002)). The sound characterization of voices has, however, many more facets than for instruments because of the sound variation in dependence of technical level and emotional expression (see, e.g., Kleber (2002)).

During a former analysis of singing performances (cp. Weihs and Ligges (2003)) it appeared that register can be identified from the spectrum even after elimination of pitch information. In this paper this observation is assessed

by means of a systematic analysis not only based on singing performances but also on corresponding tones of high and low pitched instruments.

The aim of this report is to investigate some basic ideas of local modelling towards achieving the goal of classification of the register by timbre, i.e. by the spectrum after pitch information is eliminated. To this end, pitch independent characteristics of spectral densities of instruments and voices are generated.

Similar to the voice prints introduced in Weihs and Ligges (2003) we use masses and widths of peaks of the first 13 partials, i.e. of the fundamental and the first 12 overtones. These characteristics are computed for representatives of all tones involved in the classical song “Tochter Zion” composed by G.F. Händel. For the singing performances the first representative of each note was chosen, for the instruments the representatives were chosen from the “McGill University Master Samples” (Opolko and Wapnick (1987), see also Section 2). A global classification analysis is performed in Section 3. In Section 4 these models are extended towards local modelling of the different instruments and singers.

2 Data

The analyses of this paper are based on time series data from an experiment with 17 singers performing the classical song “Tochter Zion” (Händel) to a standardized piano accompaniment played back by headphones (cp. Weihs et al. (2001)). The singers could choose between two accompaniment versions transposed by a third in order to take into account the different voice types (Soprano and Tenor vs. Alto and Bass). Voice and piano were recorded at different channels in CD quality, i.e. the amplitude of the corresponding vibrations was recorded with constant sampling rate 44100 hertz in 16-bit format. The audio data sets were transformed by means of a computer program into wave data sets. For time series analysis the waves were reduced to 11025 Hz (in order to restrict the number of data), and standardized to the interval $[-1, 1]$. Since the volume of recording was already controlled individually, a comparison of the absolute loudness of the different recordings was not sensible anyway. Therefore, by our standardization no additional information was lost.

Since our analyses are based on characteristics derived from tones corresponding to single notes, we used a suitable segmentation procedure (Ligges

et al. (2002), Weihs and Ligges (2005)) in order to get data of segmented tones corresponding to notes. The periodograms (cp. Brockwell and Davis (1991)) used for the analyses described in this paper were calculated from overlapping sections of 2048 observations, overlap starting in the middle of the preceding section. This way, we get roughly $11 (= 2 \cdot (11025/2048))$ periodograms per second of sound, whereas the duration of the whole song is roughly 60 seconds. These periodograms are classified to notes, and the notes are smoothed by means of double median smoothing. Based on the smoothed series of notes, begin and end of sung notes are decided upon. For further analysis the first representative of the notes with identical pitch in the song was chosen. This leads to 9 different representatives per voice in “Tochter Zion”.

The notes involved in the analyzed song were also identified in the “McGill University Master Samples” either in the Alto or in the Bass version for the following instruments:

Alto version (McGill notation, referred to as ‘A’): *aflute-vib, bells, cello-bv, clari-bfl, clari-efl, elecuitar1, elecuitar4, enghorn, flute-flu, flute-vib, frehorn, frehorn-m, marimba, oboe, piano-ld, piano-pl, piano-sft, sax-alt, tromb-ten, trump-ba, trump-c, trump-csto, vibra-bow, vibra-hm, viola-bv, viola-mv, violin-bv, violin-mv.*

Bass version (referred to as ‘B’): *bassoon, bflute-flu, bflute-vib, cello-bv, elecbass1, elecbass5, elecbass6, elecuitar1, elecuitar2, elecuitar4, frehorn, frehorn-m, marimba, piano-ld, piano-pl, piano-sft, tromb-ten, tromb-tenm, tuba, viola-mv.*

Thus, 28 high instruments and 20 low instruments were chosen together with 10 high female singers and 7 male.

From the periodogram corresponding to each tone corresponding to an identified note voice characteristics are derived (cp. Weihs and Ligges (2003)). For our purpose we only use the size and the shape corresponding to the first 13 partials, i.e. to the fundamental frequency and the first 12 overtones, in a pitch independent periodogram.

In order to measure the size of the peaks in the spectrum, the mass (weight) of the peaks of the partials are determined as the sum of the percentage shares of those parts of the corresponding peak in the spectrum which are higher than a pre-specified threshold. The shape of a peak cannot easily be described. Therefore, we only use one simple characteristic of the shape, namely the width of the peak of the partials. The width of a peak is measured by the range (i.e. maximum - minimum) of the Fourier frequencies (in

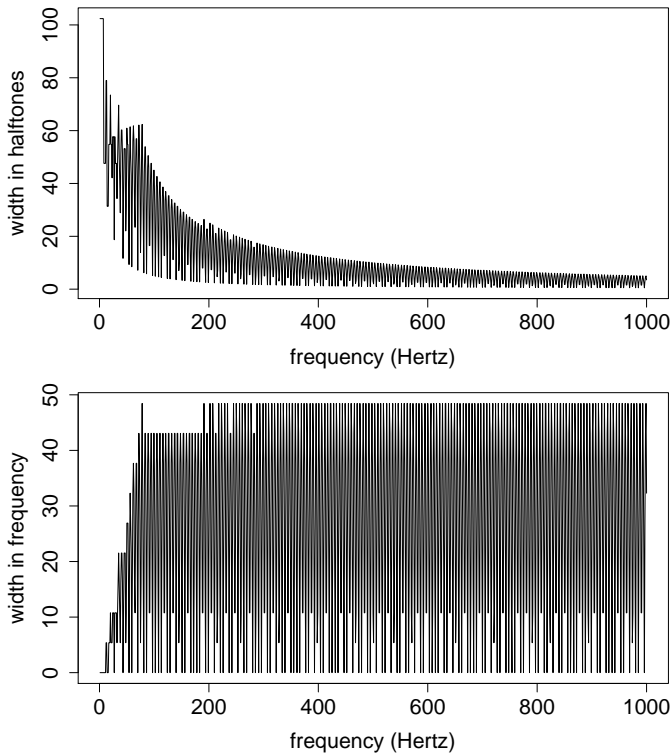


Figure 1: Widths of peaks measured in (a) halftones and (b) frequency.

Hertz) of the peak with a spectral height above a pre-specified threshold.

A former idea for measuring the width of peaks in voiceprints was to use a unit of halftones (as described in Weihs et al. (2005b)). It was our intuition that this measure would be independent of frequency and hence appropriate for our analyses. Unfortunately, it turned out that this measure is *not* independent of the frequency. Instead, the width measured by the number of Fourier frequencies (or proportionally just by the difference of frequencies) appears to be much more appropriate.

Figure 1 shows the two widths (measured in halftones and in difference of frequencies, respectively) of the fundamental frequency of a simulated combined sine wave consisting of two partials. The first partial is weighted by 0.6 and the second by 0.4. This was simulated for all integer frequencies between 1 and 1000 Hertz using a sampling rate of 11025 Hertz. Measures are based on a periodogram consisting of 2048 observations. Therefore, the

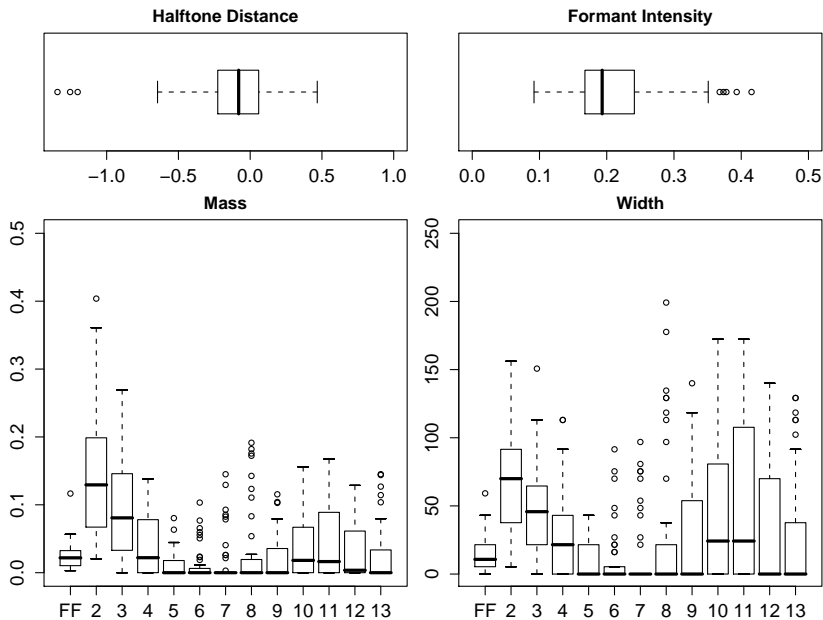


Figure 2: Voice print of professional bass singer.

distance between two Fourier frequencies is 5.38 Hertz.

Overall, every tone is characterized by the above 26 characteristics which are used as a basis for classification. For details on the computation of the measures see Güttner (2001). Mass is measured as a percentage (%), whereas width is measured in Hertz.

Figure 2 illustrates the voice print corresponding to the whole song “Tochter Zion” for a particular singer. For masses and widths boxplots are indicating variation over the involved tones. For the analyses of this paper we ignore halfnote distance and formant intensity (cp. Weihs and Ligges (2003)), and use the other characteristics of the voice print for individual tones, as well as averaged characteristics over all involved tones, leading to only one value for each characteristic per harmonic and singer or instrument.

3 Global modelling

Classification of the register of different instruments and singers is performed using two very common techniques: the classical linear discriminant analysis [LDA] (Fisher (1936)) as well as Classification trees (more specifically the

rpart by Therneau and Atkinson (1997)). All error rates are estimates using ten-fold cross validation. In the analysis of the following two subsections every note is used once. The variables are generated both for every note separately and averaged over all notes which yields one single value of mass and width per harmonic and singer/instrument. In the following, LDA and rpart are compared. Only the result of the better method is shown.

3.1 Mass and width not averaged over different notes

3.1.1 Analyzing singers only

A first analysis includes only singers. Performing classification based on all variables shows best results for LDA (error: 0.222, see table 1).

Voice	high	low	error
Soprano	35	1	0.028
Alto	45	9	0.167
Tenor	14	13	0.519
Bass	10	26	0.278

Table 1: Classification of voices on all variables (masses and widths for all harmonics) shows best results for LDA.

Classification with rpart using only mass of the fundamental slightly improves the error rates to 0.216 (table 2).

Voice	high	low	error
Soprano	35	1	0.028
Alto	43	11	0.204
Tenor	12	15	0.444
Bass	9	27	0.250

Table 2: Classification of voices with rpart using only mass of the fundamental.

Taking the mass and width of the fundamental and the first harmonic into the model further improves the missclassification-rate to 0.157 using LDA (table 3).

Voice	high	low	error
Soprano	35	1	0.028
Alto	43	11	0.204
Tenor	9	18	0.333
Bass	3	33	0.083

Table 3: Misclassification rates using LDA on first two harmonics.

Generally, the female voices were easier to classify into the high register than those of the male singers into the low register. The most confusing are the lower female and the higher male voices.

As can be seen, the reduction of the model enables a more correct classification of the male voices while classification-rates of the female voices stay at the same level. A final approach is dedicated towards finding an optimal variable subset. A variable selection is performed using *stepclass* algorithm (see Weihs et al. (2005a)). Variables are added stepwise to the model until the *Leave one out cross validated error rate [L1o]* can not be improved by more than 1%.

Best results are obtained keeping mass of the fundamental and the first harmonics as well as width of the first and fourth harmonic in the model when using Linear Discriminant Analysis. The total error rate was 0.144 (table 4).

Voice	high	low	error
Soprano	35	1	0.028
Alto	44	10	0.185
Tenor	8	19	0.296
Bass	3	33	0.083

Table 4: Variable selection by *stepclass* using LDA.

3.1.2 Taking instruments into the analysis

The analysis is now re-done also taking the instruments into the study. This leads to a strongly increased error rate, for the all-variables-model there is an error rate of 0.352 for LDA (see table 5).

instrument	high	low	error
A.bells	7	2	0.222
A.cello	7	2	0.222
A.clarinet	16	2	0.111
A.elecguitar	13	5	0.278
A.flute	19	8	0.296
A.frenchhorn	18	0	0.000
A.marimba	9	0	0.000
A.oboe-enghorn	11	7	0.389
A.piano	26	1	0.037
A.saxophone	5	4	0.444
A.trombone	6	3	0.333
A.trumpet	12	6	0.333
A.trumpet-csto	6	3	0.333
A.vibraphone	18	0	0.000
A.violin-viol	31	5	0.139
Alto	43	11	0.204
Soprano	25	11	0.306
B.bassoon	1	8	0.111
B.cello	4	5	0.444
B.elecbass	16	11	0.593
B.elecguitar	12	6	0.667
B.elecguitar-dist	9	0	1.000
B.flute	13	5	0.722
B.frenchhorn	3	15	0.167
B.marimba	9	0	1.000
B.piano	14	13	0.519
B.trombone	6	12	0.333
B.tuba	4	5	0.444
B.violin-viol	7	2	0.778
Bass	20	16	0.556
Tenor	18	9	0.667

Table 5: Misclassification rates of LDA for both instruments and voices using all variables.

Reducing to only the values of mass and width of the fundamental and the first harmonic could not further improve the classification errors as in the previous section when only classifying singers. The error rates of some instruments, e.g. bells or the trumpet strongly grow using less variables while for example the low piano sounds or the tenor voices can be classified better into their register. The register of some instruments like electric guitars are identifiable in none of the attempts. The results of both LDA (total error rate: 0.451) and rpart (total error rate: 0.344) are shown in table 6.

instrument	LDA			rpart		
	high	low	error	high	low	error
A.bells	1	8	0.889	1	8	0.889
A.cello	8	1	0.111	2	7	0.778
A.clarinet	17	1	0.056	12	6	0.333
A.elecguitar	13	5	0.278	3	15	0.833
A.flute	24	3	0.111	12	15	0.556
A.frenchhorn	17	1	0.056	4	14	0.778
A.marimba	9	0	0.000	9	0	0.000
A.oboe-enghorn	11	7	0.389	5	13	0.722
A.piano	24	3	0.111	22	5	0.185
A.saxophone	8	1	0.111	5	4	0.444
A.trombone	6	3	0.333	2	7	0.778
A.trumpet	8	10	0.556	1	17	0.944
A.trumpet-csto	8	1	0.111	3	6	0.667
A.vibraphone	12	6	0.333	16	2	0.111
A.violin-viol	30	6	0.167	23	13	0.361
Alto	36	18	0.333	22	32	0.593
Soprano	25	11	0.306	18	18	0.500
B.bassoon	2	7	0.222	2	7	0.222
B.cello	8	1	0.889	4	5	0.444
B.elecbass	15	12	0.556	16	11	0.593
B.elecguitar	12	6	0.667	9	9	0.500
B.elecguitar-dist	7	2	0.778	9	0	1.000
B.flute	10	8	0.556	9	9	0.500
B.frenchhorn	2	16	0.111	2	16	0.111
B.marimba	5	4	0.556	6	3	0.667
B.piano	10	17	0.370	7	20	0.259
B.trombone	5	13	0.278	1	17	0.056
B.tuba	7	2	0.778	4	5	0.444
B.violin-viol	5	4	0.556	5	4	0.556
Bass	13	23	0.361	5	31	0.139
Tenor	15	12	0.556	3	24	0.111

Table 6: Classification error of LDA and rpart using only fundamental and first harmonic.

3.2 Mass and width averaged over different notes

The analysis in this section differs from the preceding section in the calculation of the variables. Here, the values of mass and width for the harmonics are averaged over all different notes leading to two values (mean mass and mean width) for each harmonic. By averaging, the number of different ob-

instrument	all variables			only mass fund.			fund. and 1 st harm.		
	high	low	error	high	low	error	high	low	error
Soprano	2	2	0.500	4	0	0.000	2	2	0.500
Alto	5	1	0.167	3	3	0.500	3	3	0.500
A	21	7	0.250	13	15	0.536	15	13	0.464
Tenor	1	2	0.333	0	3	0.000	0	3	0.000
Bass	1	3	0.250	0	4	0.000	1	3	0.250
B	9	11	0.450	9	11	0.450	7	13	0.350

Table 7: Classification error of LDA averaging over different notes.

jects to classify is reduced. The number of different singers’ objects reduces to only 17. For this reason the detailed results for singers only are omitted here. LDA shows best results for the classification of both singers and instruments together. The analysis is performed using all variables (error rate using leave-one-out cross validation: 0.323), only the mass of the fundamental (0.415) and both mass and width of fundamental and the first harmonic (0.4). The register classification of the instruments still shows unsatisfying results. The best overall-error is obtained using all variables. All variables appear to contain some information for the classification of at least some of the instruments, while one can state that the voices can be classified into register quite well using only the mass of the fundamental. The classification rule for singers when only using the fundamental mass assigns to low register if the mass is below 0.1072. All results are shown in table 7, where *A* denotes alto (high register) and *B* denotes bass (low register) for the instruments.

3.3 Summarizing the results

Timbre characteristics mass and width of the fundamental and the first up to the 12th harmonics enables to classify voices into register. Female voices seem to be easier classified than male voices. The most confounding are low female voices (alto) and high male voices (tenor).

When taking instruments into the analysis the classification error rate increases dramatically. Also variable selection cannot improve the results to an acceptable magnitude. This may be due to different timbre characteristics that characterize the register of different instruments. For this reason, in the following section, local models for each of the instruments (or voices) are built.

instrument	selected variables	# cases	L1o error
cello	mass02 + mass05	18	0.056
elecguitar	mass02 + mass03 + mass04 + width02	36	0.111
flute	mass06 + width04	45	0.200
frenchhorn	mass08	36	0.056
marimba	width01	18	0.111
piano	mass01 + mass07	54	0.148
singer	width01	153	0.209
trombone	mass01	27	0.074
violin, viola	mass13	45	0.111
weighted L1o error		432	0.150

Table 8: Best performing local models using LDA for each instrument group.

4 Local modelling

In this section, local models are built for all instruments separately (indexed by l , where $\{1, \dots, L\}$ represent all instruments). For each instrument a local classification rule for the register is developed, returning posterior probabilities $p_l(k|x)$ given the instrument l and an observed tone with masses and widths x to belong to either high or low register, denoted as class k .

Since the goal consists in classification of *any* observed note to the correct register, the instrument (or voice) playing this note can not assumed to be known in advance. For this reason, the problem consists in finding a classification rule out of all L local classification models.

All local models are generated on variable subsets, found by the stepclass-algorithm using LDA (see Weihs et al. (2005a)). Variables are added stepwise to the data set until the *L1o* cross validated error rate can not be improved by more than 1%.

How classification rules may be constructed out of all local (classification) models l , $l = 1, \dots, L$, will be explained in the following subsections.

For each local model l parameters are estimated and all observations of the complete data set were predicted with all local models. Of course, only those instruments are used in the analysis for local modelling that have both register classes, namely: *cello*, *elecguitar*, *flute*, *frenchhorn*, *marimba*, *piano*, *singer*, *trombone*, *violin-viol*a.

Table 8 shows an example of LDA based local models for various instruments. For each model a variable selection was performed by stepclass as

mentioned above. Selected variables are shown in the table where *mass01* denotes the mass of the first partial (fundamental frequency) and *width02* the width of the second partial. Moreover, the number of cases used for L1o cross validation is shown as well as the L1o error rate. The L1o error weighted by the number of observations falling in the different local models is 0.15. This is a lower error bound for the error rate that would be possible in case of a perfect choice of the correct local models for a new observation of an unknown instrument.

instrument	# cases	L1o error
cello	18	0.333
elecuitar	36	0.278
flute	45	0.222
frenchhorn	36	0.139
marimba	18	0.111
piano	54	0.241
singer	153	0.150
trombone	27	0.222
violin, viola	45	0.200
weighted L1o error	432	0.194

Table 9: Best model depending on variables *mass01*, *mass02*, *width01*, and *width02* while each local model fitted separately using LDA.

In a second idealized step, let us assume that the correct instrument of a new observation is still known, but we do want a global variable selection used for all local models. In this case, *stepclass* variable selection results in a model including the variables *mass01*, *mass02*, *width01*, and *width02*, i.e. the four measures of the first two partials. The corresponding L1o error rate is 0.194 (see table 9).

4.1 Comparing posterior probabilities

A first straightforward intuition consists in comparing the posterior probabilities of the different local models. This is possible by classifying an object to the class \hat{k} with the highest posterior probability (*maximum-posterior rule*) of all local models l

$$\hat{k} = \arg \max_k \left(\max_l p_l(k|x) \right) \quad (1)$$

chosen local models									
instrument	cel.	el.guit.	flute	french.	marim.	piano	sing.	tromb.	viol.
cello	0	5	0	1	11	0	0	1	0
elec guitar	8	0	0	0	25	0	0	3	0
flute	4	1	0	0	21	0	0	19	0
french horn	0	0	4	0	32	0	0	0	0
marimba	0	0	0	0	0	0	0	18	0
piano	1	1	0	0	32	0	0	20	0
singer	30	16	11	1	64	1	0	25	5
trombone	2	0	0	0	23	0	0	2	0
violin-viol	2	9	1	0	23	1	0	9	0

Table 10: Frequency of chosen local models when using the maximum-posterior-rule.

or to that class \hat{k} possessing the largest average posterior probability averaged over all local models (*average-posterior rule*).

$$\hat{k} = \arg \max_k \sum_l p_l(k|x) \quad (2)$$

A third competitor is *majority voting* of the local classification rules

$$\hat{k} = \arg \max_k \sum_l I_{[p_l(k|x) > p_l(j|x) \forall j \neq k]}(k) \quad (3)$$

The variable subset selection is performed for every local model separately, leading to possibly different variable subsets of the different local models (= instruments).

Using the ‘maximum-posterior’ classification rule (equation 1) results in an error rate of 0.479. Both averaging the local posterior probabilities and majority voting lead to an improvement of the total error rate to the same value of 0.391. Table 10 indicates, how often each of the local models is chosen, when using the ‘maximum-posterior’-rule.

Surprisingly, using the maximum-posterior classification rule (see equation 1) more than 50% of the chosen local models are of class ‘marimba’. On the other hand, the singer’s local model is never chosen. Ignoring the possibly misleading local marimba models for the analysis can slightly but not dramatically improve the total error rates up to 0.42 (0.365 for averaging and 0.388 for majority voting).

4.2 Comparing densities

In some situation an observation may have a large posterior probability of class k close to one in some local model l even if its density $f_l(x|k)$ given this class (here: the register) and the instrument is very small.

Assume for simplicity equal prior probabilities $p_l(k) = p_l(1) \forall k, l$. Then,

$$p_l(k|x) = \frac{f_l(k|x)}{\sum_k f_l(k|x)} = \frac{f_l(x, k)}{f_l(x)p_l(1)} \frac{p_l(1)f_l(x)}{\sum_k f_l(x, k)} = \frac{f_l(x|k)}{\sum_k f_l(x|k)}. \quad (4)$$

This means $p_l(k|x) \sim 1$ if $f_l(x|k) \gg f_l(x|j), \forall j \neq k$. For some other class $j_1 \neq k$ both inequalities $f_m(x|j_1) \gg f_l(x|k)$ and $f_m(x|j_1) > f_m(x|k)$ may still hold for a different local model $m \neq l$ even if the posterior probability of this class in model m is smaller than those of class k in model l .

For this reason, densities may be better to be compared rather than posterior probabilities.

Classification rules may then be derived from (*maximum-density rule*)

$$\hat{k} = \arg \max_k \left(\max_l f_l(x|k) \right) \quad (5)$$

or (*average-density rule*)

$$\hat{k} = \arg \max_k \sum_l f_l(x|k). \quad (6)$$

Comparing the densities of the different local models is questionable if they are built on different variables. Therefore in this situation the variable selection has to be performed simultaneously for all local models in common.

Classifying according to the local model with maximum density (see equation 5) yields a total L1o error rate of 0.354 (see table 11). Averaging over all local model densities results in an error rate of 0.301.

Note that comparing posteriors and densities yields in identical results in the case of known subclasses, hence we refer to table 9 for lower error bounds.

Classifying on the whole data set of all 26 variables without any variable selection here leads to error rates of 0.368 (maximum-density rule, equation 5) or 0.366 (averaging, equation 6).

4.3 Global weighting of local models

Simply comparing posterior probabilities or densities of local models may work well if all instruments have the same probability $\pi(l)$ to appear. But

instrument	chosen local models								
	cel.	el.guit.	flute	french.	marim.	piano	sing.	tromb.	viol.
cello	0	0	2	0	0	1	2	2	11
elecuitar	0	3	9	0	0	2	7	7	8
flute	1	3	14	0	1	6	8	0	12
frenchhorn	0	2	4	0	0	0	0	16	14
marimba	3	2	6	0	6	1	0	0	0
piano	0	4	17	1	0	3	4	6	19
singer	1	6	40	4	0	21	29	24	28
trombone	0	1	5	0	0	2	2	6	11
violin-viol	0	0	11	0	1	0	4	5	24

Table 11: Frequency of chosen local models at maximum-density rule.

these measures do not take into account that on the other hand the probability of an observed object to belong to this local partition (corresponding to a certain local model) of the population can be very small. Therefore a weighting of the posterior probabilities is done in this attempt by performing a double classification. Besides predicting by the local models, a *global step* is added that classifies an observation into one of the local models, returning posterior probabilities $\pi(l)$ that the observed tone comes from instrument l . The result can be used to choose the local model (*global model choice rule*):

$$\hat{k} = \arg \max_k p_l(k|x), \quad (7)$$

$$\hat{l} = \arg \max_l \pi(l). \quad (8)$$

The local results (in form of local posterior probabilities) also can be weighted according to the posterior probabilities of the global modelling step (*global model weighting rule*):

$$\hat{k} = \arg \max_k \sum_l p_l(k|x)\pi(l). \quad (9)$$

Since again posteriors and not densities are compared in this section, the variable selection can be performed for every local model separately. Also the variables used for calculating the $\pi(l)$ in the *global step* must not be the same as in any of the local models.

When choosing the local model by LDA and variable selection in the global step (*width02*, *mass01*) as in the local step (as used in table 9: *mass01*, *mass02*, *width01*, and *width02*) this led to an overall error rate of 0.285. But

the error in choosing the right local model is 0.564. Weighting the local posteriors by the global posteriors reduces the error to be 0.269.

5 Summary

Classification of register in music is performed using timbre characteristics of the first 13 harmonics. Especially global classification rules are replaced by several local ones. Different aspects are shown, how this can be used for classification, yielding ameliorations of the misclassification rate.

The presented proceeding can be understood as some basis for future work allowing a more comprehensible and more effective classification of musical data by issues of local modelling. One topic of further research may take into account that variable subsets of different local models may differ. At the moment, comparison of such local models is difficult.

Acknowledgment. This work has been supported by the Collaborative Research Center ‘Reduction of Complexity in Multivariate Data Structures’ (SFB 475) of the German Research Foundation (DFG).

References

- BROCKWELL, P.J. and DAVIS, R.A. (1991): *Time Series: Theory and Methods*. Springer, New York.
- FISHER, R.A. (1936): The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 7, 179–188.
- GÜTTNER, J. (2001): *Klassifikation von Gesangsdarbietungen*. Diploma Thesis, Fachbereich Statistik, Universität Dortmund, Germany.
- KLEBER, B. (2002): *Evaluation von Stimmqualität in westlichem, klassischen Gesang*. Diploma Thesis, Fachbereich Psychologie, Universität Konstanz, Germany
- LIGGES, U., WEIHS, C. and HASSE-BECKER, P. (2002): Detection of Locally Stationary Segments in Time Series. In: W. Härdle and B. Rönz (Eds.): *COMPSTAT 2002 - Proceedings in Computational Statistics - 15th Symposium held in Berlin, Germany*. Physika, Heidelberg, 285–290.

- OPOLKO, F., and WAPNICK, J. (1987): McGill University Master Samples. McGill University, Quebec, Canada. URL: <http://www.music.mcgill.ca/resources/mums/html/index.htm>
- REUTER, C. (1996): *Die auditive Diskrimination von Orchesterinstrumenten - Verschmelzung und Heraushörbarkeit von Instrumentalklangfarben im Ensemblespiel*. Peter Lang, Frankfurt/M.
- REUTER, C. (2002): *Klangfarbe und Instrumentation - Geschichte - Ursachen - Wirkung*. Peter Lang, Frankfurt/M.
- THERNEAU, T.M. and ATKINSON, E.J. (1997): *An Introduction to Recursive Partitioning Using the RPART Routines*. Technical Report, Mayo Foundation.
- WEIHS, C., BERGHOFF, S., HASSE-BECKER, P. and LIGGES, U. (2001): Assessment of Purity of Intonation in Singing Presentations by Discriminant Analysis. In: J. Kunert and G. Trenkler (Eds.): *Mathematical Statistics and Biometrical Applications*. Josef Eul, Köln, 395–410.
- WEIHS, C. and LIGGES, U. (2003): Voice Prints as a Tool for Automatic Classification of Vocal Performance. In: R. Kopiez, A.C. Lehmann, I. Wolther and C. Wolf (Eds.): *Proceedings of the 5th Triennial ESCOM Conference*. Hanover University of Music and Drama, Germany, 8-13 September 2003, 332–335.
- WEIHS, C. and LIGGES, U. (2005): From Local to Global Analysis of Music Time Series. In: K. Morik, J.-F. Boulicaut and A. Siebes (Eds.): *Local Pattern Detection, Lecture Notes in Artificial Intelligence, 3539*, Springer, Berlin, 217–231.
- WEIHS, C., LIGGES, U., LUEBKE, K. and RAABE, N. (2005a): klaR Analyzing German Business Cycles. In: D. Baier, R. Becker and L. Schmidt-Thieme (Eds.): *Data Analysis and Decision Support*, Springer, Berlin, 335–343.
- WEIHS, C., REUTER, C. and LIGGES, U. (2005b): Register Classification by Timbre. In: C. Weihs and W. Gaul (Eds.): *Classification: The Ubiquitous Challenge*, Springer, Berlin, 624–631.