Appendix A – Detailed Methods

Latent Class Analysis

Latent variables are hidden and random variables that are not directly observable (nor measurable), yet could be extracted from known variables. Latent Class Analysis (LCA) is a clustering method that reveals these latent variables, also called classes. LCA has several powerful advantages over traditional cluster analysis techniques (Magdison and Vermunt 2002). For instance, it is a model-based clustering method and unlike other types of cluster analysis, LCA is not distance-based (distances between categorical data levels are often not clearly defined). Variables that are used as an input to LCA may be dichotomous, categorical, continuous or any combination. Resulting clusters (or classes) are based on membership probabilities estimated from the model, unlike the all-or-none based classification seen in traditional cluster analysis.

The basic latent class analysis (for discrete data types) is defined as follows:

$$P(\boldsymbol{y}_i) = \sum_{k=1}^{K} P(X = k) \prod_{j=1}^{J} P(y_{ij} | X = k)$$

(Eq. 1)

Here y_{ij} represents the (transformed) data of a hearing aid i on the j^{th} of J (categorical) hearing aid feature variables. LCA computes the probability of observing $\boldsymbol{y}_i$, which is denoted by $P(\boldsymbol{y}_i)$. K is the number of latent classes and k represent a specific latent class, whereas X represents the discrete latent class variable defined by k. The probability for a hearing aid feature variable being a certain level of potential (e.g. not present or a low level of potential), conditional on belonging to class k, is denoted by $P(y_{ij}|X = k)$. Lastly, $P(X = k)$ represents the (unconditional) probability of belonging to latent class k. A fundamental assumption of the LCA model is local independence, here characterized by the product over the class-specific response probabilities. However, relaxing the local independence assumption has several advantages when LCA is used as a clustering tool. A very detailed and comprehensive summary regarding the implementation of LCA and its model assumptions is given by Vermunt and Magidson (2002).

The LCA clustering method involves finding the values of the unknown parameters by maximizing a log-likelihood function. This means, maximizing some model criteria (e.g. Bayesian Information Criteria) given a specific dataset. This results in a better fitted model for the assignment of hearing aids to clusters.

$$\log L\,(\boldsymbol{\theta}; \boldsymbol{y}) = \sum_{i=1}^{N} \log P(\boldsymbol{y}_i)$$

(Eq. 2)

Here, $P(\boldsymbol{y}_i)$ represents the definition of a basic latent class analysis as shown formulated in Equation 1 and N is the total sample size. The unknown parameters are denoted by θ.

Maximization of the log-likelihood function is typically done by means of the expectation-maximization (EM) algorithm. A known problem with this method is the occurrence of a local solution, which could be prevented by using multiple starting values (Vermunt and Magidson 2002). However, changes in starting conditions do result in changes to the clusters found, but, once averaged across multiple starts, the cluster pattern is representative of these starts.

In an exploratory setting deciding on the number of classes that best fits the data is done using a fit measure obtained by the log-likelihood optimization. A typical measure is the Bayesian Information Criteria (BIC), which penalize model complexity; BIC is defined as -2log$L$ + log(N)P (P is the number of parameters). Yet, as LCA is used to model data that has a large number or records and/or many variables, it is not uncommon the end up with a large number of classes (van den Bergh, van Kollenburg et al. 2018). Such a result can therefore no longer be interpreted properly, rendering the model impractical. To overcome this problem van den Bergh, Schmittmann et al. (2017) suggested an alternative to LCA, which they called Latent Class Trees (LCT).

Latent Class Trees

LCA is a very powerful method for revealing hidden structures in complex datasets, however there are some disadvantages concerning the interpretation of LCA results. One previously stated issue, is the possibility that there is not a distinct optimum of number of classes (i.e. clusters) that fits a model. Another problem has to do with the fact that is often unclear how different model results are connected. A solution to these problems is the use of a hierarchical structure imposed on the latent classes (van den Bergh, Schmittmann et al. 2017), which is the core concept of LCT. In short, LCT is defined by a structure of mutually linked classes that are formed by sequentially splitting classes into two subclasses. This allows for a substantive interpretation of the relation(s) between classes of different levels, and so how classes are formed and related. The initial split of a LCT structure does not necessarily have to be a binary split; in their paper van den Bergh, van Kollenburg et al. (2018) argues that there are good reasons to have more than two primary classes at the root node of the tree, they also provide methods on how to determine an appropriate number of primary classes. An example of such a structure is shown in Figure A.1.

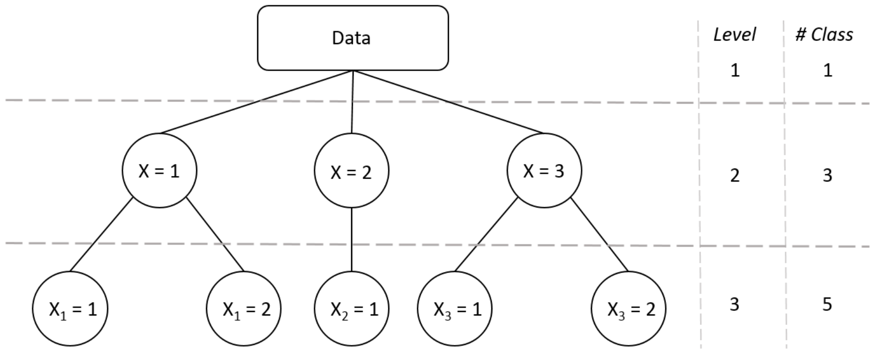

Figure A.1 – Example of a LCT structure.

When building a LCT it should be noted that LCT is based on proportional class assignment, which implies that every hearing aid is present in each node with a weight resulting from the local latent class model. Thus, the weight at a particular node level equals the weight at the parent node times the probability of belonging to the child class; the weights at the first (primary) split are equal to 1 per definition. A detailed explanation on how to perform a split at a parent node is given by Van der Palm, Van der Ark et al. (2016). The general definition of a two class latent class model in the context of LCT for a given level is

$$P(\boldsymbol{y}_i|X_{parent}) = \sum_{k=1}^{2} P(X_{child} = k|X_{parent}) \prod_{j=1}^{J} P(y_{ij}|X_{child} = k, X_{parent})$$

(Eq. 3)

Here X_{parent} denotes the parent class at the parent level L and X_{child} is the child class k (at level L + 1), where k is either 1 or 2. The probability of observing $\boldsymbol{y}_i$ is now conditional on belonging the parent class. Similarly, an adapted log-likelihood function is used to estimate the best fitted latent class model, dependent on the parent weight $w_{i,parent}$

$$\log L\left(\boldsymbol{\theta}; \boldsymbol{y}, X_{parent}\right) = \sum_{i=1}^{N} w_{i,parent} \log P(\boldsymbol{y}_i | X_{parent})$$

(Eq. 4)

Interpretation/evaluation

As the LCT method was applied in an exploratory setting, the aim was not to find the 'true' number of hearing aid modalities, but to define a set of modalities that describes the data reasonably well and, moreover, is easy to interpret. There are several statistics involved in building and evaluating a LCT structure. First, the root split needs to be evaluated to determine the number of primary classes. The authors van den Bergh, van Kollenburg et al. (2018) proposed to use a relative improvement in fit measure to decide on the number of primary classes, based on the statistical fit index. The relative improvement for a model with one split over a model with no splits is per definition 1. Thus, the benefit of subsequent splits is therefore expressed as the ratio of BIC improvement, relative to a model with one split. We included this measure in our implementation of the LCT method, which is defined as the ratio of improvement when splitting the root node into two or more splits, relative to a model with no split (one class).

Secondly, successive splits following the root node were based on a combination of statistical information and on whether a particular split was a logical result that could be explained by the context of previous divisions. Several decisive criteria were used to asses a split in terms of model fit, such as the difference between the BIC of a one and a two split model (ΔBIC = $BIC_1 - BIC_2$). In the event that ΔBIC is 0 or smaller a split is no longer justified as it does not contribute to a better model fit at that particular level. In addition, a split was also evaluated by the local improvement in model fit at a particular node relative to the improvement of the primary split; the primary split always yields the largest local improvement in model fit. The evaluation of the quality of a particular split is another important piece of information which indicates how well a latent class could be predicted given a set of hearing aid variables; in other words, how well the latent classes could be separated (Vermunt and Magidson 2016). Entropy is typically redefined as a R^2-type measure by a rescaling procedure to lie between 0 and 1, where a fraction close to one indicates a better separation of classes (Celeux and Soromenho 1996, van den Bergh, Schmittmann et al. 2017).

To consider the quality of a specific branch of consecutive splits we evaluated the product of $R^2_{Entropy}$ results at each node. The model measures, total entropy and relative BIC, were used to examine consecutive model results. However, for practical reasons it was decided to use the size of the terminal node as the main stopping criteria. Using (fixed) group sizes was suggested by van den Bergh, Schmittmann et al. (2017) and further investigated by Pelaez, Levine et al. (2019). Nasserinejad, van Rosmalen et al. (2017) proposed that the termination of a model should depend on a minimum group size that must be greater than a predetermined cut-off percentage of the total group size, ranging between 2% and 5%. We adopted this rule and set the cut-off percentage at 5%, hence the minimum group size of a LCA class should not exceed 5% of the input data, otherwise the LCA model would be discarded. Additionally, a maximum depth size was used to prevent splits based on increasingly minor details. The maximum depth size of the LCT was set at 4 levels. Beyond this point we stopped the LCT, regardless of all other measures.

Lastly, for visual inspection of the different levels of the hierarchical LCT structure, hearing aid feature profiles plots were created for each node; the number of profiles per node depending on the appropriateness of a split. Figure A.2 shows an example of such a profile plot; in this example rescaled mean values for several hearing aid variables were shown. Mean Hearing aid feature data was rescaled between 0 and 1 to enable a straightforward comparison between different scaled variables. Moreover, the final hearing aid modalities were evaluated and labeled using such feature profile plots.

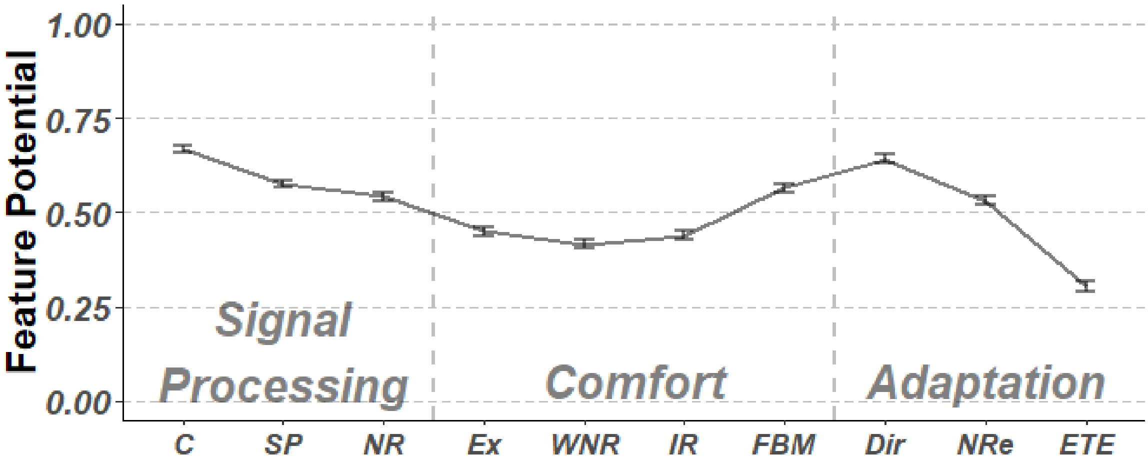

Figure A.2 – Profiles plot total dataset. Mean Hearing aid feature data were rescaled between 0 and 1 to enable a straightforward comparison between different scaled variables. Features were ordered according the three domains: Signal Processing, Comfort, and Adaptation. Whiskers show a 95% confidence interval for the specific feature. The lines between the points do not refer to a dependency between adjacent features, but were included to interpret and compare the feature profiles between modalities. Labels *x*-axis: Compression (**C**); Sound Processing (**SP**); Noise Reduction (**NR**); Expansion (**Ex**); Wind Noise Reduction (**WNR**); Impulse (Noise) Reduction (**IR**); Active Feedback Management (**FBM**); Directionality (**Dir**); Noise Reduction Environments (**NRe**); Ear-to-Ear Communication (**ETE**).

References

Celeux, G. and G. Soromenho (1996). "An entropy criterion for assessing the number of clusters in a mixture model." Journal of Classification **13**(2): 195-212.

Magdison, J. and J. K. Vermunt (2002). "A Nontechnical Introduction to Latent Class Models." White Paper, Tilburg: Statistical Innovations.

Nasserinejad, K., et al. (2017). "Comparison of criteria for choosing the number of classes in Bayesian finite mixture models." PLoS One **12**(1): e0168838.

Pelaez, K., et al. (2019). "Using a Latent Class Forest to Identify At-Risk Students in Higher Education." JEDM| Journal of Educational Data Mining **11**(1): 18-46.

van den Bergh, M., et al. (2017). "Building latent class trees, with an application to a study of social capital." Methodology: European Journal of Research Methods for the Behavioral and Social Sciences **13**(S1): 13.

van den Bergh, M., et al. (2018). "Deciding on the starting number of classes of a latent class tree." Sociological Methodology: 0081175018780170.

Van der Palm, D. W., et al. (2016). "Divisive latent class modeling as a density estimation method for categorical data." Journal of Classification **33**(1): 52-72.

Vermunt, J. and J. Magidson (2016). "Technical guide for Latent GOLD 5.1: basic, advanced, and syntax. ." Belmont, MA: Statistical Innovations Inc.

Vermunt, J. K. and J. Magidson (2002). "Latent class cluster analysis." Applied latent class analysis **11**: 89-106.

Appendix B – Detailed Results

Complete list of database attributes

Database attributes			
Name	Type	Name	Type
(1) Unique ID	ID	(26) Binaural beamforming	Y/N
(2) Brand	ID	(27) Noise reduction levels	Number
(3) Manufacture ID	ID	(28) Noise reduction environments	Number
(4) Model name	ID	(29) Wind Noise reduction levels	Number
(5) Introduction	Date	(30) Passive Feedback levels	Number
(6) Availability	Status	(31) Active Feedback levels	Number
(7) Style	Type	(32) Expansion levels	Number
(8) Ear-coupling	Type	(33) Impulse Reduction levels	Number
(9) Water resistance	IP-Class	(34) Tele-coil	Y/N
(10) CROS	Y/N	(35) Audio-input DAI	Y/N
(11) OSPL 90 (Max Output)	dB	(36) Wireless audio input	Y/N
(12) OSPL 50 (Max Gain)	dB	(37) Wireless TV input	Y/N
(13) OSPL 60 (Max Reference)	dB	(38) Wireless phone input	Y/N
(14) Lower limit bandwidth	Hz	(39) Wireless mic input	Y/N
(15) Upper limit bandwidth	Hz	(40) Ear to Ear communication	Y/N
(16) Volume control	Y/N	(41) Ear to Ear feature synchronization	Y/N
(17) Program slots	Number	(42) Ear to Ear sound streaming	Y/N
(18) Remote control	Y/N	(43) Adaptation manager	Y/N
(19) Adjustable compression channels	Number	(44) Automatic learning	Y/N
(20) Adjustable MPO channels	Number	(45) Data-logging	Y/N
(21) Signal Processing channels	Number	(46) Frequency lowering levels	Number
(22) Directionality	Y/N	(47) Tinnitus Noiser	Y/N
(23) Automatic directionality	Y/N	(48) Environmental steering	Y/N
(24) Adaptive directionality levels	Number	(49) Environmental Steering levels	Number
(25) Natural ‘ear like’ directionality	Y/N	(50) Environmental Steering manual adjustable	Y/N

Table B.1 – All ‘*ZN-hoortoestellen database*’ attributes, including the type of variable.

LCT Relative Improvement Primary Node

n primary node split	**2**	**3**	**4**	**5**
BTE	1.000	0.359	0.178	0.272
ITE	1.000	0.411	0.401	0.297

Table B.2 – Relative improvement ratio for the primary node, for the BTE and ITE data. The first split (n=2) always yields the largest improvement in model fit. Additional splits were computed as the ratio of the local improvement in model fit relative to the improvement of the first split.