Automatic classification of audiological expert knowledge summarized by Common Audiological Functional Parameters (CAFPAs)

Jens Schröder[1,2], Mareike Buhl[1,2], Marc René Schädler[1,2], Birger Kollmeier[1,2]
[1] *University of Oldenburg, 26111 Oldenburg, Germany,*
[2] *Cluster of Excellence 'Hearing4all', Germany, email: mareike.buhl@uni-oldenburg.de*

Introduction

The knowledge of experienced audiological experts comprises more than factual knowledge as found in audiological textbooks. In addition, all examined patients as well as communication with other practitioners contribute to building up the implicit knowledge of the ENT specialist. Particularly less experienced ENT specialists could benefit from a data-driven assistance system that represents and provides medical knowledge about a large number of patients.

Thus, we developed an automatic classification system that uses audiological data with the purpose of supporting audiological diagnostics. Data collected by an expert survey was used in this study. From this data, results for audiological tests and Common Audiological Functional Parameters (CAFPAs) were simulated to obtain a number of artificial patients. These patients were evaluated and classified using a naïve Bayes classifier (NBC) and an artificial neural network (ANN) (only for CAFPAs).

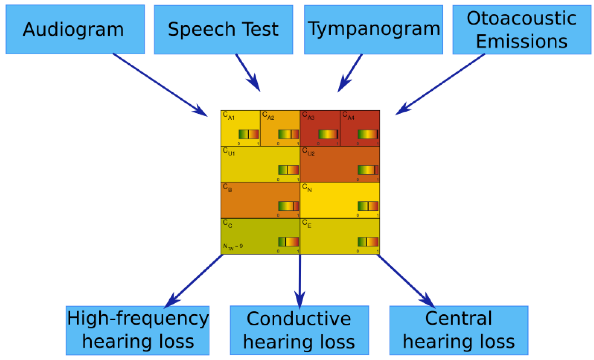

Figure 1: Schematic representation of relationships from Common Audiological Functional Parameters (CAFPAs) as intermediate layer between audiological tests and diagnostic cases.

Common Audiological Functional Parameters (CAFPAs)

The Common Audiological Functional Parameters (CAFPAs) aim at providing an abstract representation of the outcomes of audiological diagnostic tests. As intermediate layer between audiological tests and audiological findings/treatment recommendations, they perform a data reduction and serve as generalization of audiological tests (Figure 1), from which an audiological finding or diagnosis could be concluded. Ten different parameters describe relevant functional aspects of the auditory system of a patient. The CAFPAs cover hearing thresholds in different frequency ranges (C_{A1}-C_{A4}), supra-threshold deficits (C_{U1}-C_{U2}), binaural hearing (C_B), neural (C_N) and cognitive components (C_C) and the socio-economic status (C_E). Describing the hearing status of a patient by CAFPAs, the representation is designed to be independent of the choice of audiological tests, which may differ among clinics.

Database

Expert survey
In the expert survey, seven audiological findings and diagnoses, as well as seven treatment recommendations were given to the experts. Based on the given diagnostic case, the task was to indicate ranges of results that are to be expected for a set of typical audiological diagnostic tests. Secondly, the experts were asked to estimate values for Common Audiological Functional Parameters (CAFPAs), based on a traffic light scale, i.e. green = normal, yellow = uncertain, and red = impaired. Additionally, normal hearing (NH) responses for the measurements were added by taking common values for NH subjects from literature. By assigning 'green' to all Common Audiological Functional Parameters (CAFPAs), CAFPAs for NH are depicted. Due to the direction of asking in the survey, the inherent models of the experts for different audiological findings/treatment recommendations were assessed, which integrate all treated patients, respectively.

Generation of individual patients
To generate artificial individual patient data from these expert data, the indicated ranges were treated as probability distributions of the measurements. For the CAFPAs (three binary values for green, yellow and red) as well as the audiological measurements tympanometry (four types) and the occurrence of otoacoustic emissions (OAEs) (binary value) that are discrete, categorical distributions are assumed by averaging. For instance, if an expert identified the CAFPA C_{A1} for a certain diagnosis to be possibly green and yellow, the likelihood for C_{A1} for green is 0.5, for yellow 0.5 and for red 0. For continuous measurements, i.e., air conduction (AC) and bone conduction (BC) audiogram, Freiburg speech test (FS), brainstem evoked response audiometry (BERA), and adaptive categorical loudness scaling (ACALOS), one-dimensional normal distributions for each performed pa-

Table 1: Statistics of the expert data. Measurement type, corresponding parameter and its parameter values and the median, minimum and maximum number of expert data contributions per measurement type are given.

measurement type	parameter	parameter values	median num. [min, max]
Audiogram (AC)	frequency [kHz]	$\{0.25, 0.5, 1, 2, 4, 8\}$	9 [2, 11]
Audiogram (BC)	frequency [kHz]	$\{0.25, 0.5, 1, 2, 4, 8\}$	9 [2, 10]
Freiburg speech test (FS)	level [dB SPL]	$\{20, 40, \ldots, 100\}$	8 [2, 10]
BERA I,III and V	level [dB nHL]	$\{0, 10, \ldots, 120\}$	0 [0, 6]
ACALOS	level [dB HL]	$\{0, 20, \ldots, 120\}$	4 [1, 6]
Tympanogram	curve type	-	6 [2, 7]
OAE	occurrence	-	8 [2, 9]
total			2 [0, 11]

rameter value, e.g., different frequencies for audiograms, are used. The upper $b^{\mathrm{u}}_{m,i}$ and lower boundaries $b^{\mathrm{l}}_{m,i}$ indicated by expert i for each measurement/parameter value m are taken as standard deviation boundaries, i.e., for the standard deviation of the normal distribution it is

$$\sigma_{m,i} = \frac{1}{2}\left(b^{\mathrm{u}}_{m,i} - b^{\mathrm{l}}_{m,i}\right) \quad \text{with} \quad b^{\mathrm{l}}_{m,i} < b^{\mathrm{u}}_{m,i}, \tag{1}$$

whilst the center between the boundaries depicts the mean, i.e.,

$$\mu_{m,i} = b^{\mathrm{l}}_{m,i} + \sigma_{m,i}. \tag{2}$$

Figure 2 examplarily illustrates the generation of a normal distribution from audiogram ranges drawn by an expert. We could use $\mu_{m,i}$ and $\sigma_{m,i}$ to randomly draw individual patient data from independent normal distributions for each parameter value m. However, these data would be highly unrealistic since, apparently, dependencies exist between parameter values within a measurement type and probably also between different measurement types. Instead, we introduce a dependency factor α that is drawn from a normal distribution $\mathcal{N}\left(\mu, \sigma^2\right)$, i.e.,

$$\alpha \sim \mathcal{N}\left(\mu = 0, \sigma^2 = 1\right). \tag{3}$$

Hence, patient data $x^{(k)}_{m,i}$ are simulated by

$$x^{(k)}_{m,i} = \mu_{m,i} + \alpha^{(k)} \cdot \sigma_{m,i} \quad \text{with} \quad k = 1...K, \tag{4}$$

with $K = 30$ denoting the number of separate draws from expert i, i.e., the number of simulated patient sheets from one expert sheet. Here, the dependency factor $\alpha^{(k)}$ shifts all measurement values towards the same direction. For instance, the audiogram values of each frequency are affected in the same way, i.e., increasing or decreasing the hearing thresholds.

By applying this method, in total $68 \cdot 30 = 2040$ patients for audiological findings and $66 \cdot 30 = 1980$ for treatment recommendations could be simulated, respectively.

Experimental Setup

The simulated data of the audiological measurements that are described in Section "Database" are evaluated by a naïve Bayes classifier (NBC), since this is a robust approach if few training data is available and it easily allows for dealing with missing data (cf. Table 1). Hence, the probability of a diagnostic case λ for a simulated patient $\boldsymbol{x}^{(k)}_i = \left[x^{(k)}_{m=0,i}, \ldots, x^{(k)}_{m=M-1,i}\right]$ is given by

$$p\left(\lambda|\boldsymbol{x}^{(k)}_i\right) = \frac{p\left(\lambda\right)\prod_m p\left(x^{(k)}_{m,i}|\lambda\right)}{\sum_\lambda p\left(\lambda\right)\prod_m p\left(x^{(k)}_{m,i}|\lambda\right)}, \tag{5}$$

with M denoting the number of measurement parameters. Likelihoods $p\left(x^{(k)}_{m,i}|\lambda\right)$ for tympanometry type and the occurrence of OAEs are modeled by discrete likelihoods while for the continuous variables normal distributions are used, i.e.,

$$p\left(x^{(k)}_{m,i}|\lambda\right) = \frac{1}{\sqrt{2\pi\sigma^2_{m,\lambda}}} \exp\left(\frac{-\left(x^{(k)}_{m,i} - \mu_{m,\lambda}\right)^2}{2\sigma^2_{m,\lambda}}\right) \tag{6}$$

$$\forall m \in \mathcal{M}_c,$$

with $\mu_{m,\lambda}$ and $\sigma_{m,\lambda}$ denoting the mean and the standard deviation for measurement m and diagnostic case λ, respectively, and $\mathcal{M}_c$ the set of continuous measurements.

To evaluate the CAFPAs, an NBC with discrete likelihoods $p\left(x^{(k)}_{m,i}|\lambda\right)$ is applied. Additionally, for CAFPAs this approach is compared to an ANN with 1 hidden layer that is implemented using [1]. The number of units for the hidden layer is varied between 1 and 40. For training, a dropout rate of 20% is used to avoid over-fitting. Since the ANN is sensitive to initialization, tests will be performed for 10 different random seeds. Results will be given as mean and standard deviation of the seeds.

To yield reliable results from the small database, evaluation is done by a leave-one-out cross-validation, where all $K = 30$ simulated patients of one expert are used for testing while the remaining patient data is used for training.

Results

To evaluate the efficiency of the NBC, it is compared to an ANN. Since a common ANN is not able to compensate for missing data, only those patient data sheets are

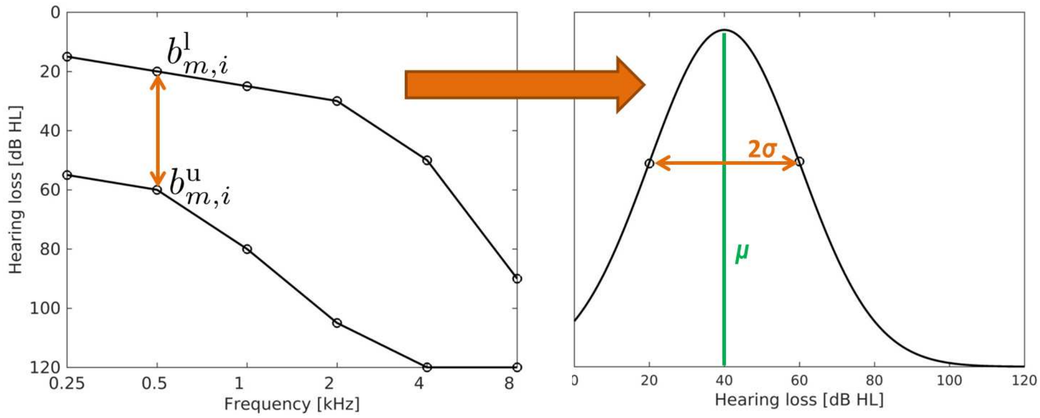

Figure 2: Illustration of the generation of normal distributions from expert distributions.

Table 2: Class-wise accuracies for the different classifier back-ends in combination with features for the audiological findings (Diag.) and treatment recommendation (Rehab.) cases.

Back-end	Features	Acc. Diag.	Acc. Rehab.
ANN	CAFPAs	(47±3)%	(37±2)%
NBC	CAFPAs	47%	40%
NBC	aud. meas.	44%	49%

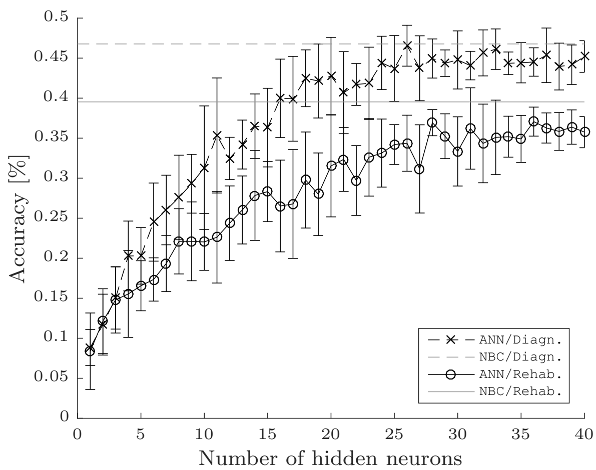

Figure 3: Class-wise accuracies (means and standard deviations of ten random seeds) for the ANN over the number of hidden neurons for CAFPAs for audiological finding/diagnosis cases (Diagn., black, dashed line), and for treatment recommendation cases (Rehab., black, solid line). The gray horizontal lines indicate the respective accuracies for the NBC.

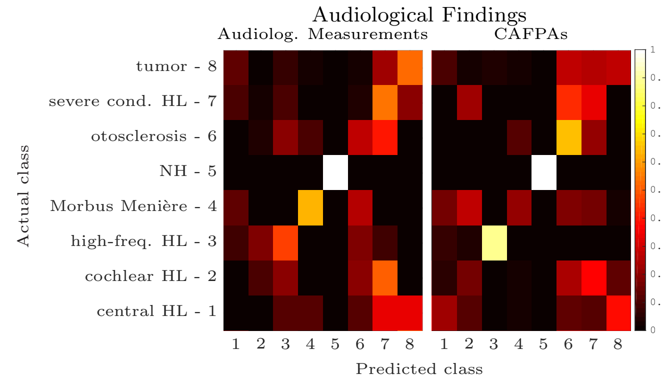

Figure 4: Confusion matrix of the eight audiological finding/diagnosis cases using audiological measurements (left) and the CAFPAs (right).

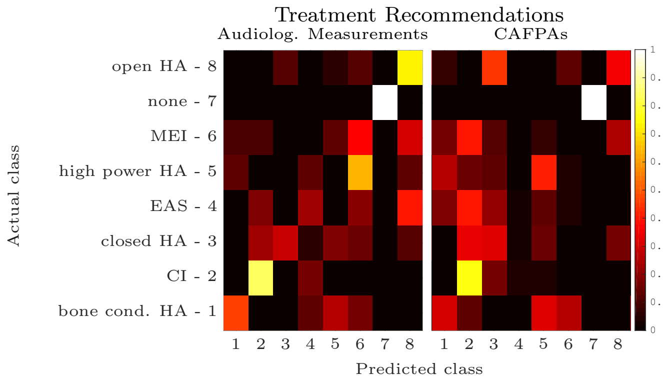

Figure 5: Confusion matrix of the eight treatment recommendation cases using audiological measurements (left) and the CAFPAs (right).

evaluated with the ANN that are complete. Hence, only $46 \cdot 30 = 1380$ of the 1980 patients are used. For CAFPAs, the means and standard deviations of the class-wise accuracies for audiological findings and treatment recommendations for ten random seeds and different numbers of hidden neurons and the respective accuracies of the NBCs are plotted in Figure 3. It can be seen that the NBCs achieve accuracies equal or above those of the best corresponding ANNs.

The adequate generalization capability of the CAFPAs is investigated by comparing the classification results of an NBC using pure audiological measurements and CAFPAs. The confusion matrices for the audiological measurements and CAFPAs are depicted for the diagnostic cases regarding audiological findings in Figure 4 and for those cases regarding treatment recommendations in Figure 5. It can be seen that the NH patients for whom no rehabilitation is necessary can be distinguished reliably from all other cases. The class-wise mean accuracy using CAFPAs is slightly better if classifying audiological findings while the audiological measurements yield higher accuracies if classifying treatment recommendations (cf. Table 2).

Conclusion

Data collected by an expert survey was used to test the generalization of audiological measurements by CAFPAs. For this purpose, ENT specialists identified common ranges of given audiological measurement types and CAFPAs for specified diagnostic cases. Artificial, individual patients were simulated by randomly drawing data points from the obtained expert distributions. The generalization capability by CAFPAs was verified using classification results of machine learning approaches. Hence, simple NBCs that can handle missing data, were compared to a widely used ANN approach when using CAFPAs. It was shown that ANNs do not improve results in comparison to NBCs for this kind of database. On the contrary, ANNs have the disadvantage of not being able to process missing data that often occur for such databases.

The decent generalization of CAFPAs was demonstrated by comparison between NBC results when using pure audiological measurements or CAFPAs, respectively. Both approaches led to similar classification accuracies for the respective tasks, hence, indicating a sufficient representation of audiological measurements by CAFPAs.

References

[1] Chollet, F. et al.: Keras. GitHub, URL: `https://github.com/fchollet/keras`

[2] Svensson, C. M., Krusekopf, S., Lücke, J., & Thilo Figge, M. (2014). Automated detection of circulating tumor cells with naive Bayesian classifiers. Cytometry Part A, 85(6), 501-511.