ECONSTOR

WWW.ECONSTOR.EU

Der Open-Access-Publikationsserver der ZBW – Leibniz-Informationszentrum Wirtschaft
The Open Access Publication Server of the ZBW – Leibniz Information Centre for Economics

Ligges, Uwe; Reuter, Christoph; Weihs, Claus

Working Paper

Register Classification by Timbre

Technical Report / Universität Dortmund, SFB 475 Komplexitätsreduktion in Multivariaten Datenstrukturen, No. 2004,71

Provided in cooperation with:
Technische Universität Dortmund

Suggested citation: Ligges, Uwe; Reuter, Christoph; Weihs, Claus (2004) : Register Classification by Timbre, Technical Report / Universität Dortmund, SFB 475 Komplexitätsreduktion in Multivariaten Datenstrukturen, No. 2004,71, http://hdl.handle.net/10419/22584

Nutzungsbedingungen:
Die ZBW räumt Ihnen als Nutzerin/Nutzer das unentgeltliche, räumlich unbeschränkte und zeitlich auf die Dauer des Schutzrechts beschränkte einfache Recht ein, das ausgewählte Werk im Rahmen der unter
→ http://www.econstor.eu/dspace/Nutzungsbedingungen
nachzulesenden vollständigen Nutzungsbedingungen zu vervielfältigen, mit denen die Nutzerin/der Nutzer sich durch die erste Nutzung einverstanden erklärt.

Terms of use:
The ZBW grants you, the user, the non-exclusive right to use the selected work free of charge, territorially unrestricted and within the time limit of the term of the property rights according to the terms specified at
→ *http://www.econstor.eu/dspace/Nutzungsbedingungen*
By the first use of the selected work the user agrees and declares to comply with these terms of use.

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Mitglied der Leibniz-Gemeinschaft

Register Classification by Timbre

Claus Weihs[1], Christoph Reuter[2], and Uwe Ligges[1]

[1] University of Dortmund*, Department of Statistics 44221 Dortmund, Germany
[2] Musikwissenschaftliches Institut, Universität Wien, A-1090 Wien, Austria

Abstract. The aim of this analysis is the demonstration that the high and the low musical register (Soprano, Alto vs. Tenor, Bass) can be identified by timbre, i.e. after pitch information is eliminated from the spectrum. This is achieved by means of pitch free characteristics of spectral densities of voices and instruments, namely by means of masses and widths of peaks of the first 13 partials (cp. Weihs and Ligges (2003b)).

Different analyses based on the tones in the classical song "Tochter Zion" composed by G.F. Händel are presented. Results are very promising. E.g., if the characteristics are averaged over all tones, then female and male singers can be easily distinguished without any error (prediction error of 0%)! Moreover, stepwise linear discriminant analysis can be used to separate even the females together with 28 high instruments ("playing" the Alto version of the song) from the males together with 20 low instruments (playing the Bass version) with a prediction error of 4%. Also, individual tones are analysed, and the statistical results are discussed and interpreted from acoustics point of view.

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 (2003b)) 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 analysis is the demonstration that the high and the low musical register (Soprano, Alto vs. Tenor, Bass) can be identified 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. As in the voice prints introduced in Weihs and Ligges (2003b) 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

* The work of Claus Weihs and Uwe Ligges has been supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 475.

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" (see section 2). These data were analysed with Linear Discriminant Analysis (LDA) and decision trees (see section 3). The results are very promising (see section 4). Some acoustics' explanations of our findings are given in section 5.

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)) 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): *aflute-vib, bells, cello-bv, clari-bfl, clari-efl, elecguitar1, elecguitar4, enghorn, flute-flu, flute-vib, frehorn, frehorn-m, marimba, oboe, piano-ld, piano-pl, piano-sft, sax-alt, tromb-ten, trumpba, trump-c, trump-csto, vibra-bow, vibra-hm, viola-bv, viola-mv, violin-bv,*

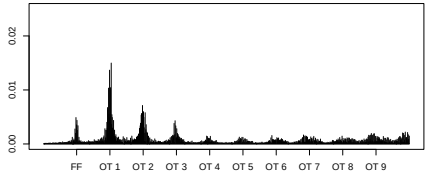

Fig. 1. Pitch independent periodogram (professional bass singer).

violin-mv.

Bass version: *bassoon, bflute-flu, bflute-vib, cello-bv, elecbass1, elecbass5, elecbass6, elecguitar1, elecguitar2, elecguitar4, 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 print characteristics are derived (cp. Weihs and Ligges (2003b)). 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 (cp. Figure 1). 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 half tone distance between the smallest and the biggest frequency of the peak with a spectral height above a pre-specified threshold. 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). Note that pitch information is eliminated in that the frequencies corresponding to fundamentals and overtones are ignored in the pitch independent periodogram. Mass is measured as a percentage (%), whereas width is measured in parts of halftones (pht). 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 halftone distance and formant intensity (cp. Weihs and Ligges (2003b)), 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 singer or instrument.

3 Classification Methods

On these data we applied supervised classification methods (see, e.g., Michie et al. (1994)) trying to reproduce the pre-defined grouping by means of classi-

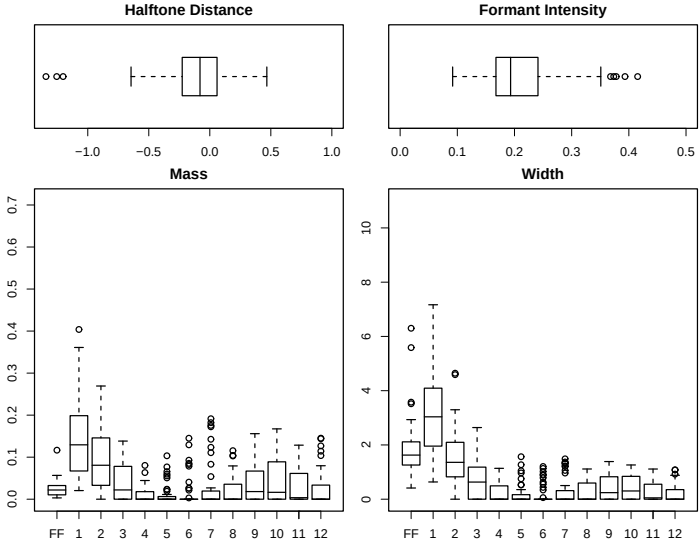

Fig. 2. Voice print of professional bass singer.

fication rules from the chosen voice print characteristics. We applied the easily interpretable classification tree (more specifically RPART by Therneau and Atkinson (1997)) and the well-known statistical linear discrimination analysis (LDA) to our data. These two classification methods are often identified to be adequate for quite different situations. For such methods the classification quality can, e.g., be measured by means of the misclassification rate, i.e. the ratio of the wrongly classified cases to the overall number of cases, which will be estimated by cross-validation.

4 Results

4.1 Individual tones, voices only

Let us start with the analysis of individual tones. If one restricts oneself to voices, then the best classification with only an error rate of 9.2% (estimated by 10-fold cross-validation) resulted from using only MassFF, MassOT01, WidthFF, WidthOT01 as predictors in LDA. The classification is detailed in Table 1. Obviously, the middle voice types Alto and Tenor generate the most errors. The results even show that the four characteristics MassFF, MassOT01, WidthFF, WidthOT01 are more appropriate for prediction of register than all 26 characteristics together (12.4 % error). Thus, there are characteristics that deliver prediction irrelevant information for the classification rule. The prediction error of 9% of the individual notes appear to be acceptable. The most important characteristics for separation of high and low

voices are MassFF and WidthFF with 8.5 % apparent error rate. However, the groups are not very well separated even for these characteristics. MassFF alone is not sufficient for prediction (21.6 % error).

In the following we will mainly concentrate on reporting of the results of `LDA(MassFF, MassOT01, WidthFF, WidthOT01)`. Other results will only be mentioned in comparison. Note, however, that decision trees were never competitive.

4.2 Individual tones, voices and instruments

Considering the voices together with the instruments, the error rate of `LDA(MassFF, MassOT01, WidthFF, WidthOT01)` is roughly doubled, namely from 9.2% to 17.1% of the individual notes (estimated by 10-fold cross-validation). The only instruments which are predominantly misclassified are bass French horn and bass-marimba with 72% and 89% error, correspondingly. Again, the characteristics MassFF and WidthFF separate high and low particularly well (20.7% apparent error rate). However, the combination MassOT01 and WidthOT01 is even somewhat better (19.9%). Separation of groups is even worse than for voices alone. MassFF alone is, again, not sufficient for prediction (38.1% prediction error). Note, however, that LDA based on all 26 characteristics leads to the distinctly best error rate (14.2%). Here only bass-marimba is particularly bad predicted.

4.3 Averaged tones, voices only

After averaging the characteristics of the individual tones, i.e. using only one value for each characteristic per voice, prediction is possible without any error (0% error estimated by 17-fold cross-validation) using the classification rule based on `LDA(MassFF, MassOT01, WidthFF, WidthOT01)`. The apparent error rate is 0% for three pairs of characteristics, namely for "MassFF, WidthFF", "MassOT01, WidthOT01", and "WidthFF, WidthOT01". Again, MassFF alone is not sufficient for prediction (error rate = 11.8%).

4.4 Averaged tones, voices and instruments

If instruments are considered also, then the error rate is only increasing to 4.6% for `LDA(MassFF, MassOT01, WidthFF, WidthOT01)` (estimated by 65-

	high	low	error
Soprano	33	3	0.083
Alto	47	7	0.130
Tenor	3	24	0.111
Bass	1	35	0.028

Table 1. Classifying individual tones of voices with `LDA(MassFF, MassOT01, WidthFF, WidthOT01)`.

fold cross-validation, i.e. by leave-one-out cross validation). Only the low instruments cannot be predicted perfectly (see Table 2). When considering all characteristics the corresponding error rate of the LDA classification rule is somewhat decreasing to 3.1%. In the case of `LDA(MassFF, MassOT01, WidthFF, WidthOT01)` only three bass-instruments are wrongly predicted as high, namely French horn (stomped and not stomped) and Marimba. Using LDA with all characteristics only Marimba and one Tenor singer was wrongly classified. The scatterplot matrix shows that the variable pair "MassOT01, WidthOT01" leads to the smallest apparent error rate (see Figure 3). Again, using only MassFF for prediction is not sufficient (41.5% error!).

5 Acoustics

Our findings are well supported by acoustics. Some explanations are the following. The relatively small opening of the human mouth acts as a high pass filter, i.e. the lower the tone the less the mass of the fundamental relative to the 1st overtone. This was already found in the middle of the last century (s. Scheminzky (1943), 428). From this it, e.g., follows that sopranos have more mass in the fundamental than basses. Moreover, synthesizing the fundamental together with a 18 dB weaker 1st overtone plus a vibrato typical for singing voices (6 Hz, 1-2% lift) leads to the impression of a soprano voice (Voigt and Reuter (1998), 18-20). Thus, the fundamental together with the 1st overtone is enough to produce voice similar tones. Overall, the fundamental and the 1st overtone appear to be important candidates for the separation of high an low register for voices.

Sopranos nearly always use head voice with strong fundamentals, basses nearly always chest voice. Altos and Tenors change between the two types of register, which leads to errors in register prediction. Therefore, overlap of registers occur for altos and tenors, and these voice cannot be attached to only one type of register in the case of individual tones.

Most music instruments are too small for a strong production of their lowest fundamentals. Thus, the fundamental has the more mass the higher

	`LDA(MassFF, MassOT01, WidthFF, WidthOT01)`			`LDA(all charact.)`		
	high	low	error	high	low	error
Soprano	4	0	0.000	4	0	0.000
Alto	6	0	0.000	6	0	0.000
A-instr.	28	0	0.000	28	0	0.000
Tenor	0	3	0.000	1	2	0.333
Bass	0	4	0.000	0	4	0.000
B-instr.	3	17	0.150	1	19	0.050

Table 2. Classification of voices and instruments based on averaged characteristics.

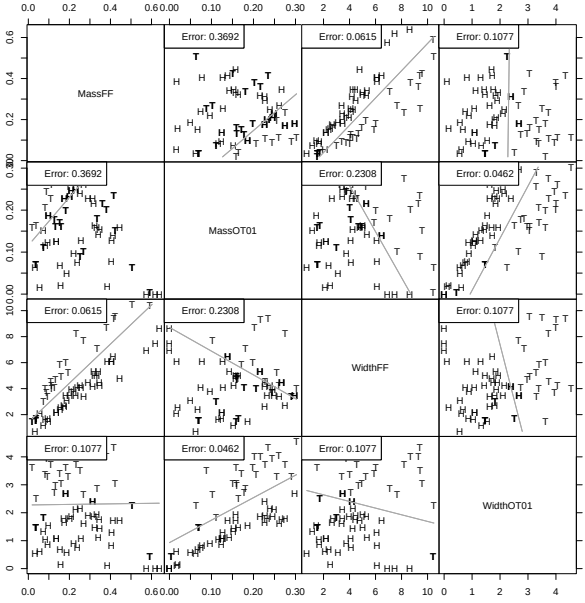

Fig. 3. Scatterplot matrix of MassFF, MassOT01, WidthFF, WidthOT01 with class separating lines and apparent error rates for voices and instruments based on averaged characteristics (H = high, T = low).

the tone is, and a strong fundamental relative to the 1st overtone indicates a high register for music instruments. The most problems occurred with French horn and Marimba. However, comparing French horn and Bassoon in their low register both instruments have similar spectral properties, e.g. a strong formant area 300-500 Hz. For both instruments the fundamental reaches the formant area with increasing pitch, however, slowly for the French horn, and abruptly for the Bassoon (Reuter (2002), 263, 327). Thus, the change between a strong fundamental and a strong 1st overtone is more exact for the Bassoon, leading to a lower error rate. For the Marimba in its low register partials are not harmonic so that the impression of the fundamental is built by a residual tone not included in the spectrum (Hall (1997), 176). This causes the problems with classification. Overall, following these arguments, except for French horn and Marimba the fundamental and the 1st overtone appear to be good indicators for register.

6 Conclusion

Altogether, the found characteristics lead to astonishingly well prediction of register. Individual tones are predicted correctly in more than 90% of the cases for the sung tones, and classification is only somewhat worse if

instruments are included in the analysis. Even better, if the characteristics are averaged over all involved tones, then voice type (high or low) can be predicted without any error, and only with at most two instruments (French horn and Marimba) severe classification problems appear, French horn not being a problem when using all characteristics for classification. Thus, there are small problems with predicting the register of individual tones, but on averages the instruments can be identified as high or low nearly without problems, with the exception of at least Marimba in its Bass version.

References

BROCKWELL, P.J. and DAVIS, R.A. (1991): *Time Series: Theory and Methods*. Springer, New York.

GÜTTNER, J. (2001): *Klassifikation von Gesangsdarbietungen*. Diploma Thesis, Fachbereich Statistik, Universität Dortmund, Germany.

HALL, D.E. (1997). *Musikalische Akustik: Ein Handbuch*. Schott, Mainz

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.

McGill University Master Samples. McGill University, Quebec, Canada. URL: `http://www.music.mcgill.ca/resources/mums/html/index.htm`

MICHIE, D., SPIEGELHALTER, D.J. and TAYLOR, C.C. (Eds.) (1994): *Machine Learning, Neural and Statistical Classfication*. Ellis Horwood, New York

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.

SCHEMINZKY, F. (1943): *Die Welt des Schalls*. Salzburg.

THERNEAU, T.M. and ATKINSON, E.J. (1997): *An Introduction to Recursive Partitioning Using the RPART Routines*. Technical Report, Mayo Foundation.

VOIGT, W. and REUTER, C. (1998): About the timbre quality in case of the Thereminvox. *Proceedings of the Russian Conference on Musicology: Organology, Petersburg, 18–20*.

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. (2003a): Automatic transcription of singing performances. *Bulletin of the International Statistical institute, 54th Session, Proceedings, Volume LX, 507–510*.

WEIHS, C. and LIGGES, U. (2003b): 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.