Sound Quality of Wind Turbines

Reinhard Weber, Christoph Eichenlaub

Physics Department, WG Acoustics/Psychoacoustics
Carl-von-Ossietzky University, D - 26111 Oldenburg

The unpleasantness of 20 s-sequences of wind turbine noises is assessed by test persons (TPs) in paired comparison experiments. Noise rankings are deduced from the TPs unpleasantness judgements as well as from the calculated parameter data and theyare analysed by several statistical methods. The unpleasantness of wind turbine noises is estimated very similar by a large group of TPs. For the objective description of the noises psychoacoustical parameters are calculated. The judgement behaviour of the TPs fits very well to the rankings which have been computed for the psychoacoustic parameters, tonality and fluctuation strength.

1 Introduction

Subjective annoyance caused by wind turbine noises shall be related to acoustical and psychoacoustical parameters which are supposed to be the objective basis for the subjective response.

In the laboratory the question has to be about the unpleasantness instaed of annoyance, because annoyance is not only due to the acoustical input but also to nonacoustical factors which are not present in the laboratory. Unpleasantness is regarded as the reaction to the acoustical noise input. In combination with a real situation this unpleasantness may be moderated by acoustical and nonacoustical influences and evoke a sensation of annoyance.

For the assessment of wind turbine noise an adequate way of noise presentation has to be checked in a first step. The extension of the acounted dynamic source "wind turbine" arises doubts whether a single channel recording where the spaciousness of the source disappears is adequate for the unpleasantness judgements. We thus decided to compare a recording technique using an artificial head to the classical technique for official noise emission sound pressure level measurements [1] using a half-inch-microphone on a groundboard in a first test.

In a subsequent experiment the assessments are carried out in a similar form, but with a new series of recordings, to check the results of the first part. Here all stimuli are artificial head recordings and the number of different sources is increased to 8 instead of 5, as in the previous experiment to cover a wide range of wind turbine noises.

2 Experimental setup

2.1 Single Channel vs. Atificial Head Recordings (5 Windturbines)

Artificial head and single microphone recordings are simultanously made at five different wind turbines and stored on Digital Audio Tapes (Sony TCD-D10 DAT, 48 kHz sampling frequency). The recording position is varied in the distance from 30 m up to 90 m and in the direction to the rotor from in-the-plane to right-angle.

From the recordings six pairs of passages of 20 s length are chosen synchronized for both techniques. The perceived unpleasantness of these 12 sounds was judged by 24 normal listeners[1]. The 24 TPs are aged between 23 and 49 (6 of them female, 18 male); 16 of the TPs have experiences in other psychoacoustical studies.

In the experiment the TPs are asked to compare the noises with regards to their unpleasantness. In every test session a TP has to compare 66 pairs, so that every noise is compared to each of the other noises: the 6 microphone noises are compaired with each other microphone noise (15 pairs) as well as the 6 artificial head noises (15 pairs) and also all microphone noises are compaired with all artificial head noises (36 pairs).

The TPs are seated in a soundproof cabin and navigate through the experiment with the help of an assessment surface on the monitor screen and the keyboard. The sounds are stored on the hard disc of a PC and are presented via Stax earphones. The microphone recordings are presented diotically. The loudness of the signals is adjusted to a calculated loudness level of 14.3 sone, according to the Zwicker model and computed by the Binaural Analysis System BAS 4.20 from Head-Acoustics. This loudness value is related to the whole sequences of 20 s.

After the test the TPs are interviewed in order to find out about their impressions, decision strategies and their coping with the assessment task. All TPs have performed the test twice or more times to check wether the decision criterion remains stable for different test sessions. An experimental session lasts from 30 minutes up to 60 minutes.

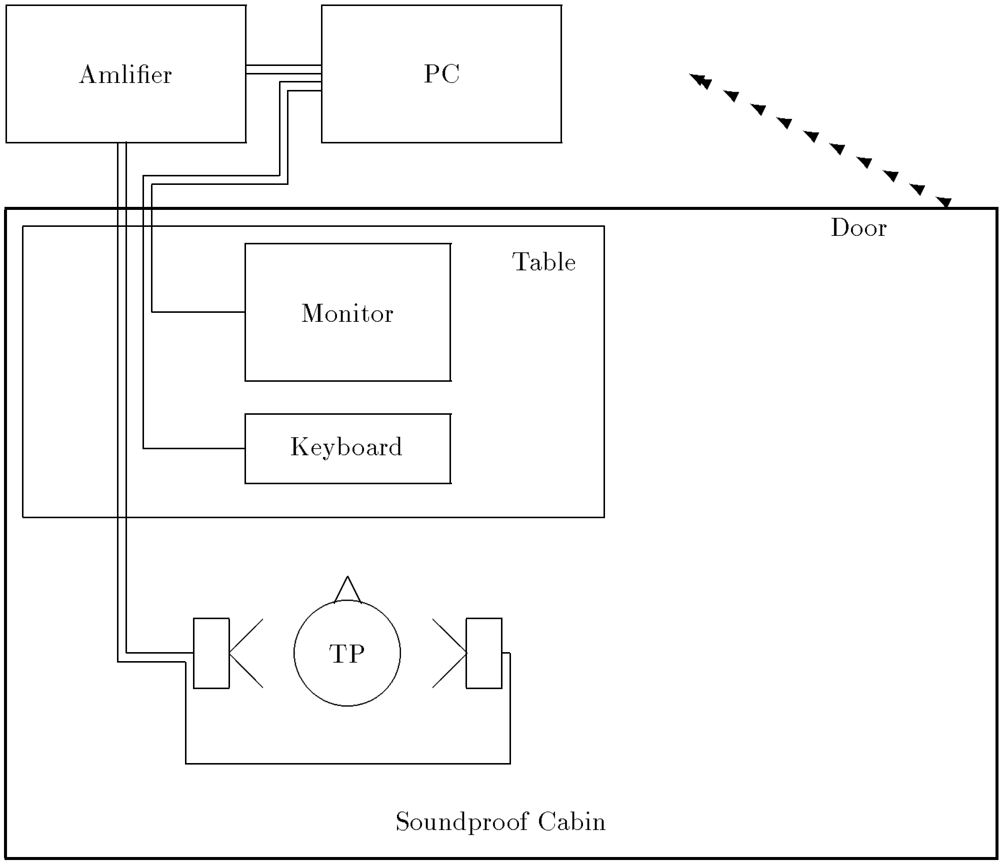

Figure 1: **Experimental Setup:** All noises are recorded to the harddisk of a PC Pentium 133 as WAV-files. The noises are DA-converted by a Turtle-Beach-Tropez-Plus-Soundcard and are presented in a soundproof cabin through STAX SR-Λ (Exp. 1) or Sennheiser HE 60 (Exp. 2) headphones via a STAX (Exp. 1) or Sennheiser HEV60 (Exp. 2) amplifier.

2.2 Artificial Head Recordings (8 Windturbines)

A new set of noises is recorded at different wind turbines at wind speeds between 5.5 and 9.5 m/s using a Head-Acoustics HMS II artificial head system with a pair of B&K 4165 1/2"-microphones and a portable Sony TCD-D10 DAT recorder at a sampling frequency of 48 kHz. To avoid disturbing wind noises the ears of the artificial head are covered with a modified[2] Head-Acoustics' windshield. The recording position is varied in the distance from 33 m to 70 m and the direction over the whole radius. From the recordings ten passages of 20 s length are chosen for the assessment tasks.

In the experiment 17 TPs estimate the unpleasantness of 10 noises using the paired comparison method. The TPs are aged between 22 and 43 (7 of them female, 10 male); 8 of them have previous experiences with other psychoacoustical tests, none of them has estimated wind turbine noises before. TPs judge 45 pairs in two test sessions in order to compare all pairs of noise twice.

The loudness of the signals is adjusted to a N_4 of 16.3 sone. The N_4 is the 4 % (upper-)

[1] TPs are accepted as normal listeners if they have no threasure suspension stronger than 15 dB at 500 Hz, 1, 2, 3, 4, 6, 8 kHz. This was tested by a Bekesy-Tracking using an B & K 1800 audiometer.

[2] Two tonal noises inducing wires were covered by foam material

percentile loudness (Zwicker method). The loudness distribution is derived again from the entire 20 s signal presented, using the BAS 4.33 Binaural Analysis System by Head-Acoustics.

In this experiment the sounds are presented using Sennheiser HE 6O headphones and a Sennheiser HEV 70 Preamplifier. Test sessions take between 8 and 20 minutes.

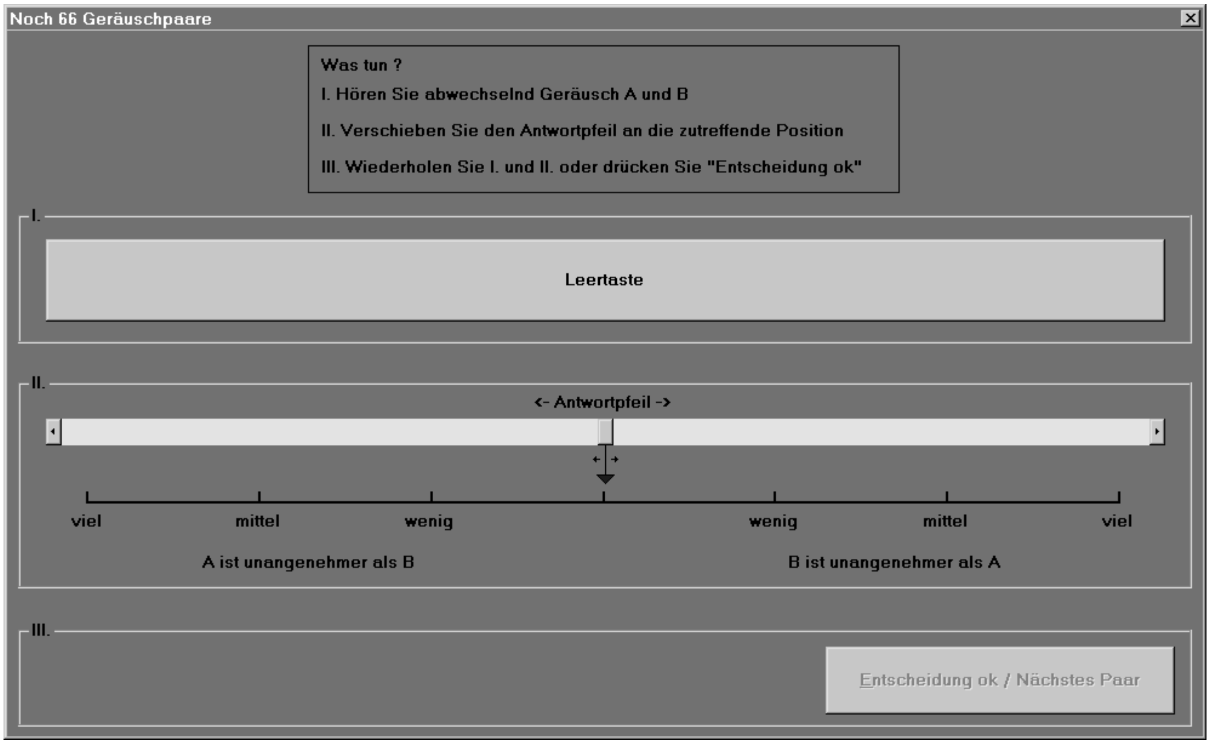

Figure 2: The assesment surface used in the experiments. The upper field contains a short instruction to the task. In field I (upper left corner) the TP is informed, which sound of one pair is presented at the moment; a hit on the space key changes to the other noise. The TPs are asked to give their judgement with the cursor-keys. A hit on the right cursor key moves the arrow in field II to the right, a hit on the left cursor key moves the arrow to the left. At the end of a comparison the TP confirms the judgement by using the return key.

3 Results

3.1 Subjective Judgements

In order to check the intra-individual reliability of the judgements the consistence of the judgements is calculated first [2]: Within the paired comparison matrix the number of circular triads in relation to the number of maximum possible circular triads is calculated[3]. Data from test sessions which show consistence levels smaller than 0.7 or correlation level between two sessions smaller than 0.75 are not considered in the exploitation.

[3]The consistence is a measure for the stability of the TPs' decision criteria. A coefficient of 1 means complete consistent, 0 means complete inconsistent.

From the judgements paired comparison matrix a ranking of the unpleasantness of the noises is calculated for each test session.

3.1.1 Single Channel vs. Atificial Head Recordings (5 Windturbines)

According to the consistence criteria the data of 18 out of 24 TPs are taken into account. Two sessions for every TP produce 36 evaluated sets of data.

A hierarchical cluster analysis of the rankings shows four different clusters. Within each cluster the TPs perform rankings which are concordant at a level of significance better than 0.005 %. For each of the four groups the rankings are calculated separately for the microphone recordings as well as for the artificial head recordings. Again the rankings are significantly concordant at a level of 0.3 % for the groups I, III, and IV for both sets of noises. For Group II the rankings for the artificial head noises are concordant at a level of 2.5 %, whereas the rankings for microphone noises are concordant at a level of 1 %.

For the groups II, III and IV the rankings of the noises are dependent on the recording technique. Group I does not show a difference.

In the interviews most of the TPs report that the test is really difficult, because unpleasantness does not depend on a single criterion, others moreover tell that it is sometimes very difficult to decide which noise is more unpleasant, either because the noises have different characteristics or sometimes because of disturbing wind noises. Some TPs also report, that they do not notice two different noises of the same wind turbine until they have to compare the microphone and artificial head recordings. Though some of the TPs do not report differences between the recording techniques there is a strong tendency to estimate the microphone recordings as more unpleasant.

As a consequence the binaural artificial head recordings prove to be better for judgements about unpleasantness under laboratory conditions than one-channel-microphone recordings. They give a more realistic impression of the acoustical real field situation. The microphone recordings give the impression of filtered an unnatural sounds; they are judged as more unpleasant than artificial head noises.

3.1.2 Artificial Head Recordings (8 Windturbines)

The 34 unpleasantness rankings (17 TPs, 2 sessions) are analysed by a factor analysis and a hierarchical cluster analysis. By three different methods[4] of cluster analysis a very stable cluster of 26 of the TPs' rankings is obtained. Using the Wards method there is a second cluster, other methods do not show this second cluster. Nevertheless in this text these 8 rankings are named as the "second group" of rankings.

In the cluster analysis the unpleasantness rankings of both sessions of 12 TPs are grouped into the same cluster by different methods as shown in the dendrogram in figure 3. A consistence check shows that the judgements of these TPs fulfill the law of transitivity very well[5]. Also the correlation between the two sessions of each TP is very high[6].

[4] Ward, Average Linkage between and within Clusters

[5] The mean coefficient of consistence of these TPs was 0.92, the lowest level was 0.75.

[6] from 0.88 to 0.99.

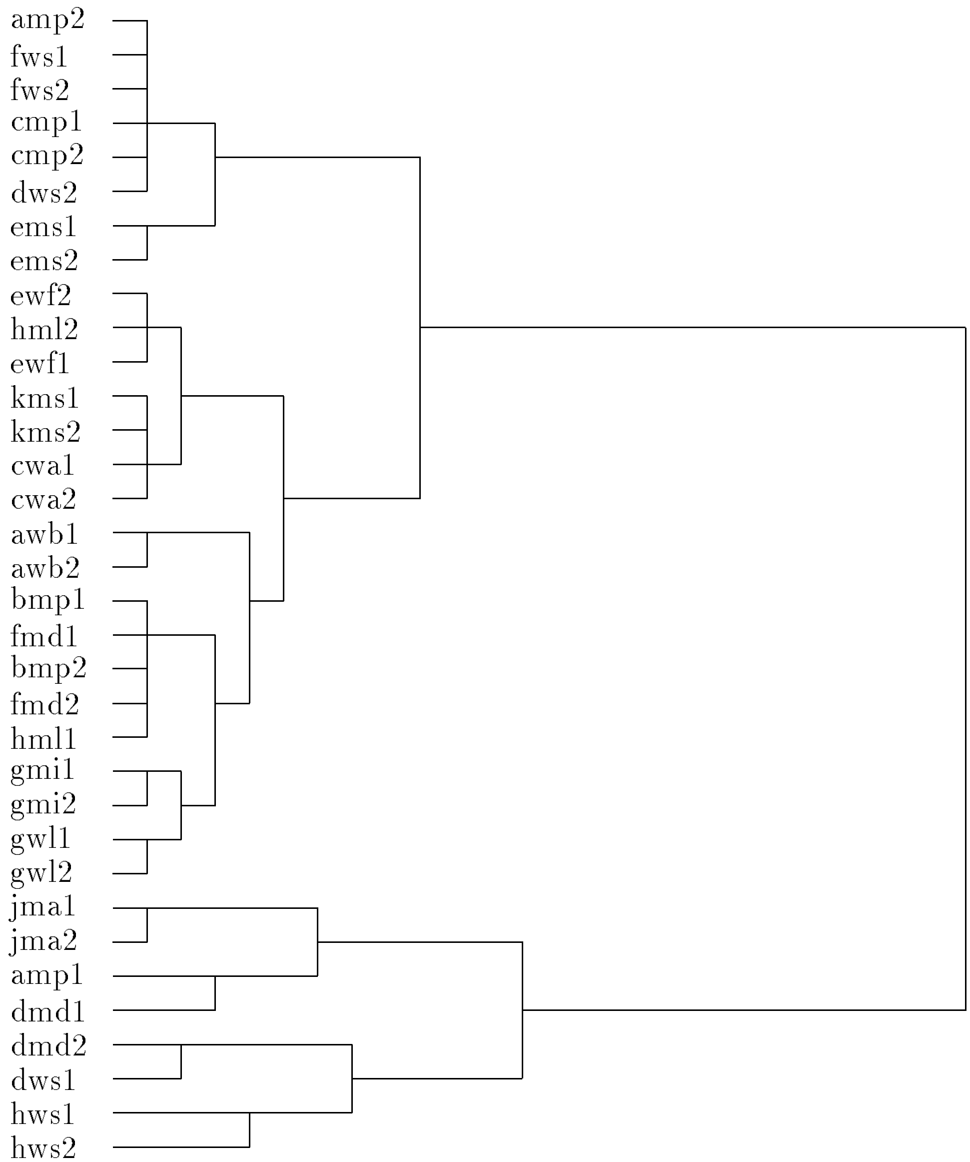

Figure 3: Dendrogram of hierarchical cluster analysis of TPs rankings using the Ward method. The distance between two rankings ist proportional to the length of the way between the two associated lines. This means from the left edge horizontal to the most far point to the right and back to the left edge again. Horizontal distances are square euclidian, vertical distances have no meaning.

The second session of two TPs also falls into this cluster though their first session fell into the second group. For both TPs the consistence of the first session is very low (0.3 and 0.45) but the second session gives complete consistent results without non-transitive judgement; the correlation between their rankings is 0.27 and 0.61. Obviously they are TPs who need training in order to judge about the unpleasantness. In further test sessions these two TPs judged in a similar way as the other 12 TPs of the first cluster.

Three TPs' rankings are in the second group both times. Only one of these persons judged consistent both times. This TP shows a correlation of 0.82 between the sessions. In an interview the TP declared noises to be a sign of life. The TP reported not to be annoyed by noises. The last two TPs show a correlation between their rankings of 0.3

and -1. The consistence improved in the second session from 0.5 to 0.7 resp. from 0.45 to 0.65. It is not clear whether these two TPs need more training than other TPs or if the test does not fit these TPs.

The factor analysis of all rankings shows that there are six factors which give explanation for 95,7 % of the variance. If the second group of TPs is removed from the analysis there are only three factors left explaining 74,9 %, 9,0 % and 5,5 % of the judgements' variance. All TPs' rankings of the first cluster were positively correlated to the first factor, for the other factors the dependency alternates.

3.2 Objective acoustical and psychoacoustical Factors

The noise sequences of the second experiment are also analysed by the Head-Acoustics BAS 4.40 Binaural Analysis System. For all sequences the L_{eq}, 4 %-, 10 %-, and 50 %- percentiles of the sound pressure level (unweighted as well as A-, B-, C-weighted), Zwicker-loudness, sharpness, tonality, roughness, fluctuation strength, degree of modulation and impulsiveness are calculated.

Rankings are deduced from these results and correlation coefficients between these rankings and the TPs' unpleasantness rankings are calculated.

It is found, that 16 rankings (out of 26) of the first cluster are correlated at rates of 0.65 to 0.95 to the computed tonality rankings at a level of significance lower than 5 % and out of these 16 there are 9 rankings on a level of significance even lower than 1 %.

Furthermore there are 8 rankings of the first cluster correlated at rates of 0.63 to 0.80 to the computed fluctuation strength rankings at a level of significance better than 5 %.

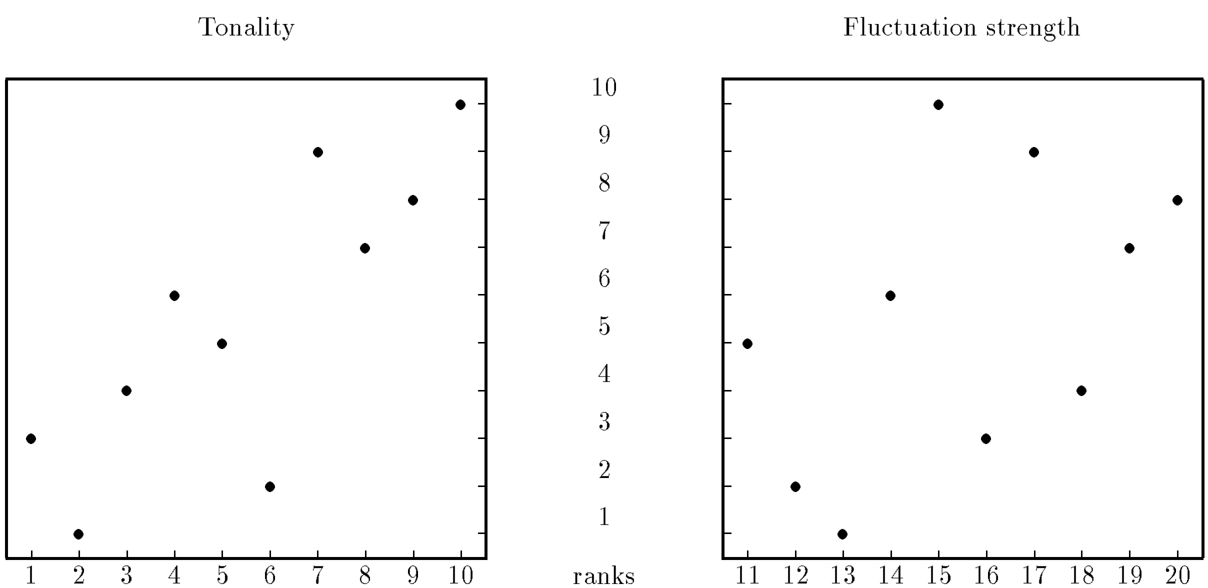

Figure 4: Rankings calculated from TPs' judgements (horizontal) vs. computed rankings (vertical) for tonality (left) and fluctuation strength (right). In the lower left corner are the lower ranks, in the upper right corner the higher ranks. The correlation of TPs' rankings and tonality rankings is 0.81, of TPs' rankings and fluctuation strength rankings is 0.46.

The TP out of the second group which has reported not to be annoyed by noises has both rankings negatively correlated to tonality and fluctuation strength on a level of significance lower than 1 %.

Figure 4 shows the dependence of tonality and fluctuation strength rankings on a ranking which is the mean of the TPs unpleasantness rankings which form cluster 1.

4 Conclusion

Comparing unpleasantness judgements of artificial head recordings to single microphone recordings of wind turbine noises in the laboratory, it turns out that the artificial head recordings should be used for this assessment task. They sound more natural in comparison to the single channel recordings, which moreover are judged to be more unpleasant than the corresponding artificial head recordings. Unpleasantness judgements of 8 wind turbines show a good agreement among a large number of TPs. An objective analysis of acoustical and psychoacoustical parameters of the wind turbine noises and their correlation with the subjective rankings shows the highest correlations for tonality and fluctuation strength.

It is unclear, whether the calculation of fluctuation strength and tonality both fit to the strong factor or whether they are correlated to a different dimension.

Acknowledgement

We wish to thank Dr. Nicolas Chouard for his most valuable comments on this paper. The work was supported by the European Union through the Joule III project Noise Imission of Wind Turbines.

Literature

[1] Recommended Practices for Wind Turbine Testing and Evaluation: 4. Acoustics Measurement of Noise Emissions from Wind Turbines; 3. Edition 1994; Edited by Sten Ljungren, Royal Institute of Technology, S-Stockholm

[2] J. Bortz, G. A. Lienert, K. Boehnke; Verteilungsfreie Methoden in der Biostatistik; Springer 1990

[3] C. Eichenlaub, R. Weber; Die Bewertung räumlich ausgedehnter Schallquellen im Labor bei unterschiedlichen Aufnahmetechniken am Beispiel von Windenergieanlagen (WEA); DAGA Kiel 1997