inter.noise 2000
The 29th International Congress and Exhibition on Noise Control Engineering
27-30 August 2000, Nice, FRANCE

I-INCE Classification: 6.6

EXTRACTING LINGUISTIC RULES TO MODEL COMMUNITY TRAFFIC NOISE ANNOYANCE

D. Botteldooren*, A. Verkeyn*, P. Lercher**

* INTEC-The Ghent University, St. Pietersnieuwstraat 41, 9000, Gent, Belgium

** University of Innsbruck, Sonnenburgstrasse 16/I, 6020, Innsbruck, Austria

Tel.: +32 9 267 35 78 / Fax: +32 9 267 35 99 / Email: botteldo@intec.rug.ac.be

Keywords:
NOISE ANNOYANCE, FUZZY LOGIC

ABSTRACT
The vagueness and uncertainty involved in community noise annoyance modeling and prediction, inspires the use of fuzzy techniques. In this paper we will focus on predicting noise annoyance using a set of linguistic rules. The fuzzy membership functions for linguistic terms such as 'high' that are used in the rules are not extracted from the database but are obtained from external sources. Among these sources are the International Annoyance Modifier Study, dosage response relations found in literature, and various expert opinion. The annoyance prediction itself is based on a set of IF-THEN rules involving the linguistic terms, fuzzy AND, and fuzzy OR. The importance of each rule in the model is tuned by comparing the outcome to the results of a social survey. The quality of a proposed decision tree is measured by the uncertainty in the prediction. This results not only in a modeling tool, but also in knowledge on the construct of community noise annoyance. The technique is applied to a detailed Austrian survey involving people exposed to road and rail noise living in one valley [1]. Several variables are included. The sound exposure related rules dominate the prediction, but also variables describing the background soundscape and the so-called enviroscape and psychscape have some impact on the performance of the model.

1 - INTRODUCTION
This paper further investigates the use of possibility theory and fuzzy logic to address the vagueness and uncertainty that is intrinsic in the relationship between sound exposure and the level of annoyance. An example of this vagueness is the meaning of linguistic terms such as "moderately", "highly", ... that are used to indicate the level of annoyance. Also many other factors that can influence annoyance are vague or uncertain by nature. By using those techniques the handling of this vagueness can be built directly into the model itself. Another advantage of fuzzy models is the way in which they make the representation of their knowledge very explicit through a set of linguistic rules.
In the first part of this paper, the fuzzy logic tools that will be used are further explored. Secondly we focus on the data from the Austrian survey that is used to test the model. In the third and fourth sections, the details of the representation of the linguistic terms and the implemented rules are fully amplified. Finally, some conclusions are drawn.

2 - FUZZY MODELING
A fuzzy model is compromised of two parts: facts and rules. Facts are of the form X = A, where X is a variable and A is a value which can be crisp or fuzzy. In possibility theory, all values are represented by possibility distributions over the domain U of the variable X. This distribution indicates the degree of possibility that a value u of the domain U belongs to A. Often, there is also a linguistic term associated with such a possibility distribution. For instance, the term "young" can be represented by a possibility distribution on the universe of age. This indicates the extent to which each age is similar to the term young. Variables for which every value is associated with a linguistic term are called linguistic variables.

The core of a model is formed by linguistic rules that express the available knowledge. A rule relates some premise to a conclusion and has the form: IF X = A THEN Y = B where X and Y are variables over the universes U and V respectively, and A and B are (linguistic terms associated with) possibility distributions over U and V. E.g.: IF (distance to road) = small THEN (road noise annoyance) = high. Rules can be seen as a possibility distribution, R(u,v), over a cartesian product of the universes. The resulting distribution can be calculated by using an implication operator. In literature different classes of fuzzy implication operators are described. Some are direct extensions of crisp implication, but the class that is most widely used in engineering applications is based on the classical conjunction operator. This model is known as the Mamdani model and is used throughout our paper [2], [3]. In fuzzy set theory, there are some choices to model such a conjunction operator. In fact, any triangular norm T can be used for that purpose. We will use the minimum norm T(x,y) = min(x,y) and the product norm T(x,y) = xy. The purpose of rules is to infer new facts. The rule IF X = A THEN Y = B allows the inference of some information about Y, given a fact about X, such as X = A'. The inferred possibility distribution of Y is calculated with the compositional rule of inference:

$$B'(v) = \sup_{u \in U} \min\left(A'(u), R(u,v)\right), \forall v \in V$$

where R is the representation of the rule. Please note that in this extension of the modus ponens from classical logic, the fact and the premise of the rule do not have to match exactly.
The real power of fuzzy reasoning lies in systems of parallel rules. In such a system, the (complex) relationship between two variables X and Y is expressed by more than one rule, for instance:

- IF X = A_1 THEN Y = B_1
- IF X = A_2 THEN Y = B_2
- ...
- IF X = A_n THEN Y = B_n

There are two different ways to infer the resulting possibility distribution on Y. First, one can use the given fact X = A' in each separate rule to infer a piece of the result with the previously described method, and then gather all those pieces with a disjunction operator. Alternatively, it is also possible to calculate the representation of each rule, R_1, R_2,..., R_n, combine them into one representation R of the whole system using a disjunction, and then use the compositional rule of inference once to infer the final conclusion. It can be proven that both possibilities always produce the same result. In a Mamdani system, the implicit "else" between the rules is thus interpreted as an OR. To model a disjunction, any triangular conorm S can be used. Here, we will use the most common one, namely the maximum conorm S(x,y) = max(x,y).
So far, only the simplest type of rule was described. It is also possible to express the confidence of a rule, or stated otherwise, the sufficiency s of having the premise true for concluding the consequent true [4]. Such a "confidence" rule, "IF X = A THEN Y = B with sufficiency = s" is equivalent to "IF X = A THEN Y = B *" with B^* defined as $B^*(v) = \max(B(v), 1-s)$, $\forall v \in V$

3 - TEST DATABASE

The constructed noise annoyance model will be tested against the results of an Austrian survey. This detailed database contains information about 2007 people that are exposed to road and railway noise, all living in one valley in Tyrol [1]. The telephone survey includes not only annoyance data, but also a lot of other psychosocial and environmental factors. Simulations of noise level for each source (L_{dn}, L_{eq}, ...) and data on air pollution and exposure to dust were added.
Noise annoyance was expressed on a four point scale labelled: "überhaupt nicht" (not at all), "teilweise" (a little), "mittelmaßig" (moderately) and "erheblich oder stark" (highly).
It should be mentioned that in this database the level of traffic noise annoyance is rather unequally distributed, in fact, very few people seem to be really annoyed: 60% of the people in the database is not at all annoyed, 20% is slightly annoyed, 10% is moderately annoyed and only 10% is highly annoyed.

4 - REPRESENTATION OF THE MODIFIERS

A very important part in the construction of a linguistic fuzzy model, is the choice of the representation of the linguistic terms that are involved. In our case, these are "überhaupt nicht", "teilweise", "mittelmaßig" and "erheblich oder stark". Assuming the modifiers were interpreted in a pure linguistic way by the

respondents, we used the data available from the International Annoyance Modifier Study [5] to model those terms as natural as possible. In that study, people were asked to indicate on a continuous numeric scale from 0 to 10 what annoyance level best corresponds to each of 21 modifiers. For the German language, the survey contains 61 records.
For each modifier, we calculated the standard deviation. Then we constructed a Gaussian distribution with this standard deviation around each individual response. Finally, the resulting functions were summed and the result was normalized [6].
First of all, the fact that "erheblich" and "stark" were considered as synonyms in the survey was confirmed: their curves were almost identical. Secondly it can be noticed that "mittelmaßig" is very sharply centered around 5 and "teilweise" is interpreted very broadly. Mainly because of this great overlap, slight changes were made to the curves to make them more appropriate for practical purposes. Also, the very small curve for "überhaupt nicht" was substituted by the fuzzy complement of "teilweise" and the "erheblich" curve was changed to a monotonic increasing sigmoid fit, based on the cumulative distribution of the scaling study. This better reflects the fact that this is the last category: the highest possible annoyance should definitely fall within this modifier. Figure 1 shows the final representations.
Furthermore, it is believed that the respondents may have decided their choice of modifier not only based on the linguistic meaning of the term, but rather on a division of the scale into four equal parts. This observation gives some rationale to our adaption of the modifiers to a more evenly distributed set.

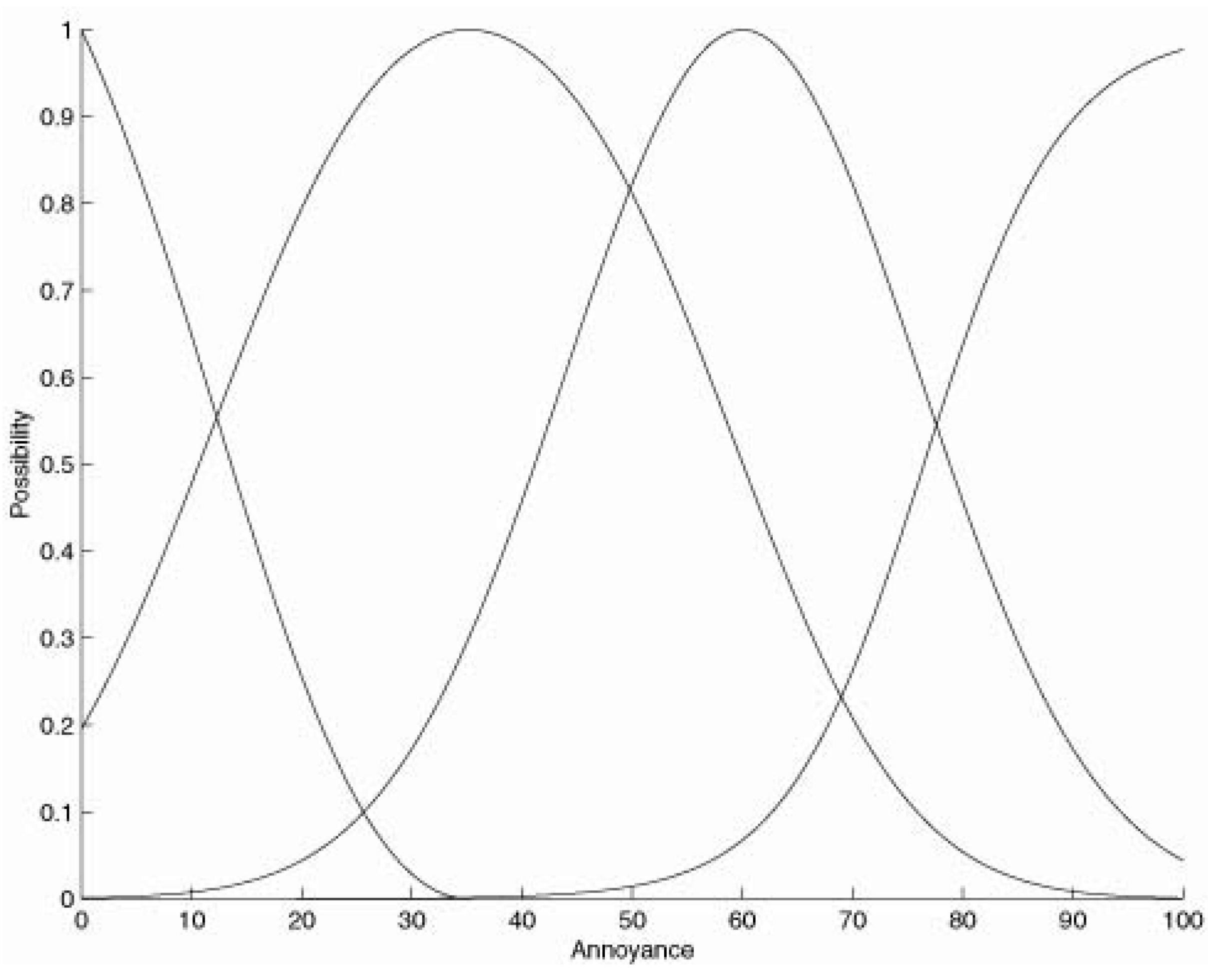

Figure 1: Final German modifiers.

5 - GENERAL STRUCTURE OF THE MODEL

In this paragraph the basic parts of the model are explained.

Starting distribution. To start with, a possibility distribution that more or less reflects the degree to which the average person in this region is annoyed, is constructed. This is based on the probability distribution of the database, and is thus slightly in favor of no annoyance at all. This initial distribution

assures that the lack of any additional information on the individual respondent, resulting in a possibility distribution of all 1's (every level of annoyance is equally possible), leads to a conclusion based on the available probabilistic information for the whole population.
Sound exposure. There has been extensive research on the extraction of dosage-annoyance relations from social surveys [7]. To include this knowledge in the rule-based system the following sound exposure rules are used:

- IF (% highly annoyed) = high THEN annoyance = high
- IF (% moderately annoyed) = high THEN annoyance = moderate

where % highly annoyed and % moderately annoyed are percentages obtained from a dosage-annoyance relation.
For the definition of "high" in the premise parts of the rules, a straight line was taken that reaches a maximum value of 1 at 7% and 20% for high and moderate annoyance respectively. The possibility distributions for "high annoyance" and "moderate annoyance" were obtained using the same procedures as for the German modifiers, this time, using the English modifier data.
Other variables. Based on expert opinion and sources from literature [8], [9], several other systems of parallel rules for those variables that possibly influence annoyance were constructed. See below. This involves the non-trivial process of defining possibility distributions for the premises part of each rule. Applying these systems of rules to a record in the database, results in a possibility distribution on the annoyance domain.
Final annoyance distribution. This process gives us some possibility distributions, all ranging over the same annoyance domain. The final resulting distribution is then obtained by combining these with an AND operator. For this conjunction operation, the product norm as described earlier was used.
Final linguistic term. The purpose of the model is to predict the annoyance modifier that is used by the respondent of the survey. To obtain this modifier, a similarity-function was calculated for the final possibility distribution against each of the four German modifiers. The one resulting in the highest similarity was chosen as the resulting linguistic term. Similarity is calculated as the overlap integral between distributions.
This model resulted in one selected modifier for each person in the database. The whole process was then put into an optimization cycle to maximize the correct hits by changing the degree of confidence of each rule. The resulting degrees of confidence indicates how important the rule is in the construct of annoyance.

6 - IMPLEMENTED RULES AND THEIR IMPACT

At the moment this paper was written, only a limited number of rules was implemented in the system. They were used to predict both road and railway noise annoyance. The confidence levels shown below were obtained when all rules were included in the optimisation.

Rules	**Confidence**
IF (% highly annoyed) = high THEN annoyance = high	0.70 (road), 0.76 (rail)
IF (% moderately annoyed) = high THEN annoyance = moderate	0.53 (road), 0.66 (rail)

Table 1: Sound exposure.

System of rules	**Confidence**
IF age = young THEN annoyance = NOT(high)	0.08 (road), 0 (rail)
IF age = old THEN annoyance = NOT(high)	0 (road), 0 (rail)
IF age = middle THEN annoyance = NOT(low)	0.25 (road), 0.17 (rail)

Table 2: Age.

System of rules	**Confidence**
IF sensitivity = stark THEN annoyance = NOT(überhaupt nicht)	0.60 (road), 0.59 (rail)
IF sensitivity = (überhaupt nicht) THEN annoyance = NOT(stark)	0.10 (road), 0.27 (rail)

Table 3: Sensitivity.

See Figure 2 for the representation of this system for road.

System of rules	Confidence
IF (distance to source) = close THEN annoyance = NOT(not at all)	0.19 (road), 0 (rail)
IF (distance to source) = far THEN annoyance = (lower half of scale)	0.21 (road), 0.07 (rail)

Table 4: Distance to nearest road, railway.

Number of children.

System of rules	Confidence
IF children = (very few or none) THEN annoyance = NOT(high)	0 (road), 0 (rail)
IF children = few THEN annoyance = (more than half of scale)	0.11 (road), 0 (rail)
IF children = many THEN annoyance = NOT(moderate OR high)	0.08 (road), 0.09 (rail)

Table 5.

Confidence in noise exposure and noise sensitivity rules is large. This is the case for both noise sources. For road traffic, the confidence in a rule based on distance is somewhat larger. We believe that this is caused by the fact that the noise simulation does not include traffic on smaller, local roads and this may influence the perception of annoyance caused by road traffic. The system replaces 'noise level' to some degree by 'distance to road' to compensate for this.

Reported sensitivity to noise is the premise in one of the most important rules. However it is not very useful in a model since it is generally not known. Moreover it is not clear to what extend the annoyance question influences the response to the sensitivity question. Therefore a model was constructed for noise sensitivity itself. It includes the same rules as above except for the noise level and distance rule. The prediction of sensitivity is worse than the prediction of annoyance. The confidence level is now quite large for the rules based on "a few children" (0.41) and "middle age" (0.43). Since this age-related rule and in particular the rule based on the number of children, are suppressed in the model for annoyance, this may indicate that these factors influence annoyance through the noise sensitivity variable.

7 - CONCLUSIONS

A fuzzy noise annoyance model based on linguistic rules is presented. At the time this paper is written a few rules were implemented and the system was tested on an annoyance database obtained in Austria. The results seem to correspond quite well to conclusions drawn from traditional models.

The approach opens however many new possibilities that could not fully be exploited yet.

ACKNOWLEDGEMENTS

The first author is a Research Associate with the Fund for Scientific Research - Flanders (Belgium). The authors would like to thank Dr. James Fields for the data of the International Scaling Study.

REFERENCES

1. **P. Lercher, G. Brauchle, U. Widmann**, The interaction of landscape and soundscape in the Alpine area of the Tyrol: an annoyance perspective, In *Proceedings of Internoise 1999, Florida (USA)*, pp. 1347-1350, 1999

2. **K. Tanaka**, *An introduction to fuzzy logic for practical applications*, Springer-Verlag, 1997

3. **R. Jager**, *Fuzzy logic in control*, Doctoral dissertation, TU Delft, 1995

4. **J. Lebailly, R. Martin-Clouaire, A. Prade**, Approximate Reasoning in Intelligent Systems, decision and control, *E. Sanchez, L.A. Zadeh, Pergamon Press*, pp. 125-144, 1987

5. **U. Felscher-Suhr, R. Guski, R. Schuemer**, Some Results of an International Scaling Study and Their Implications for Noise Research, In *In Proceedings of Noise Effects 1998*, pp. 733-736, 1998

6. **R. Cleeren, R. Vandenberghe, N. Van Gyseghem, R. De Caluwe**, The Modelling of Vague Predicates Used in Linguistic Expressions by Means of Fuzzy Set Theory, In *In Proceedings of the Fifth International Fuzzy Systems Association World Congress, Seoul (Korea)*, pp. 54-57, 1993

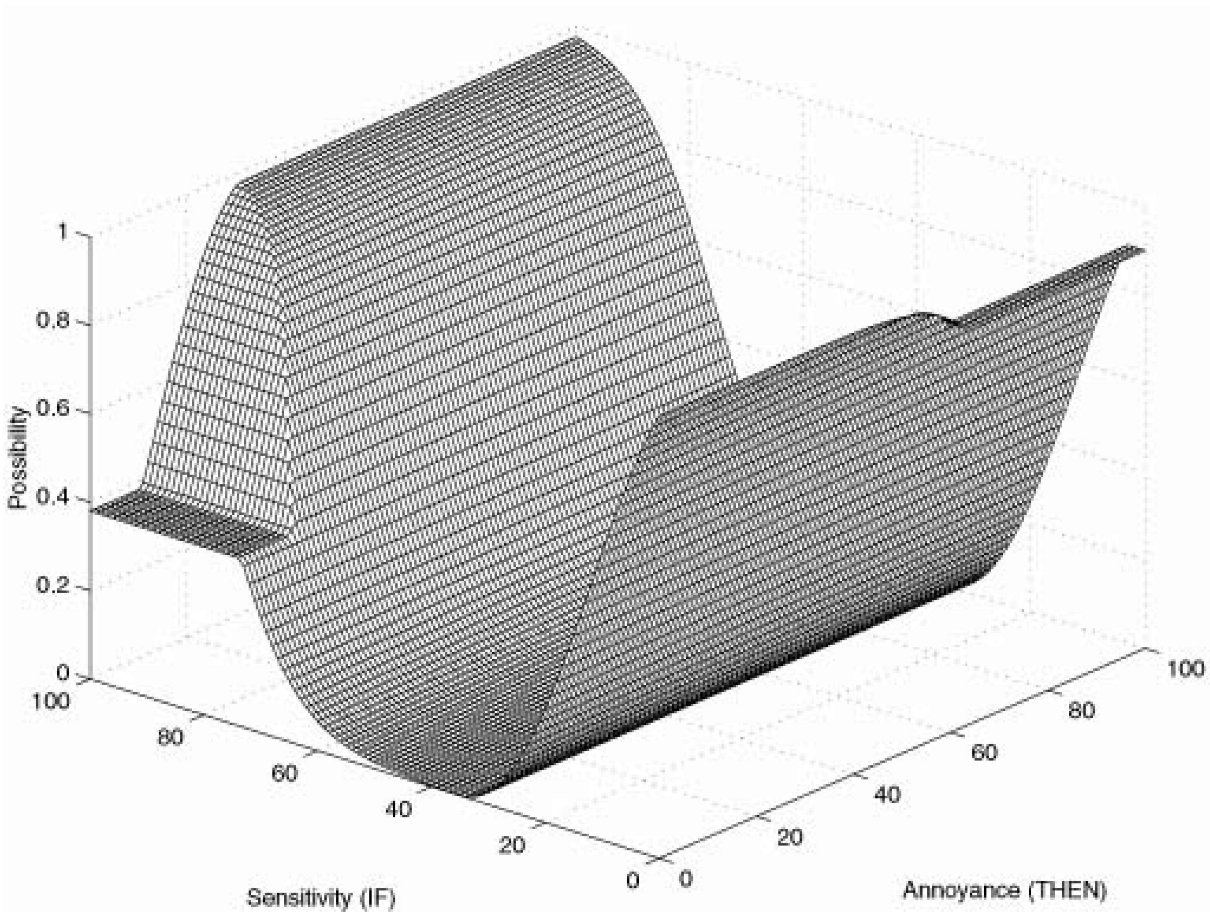

Figure 2: Sensitivity rule system (road).

7. **H.M.E. Miedema, H. Vos**, Exposure-response relationships for transportation noise, *J. Acoust. Soc. Am.*, Vol. 104 (6), pp. 3432-3445, 1998

8. **H.M.E. Miedema, H. Vos**, Demographic and attitudinal factors that modify annoyance from transportation noise, *J. Acoust. Soc. Am.*, Vol. 105 (6), pp. 3336-3344, 1999

9. **J. Fields**, Effect of personal and situational variables on noise annoyance in residential areas, *J. Acoust. Soc. Am.*, Vol. 93, pp. 2753-2763, 1993