

A word elicitation study including the development of scales characterizing aided listening experience

DORTE HAMMERSHØI^{1,*}, ANNE WOLFF², LYKKE J. ANDERSEN³, RIKKE L. MORTENSEN³, MADIS D. NIELSEN³, AND STEFANIE A. S. LARSEN³

¹ *Department of Electronic Systems, Signals and Information Processing, Aalborg University, Aalborg, Denmark*

² *Department of Otolaryngology, Head and Neck Surgery and Audiology, Aalborg University Hospital, Aalborg, Denmark*

³ *Master program in Engineering Psychology, E-Study Board, Aalborg University, Denmark*

The purpose of the present study was to identify the terms hearing aid professionals and their patients use in the communication about the aided listening experience and develop scales that would help characterize this experience in the domain of corrective actions that a hearing care professional may apply. The study comprised a word elicitation task based on observations and interviews from consultations at the Aalborg University Hospital. The results were analyzed by developing an affinity diagram. The resulting 80 words were then sorted by three hearing professionals in a supervised card sorting session. The resulting attributes were included in a 63-point scale design, which (in a usability test including eight hearing-aid users) were considered easy to survey and use, but also including some redundancy and ambiguities. The results suggest that it is possible to develop scales based on the voluntary statements expressed during actual consultations, but it remains uncertain whether the expressions will be interpreted the same way by other patients and professionals.

INTRODUCTION

Literature suggests (see McCormack and Fortnum (2013) for a review) that many patients fitted with hearing aids don't use them and that it is a complicated process to find the reason for the lack of success. A letter from a frustrated hearing aid (HA) user explains it well (Fig. 1). The user has congenital hearing impairment and now has a cochlear implant (CI) in the left ear after many years of HA experience. A successful fitting experience for her includes good communication with the hearing care professional (HCP), which among other things entails that you understand what each other is saying, and have the same vocabulary for sound and hearing. The user has also observed the frustration among new users, who often find it difficult to describe their experience to the HCP.

*Corresponding author: dh@es.aau.dk

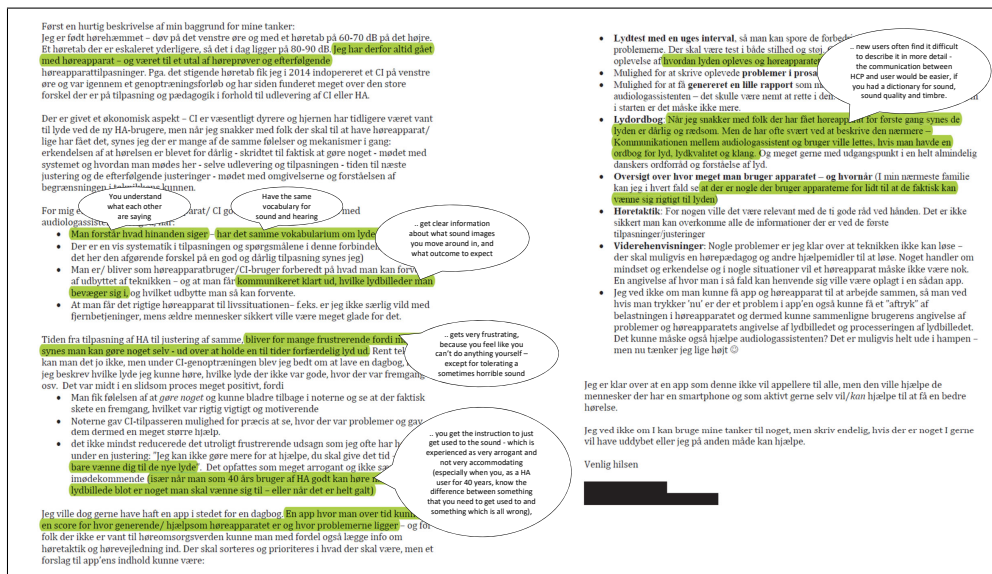


Fig. 1: Letter from frustrated HA user.

The scope of the study was to take a novel approach to 1) identifying the terms HCPs and their patients use in the communication about the aided listening experience, and 2) to develop scales that would help characterize this experience in the domain of corrective actions a HCP may apply.

METHODS

The overall method comprised of two steps, where the first was intended to elicit the words HA users and the HCPs use in the communication. The second step included the development of scales based on this, and a test trial that would provide qualitative experience with the use of such customized scales.

Word Elicitation

Word elicitation as understood by, for example, Francombe *et al.* (2014) was carried out to find suitable attributes for describing the aided listening experience for HA users. The process is depicted in Fig. 2 and includes observations during eight examinations at the audiological department of Aalborg University Hospital, involving fifteen patients, seven audiologists, and two medical doctors. The observations were succeeded by interviews for four HA users, and six interviews with two audiology assistants, which evoked further verbalization of relevant terms for the aided listening experience.

The central words that patients and professionals used for describing the aided listening experience were itemized, noted on cards (394 in total), and analyzed by

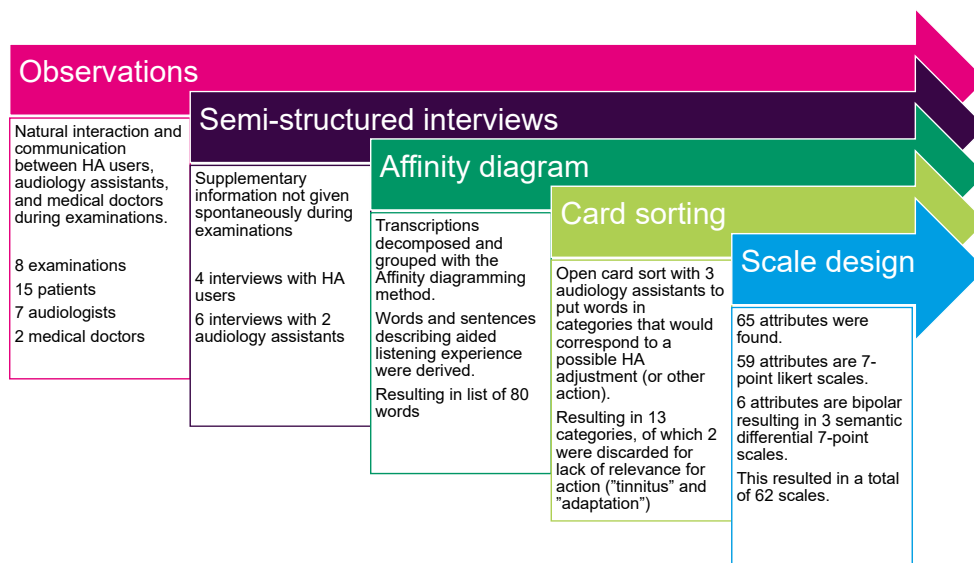


Fig. 2: Process used for word elicitation.

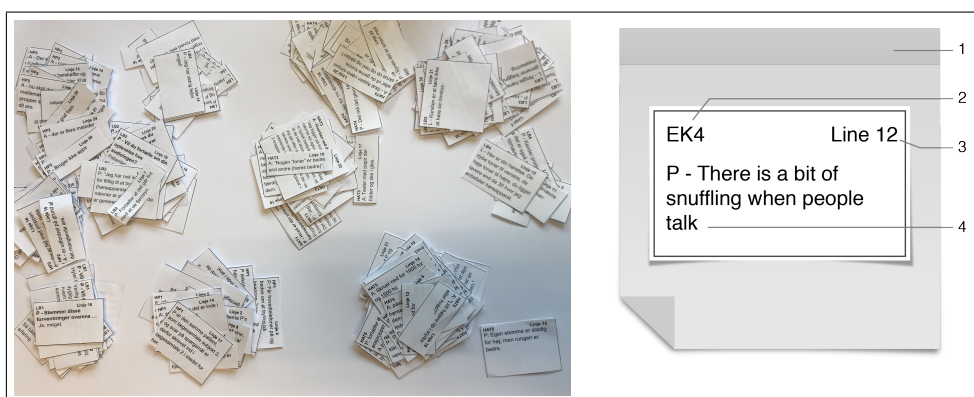


Fig. 3: **Left:** Notes decomposed from transcripts. **Right:** Example note card for the Affinity diagram: 1) post-it note, 2) label showing to which examination the quote was observed, i.e. “EK” represents follow-up (“afterkontrol”), 3) label showing the line in the transcription, where the quotation was found, 4) the quotation.

developing an Affinity diagram using inspiration from Goodwin (2009), Kuniavsky *et al.* (2013), and Dan and Siang (2018), see Fig. 3. These notes were sorted by the four junior scientists into 80 relevant words, which were eventually grouped by two HCPs in a supervised open card sorting, according to Albert *et al.* (2013) for example, in groups that were operational in terms of what actions the HCPs would and could perform during consultations. This resulted in 13 categories, of which two were discarded for lack of direct relevance to relevant actions (“tinnitus” and “adaptation”).

	strongly disagree						strongly agree
37) I experience problems with female voices	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
38) I experience problems with bright tones	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Fig. 4: Example of scale design. An index number is printed to the left, and then the item. The 7-point Likert scale ranging from “strongly disagree” to “strongly agree” is printed to the right.

Scale design and test

All 65 attributes identified in the Affinity diagram were included in a 63-point scale design (including three semantic differential scales). Fig. 4 shows an example of the scale design.

The scales were evaluated in a test with the eight HA users and two HCPs (audiology assistants). The HA users completed the scales, under the instructions following a think-aloud protocol, Mathison (2005). The results were subsequently shown to the HCPs, who were also instructed to follow the think-aloud protocol. Both HA users and HCPs participated in semi-structured interviews after their sessions.

RESULTS

The pooled results for the eight HA users are shown in Fig. 5. Most HA users were positive about using the scales as a basis of the dialogue in the context of examinations. Examples of positive statements are “*easy to survey and use*”, “*quick to fill out*” and “*the check-boxes are good so no self written text is needed*”. Six HA users stated that many of the scales were similar, but five of the six subjects also commented that this wasn’t an issue and stated that “*it made sense to ask about the aspects from many different angles*” or “*to assess the validity of the answers*”. One HA user had problems understanding the scales and didn’t complete the questionnaire. This HA user stated that binary yes/no questions would make more sense. One HCP also stated that for some of the questions, this would make more sense (i.e., either-or questions).

A qualitative examination of the response distribution reveals that items 1-4 for *Loudness*, 18-20 for *Occlusion*, 32-34 for *Low frequencies* are very similar in their score distribution, and might prove redundant by further examination. These items are also grouped in the same category from the HCP card sorting, whereas other items do not compare well within the assigned category. This is the case, for example, for items 52-56 for *2 kHz*, where the scores vary considerably between items. Since the category represents a given action, which an HCP would exert (e.g., adjustment of the amplification at 2 kHz), it could also suggest that similar actions are made in response to very different patient experiences, and if so, carry options for misunderstandings. This suggests that the grouping of the attributes may not adequately represent relevant actions of adjustments, or that the items do not adequately represent the patient

Word elicitation for scales of aided listening

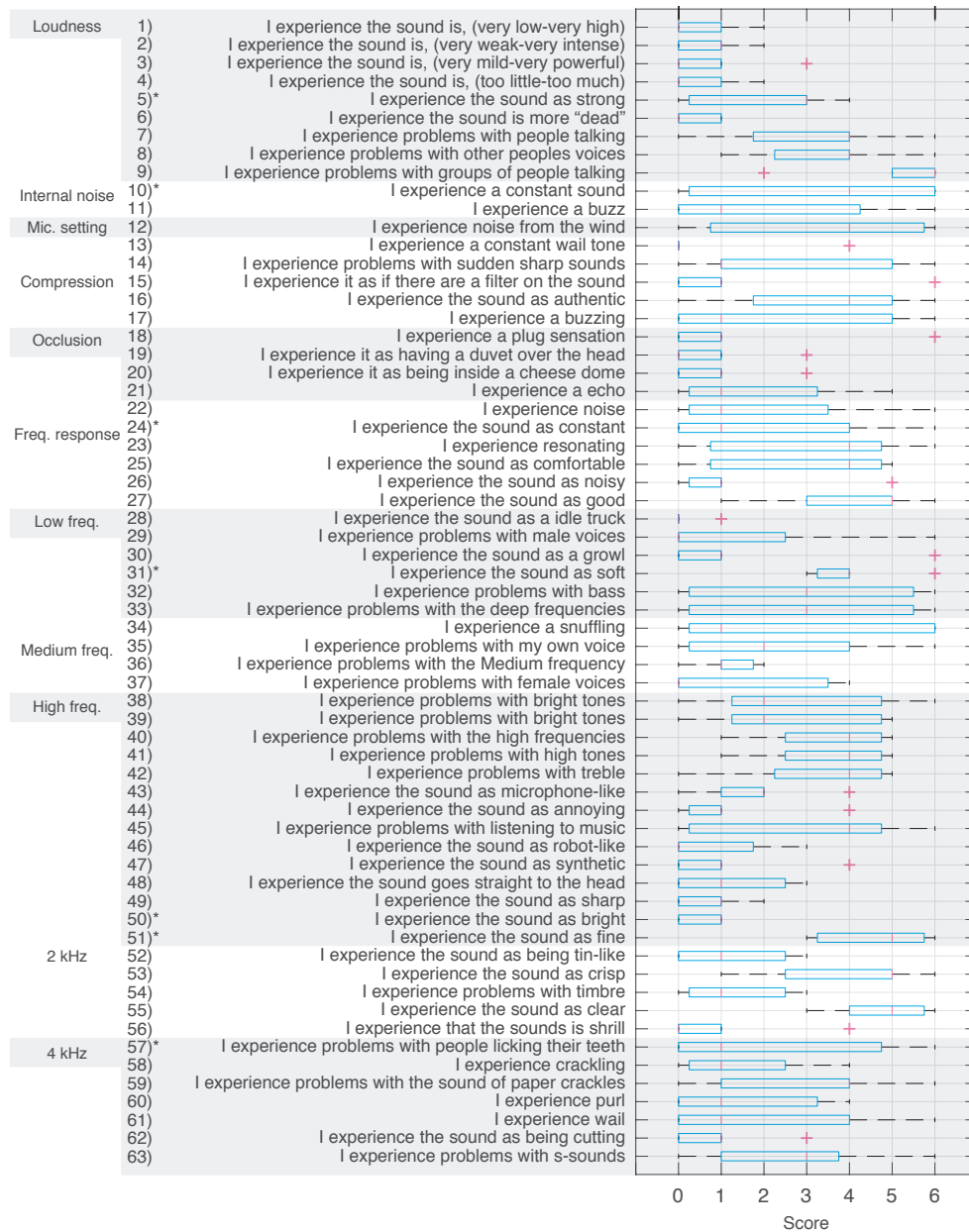


Fig. 5: Data from 8 HA users (numerical scores assigned corresponding to the categories in the 7-point Likert scale, hence “0” corresponds to “strongly disagree”, and “6” to “strongly agree”, except for the top four items, which had the end labels listed in brackets after the item). The categories (our translation) to the left are the result of the card sort. The items (our translation) are shown on the y-axis and the score on x-axis. The items that were found ambiguous or lacking in context in the interview feedback are marked with *.

experience that should result in the given action.

Items 16, 25, 27, 53 and 55 are positive aided listening experiences as opposed to the other, negative aided listening experiences. This means that the positive scores are represented to the right (agreeing to the statement), whereas all other items have positive scores at the left (disagreeing with the statement). Although all scales were presented with appropriate markers for the patient, it can, however, not be ruled out that this may have contributed to the variation between items within the 2 kHz category.

Items 5, 10, 24, 31, 50 and 51 were found by both HA users and HCPs to be ambiguous or lacking adequate context. This suggests that such statements may only be relevant in particular contexts, and the situation where the HA users responded to the questionnaires did not provide the environment for such assessment. It is possible that establishing a listening scenario, where the patients are stimulated with relevant listening experiences, will improve the reliability of such assessments. If the questionnaires should be filled out in out-of-clinic situations (e.g., prior to a clinical visit), it would seem necessary to develop instructions to accompany some of the items, as they would otherwise also lack context in the out-of-clinic environment.

DISCUSSION

The HCPs pointed out that the HA users would not complete a scale questionnaire with 63 scales. None of the HA users expressed that the amount of scales was too large, only that some of them were similar and stated that “*it is something you fill out believing that it is important for the treatment*”. It is possible that the HA users experience the completion of large questionnaires as satisfying, as long as the activity makes sense to them. In the present study, they completed the questionnaires in the clinic, and the activity provided extra attention.

If the scales are investigated for redundancy this could decrease the sense of similarities and amount of scales. On the other hand, an advantage of the redundancy is the possibility to investigate the validity of the answers given. Two items were by mistake identical (items 38 and 39), which was pointed out to the HA users. Yet the answers differ for one patient. Since scales are inherently only relevant for assessment of perceptive dimensions, which by nature can have a range of options, there will be “noise” in the responses. A test-retest could probe the magnitude of this uncertainty and reveal if some of the dimensions investigated (certain items or categories) are more prone to intrinsic variance. This was not done in the present study.

Another benefit from including all attributes is a stronger probability of including attributes HA users can relate to, as they use a richer vocabulary to describe their experiences (Tab. 1). Further studies would be required to better understand which of the scales best characterize the variance of the aided listening experience, which links efficiently to the actions possible by the HCPs. The observed vocabulary of the HCPs (the audiology assistants in particular), may constitute a desired refinement

Audiology assistant		Patient		Patient (cont.)	
Bas/base	4	Autentic sound	1	Idle truck	1
Bas amplification	1	Better	4	Lav	1
Bas sound	1	Softer	1	Male voices	1
Pleasant	1	Humming	3	More dead	1
High tones	2	Female voices	2	Like a microphone	1
Bright tones	2	Pillow over head	1	Mild	1
Echo	2	Own voice	4	Music	1
Pork roast with crisp skin	2	Some adults talking	1	(see talking)	
Frequency	1	Filter	2	Paper scratching	1
High frequencies	1	Fine	1	Clotted sensation	3
High tones	1	People talking	2	Radiator turned on	1
Enclosed	1	Powerful	1	Licking teeth	1
Bright	1	Cork	1	Talk	3
Purling	2	Stab	1	Snuffling	2
Robotic	1	Straight to the head	1	Som en vindmølle	1
S'sounds	1	Said in church	1	As a filter (see filter)	
Spoon in glas	1	Bottom	1	Voices	2
Scratching	3	Inside head	1	Strong	1
Scratching	1	Annoying	1	Noise from wind	1
Sharp	1	Smooth	1	Weak	1
Strength	1	(see female voices)		Syntetisk	1
Woom woom	1	Clearer	1	Heavy in head	1
Medical doctor		Click in ear	1	Clear	2
Bas	2	Constant howling	1	Dense filter	1
Deep tones	1	Constant sound	1	Intense	2
High tones	1	Noise	1	be in cheese dome	1
Light tones	1				
Elevate the sound	1				
Mid tones	2				

Table 1: Prevalence of central words used in the description of the aided listening experience during consultation, HA-fitting and follow-up.

guided by previous experience relating to which terms most efficiently excite which reflections in the patient. One phrase “pork roast with crisp skin” (“flæskesteg med sprøde svær”) was, for example, deliberately used in the interaction as an instrument to trigger a listening experience, which was considered to reveal sub-optimal fitting, because of a high density of transient and unvoiced consonant combinations. Also the “woom woom” was a self-engineered stimulus to bring attention to the specifics of the qualities of the listening experience at low frequencies.

CONCLUSION

The results suggest that it is possible to develop scales based on the voluntary statements expressed during actual consultations, but that the expressions may not be interpreted the same way by other patients and professionals. The results also suggest that standardized scales (e.g., MUSHRA by ITU-R (2015)) may be interpreted differently by different users.

ACKNOWLEDGEMENTS

Sincere thanks to the HA users and the staff at the audiological department at Aalborg University Hospital for their participation in experiments and interviews. Collaboration and support by Innovation Fund Denmark (Grand Solutions 5164-00011B), Oticon, GN Hearing, Widex Sivantos Audiology and other partners (University of Southern Denmark, Aalborg University, the Technical University of Denmark, Force, and Aalborg, Odense and Copenhagen University Hospitals) is sincerely acknowledged.

REFERENCES

- Albert, W., and Tullis, T. (2013). “Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics,” Ch. 6: Self-reported Metrics; Ch. 7: Behavioral and physiological metrics; Ch. 9: Special Topics in *Interactive Technologies Ser. Elsevier Science & Technology*, second edition.
- Dam, R. and Siang, T. (2018). “Affinity diagrams. Learn how to cluster and bundle ideas and facts” online at *Interaction Design Foundation*.
- Francombe, J., Mason, R., Dewhurst, M., and Bech, S. (2010). “Elicitation of attributes for the evaluation of audio-on-audio interference,” *J. Acoust. Soc. Am.*, **136**(5), 2630-2641.
- Goodwin, K. (2009). “Designing for the Digital Age: How to Create Human-Centered Products and Services,” Ch. 10: Making Sense of Your Data: Modeling. Wiley Publishing Inc.
- ITU-R (2015). “RECOMMENDATION ITU-R BS.1534-3: Method for the subjective assessment of intermediate quality level of audio systems,” International Telecommunication Union, Geneva, Switzerland.
- Kuniavsky, M., Goodman, E., and Moed, A. (2012). “Observing the User Experience: A Practitioner’s Guide to User Research,” Elsevier Science & Technology, second edition, 2012. Ch. 8 More than words: Object-Based technique; Ch. 9 Field Visits: Learning from Observation; Ch. 15 Analyzing qualitative data.
- Mathison, S. (2005). “Think-Aloud Protocol,” In: *Encyclopedia of Evaluation* edited by S. Mathison, by Sage.
- McCormack, A., and Fortnum, H. (2013). “Why Do People Fitted With Hearing Aids Not Wear Them?,” *Int. J. Audiol.*, **52**, 360-368.