

PARTICIPATORY ENQUIRY FOR A BIONIC VOICE

M. Hagmüller¹, A. K. Fuchs¹, C. Bath²

¹Signal Processing and Speech Communication Laboratory, Graz University of Technology

²Technische Universität Braunschweig, Germany

{[hagmueller.anna.fuchs](mailto:hagmueller.anna.fuchs@tugraz.at)}@tugraz.at, c.bath@tu-braunschweig.de

Abstract: People who have lost their larynx and thus speech functionality need a substitution voice to regain speech. Three main approaches exist, all of which have severe disadvantages. Previously, we have been working on improving the state-of-the-art for an electronic speaking aid. The current stage of our project has a special focus on a gender appropriate voice for laryngectomised speakers. To better understand the needs of the potential users of a bionic voice we adopted a participatory inquiry that involved interaction with 17 people without a larynx, of which 9 were female. All common substitution voices were used in the test sample. We spent between 1.5 and 6 hours with the individuals per session and had one to four visits. We learned that for all of them a natural voice is important. Most of the laryngectomees reject the use of a speaking aid, because of its bad sound. Women were specifically against the speaking aid. Desired properties of a bionic voice were an assertive voice, a voice matching ones personality. Women want to be recognized as female and have an attractive voice. They suffer from the low fundamental frequency of all substitution voices.

Keywords: alaryngeal speech, bionic voice, participatory enquiry

I. INTRODUCTION

For people who suffer from laryngeal cancer or similar diseases, the last resort is a total laryngectomy, which results in a loss of speech. Currently there are around 25.000 people who have undergone a laryngectomy in Germany, around 10% of which are female.

After the larynx is removed surgically, the anatomy has changed dramatically, as depicted in Figure 1 (a) and (b). The trachea ends at the so called tracheostoma at the neck and the vocal tract is shortened. The vocal folds are missing, and thus the possibility to produce voiced speech.

There are three alternatives for people to regain their speech. (1) For esophageal voice air is gulped and then released in a controlled manner and the tissue of

the pharyngo-esophageal segment in the pharynx vibrates. (2) A Tracheo-esophageal shunt valve is placed between the trachea and esophagus and therefore speech can be generated as above but with the air coming from the lungs (Fig. 1c). Although in Western Europe the tracheo-esophageal voice is the primary method of speech rehabilitation the situation is different in other countries and often it causes problems due to a leaking valve [1]. (3) The transcutaneous electronic speaking aid device (EL) is a small, hand-held and battery-driven device. The vibrating coupler disk of the device is held against the neck. The signal of the coupler disk is carried into the vocal tract. The EL is the focus of our research.

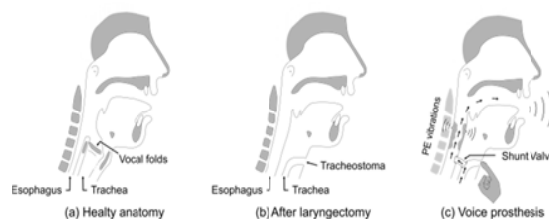


Figure 1: Anatomical details of (a) a healthy neck, (b) after laryngectomy, and (c) speaking with a voice prosthesis (from [2])

Major drawbacks of the resulting speech using the EL are the directly radiated noise of the device itself, the unnatural, monotonous quality of speech and the need of one hand to operate the device [3]. For the past years we have been working to improve the EL in order to increase the communication quality of the users. Regarding device operation, it is inconvenient to use one hand to operate the device. The main disadvantage is the inadequate quality of the resulting speech. The current technology of electronic speaking aids has been available for more than half a century [4] and there has been no major improvement of intelligibility and naturalness since then. An overview of the state-of-the-art and our research results has been published in [5]. In the scientific literature, we encounter two streams of approaches to improve the situation. (1) Technical approaches: the properties of EL speech and its differences to healthy speech are analysed (e.g. [6]); filtering techniques or similar

approaches to reduce the differences are applied (e.g. [7], [8], [9]). The resulting speech is evaluated objectively or subjectively with more or less appropriate listeners. (2) Researchers try to learn about the situation of affected people by sending out questionnaires, analyse and draw conclusions from the answers (e.g. [1]).

Female and male laryngectomees have different needs and requirements. Women are much less likely to be laryngectomised, therefore there is only limited research focused on women and much of the research done on men cannot be generalized to also include women [10], [11]. Much more data seems to be available on research about transgender women (e.g. [12]). The challenge to acquire a new voice seems to be similar to our topic of research. For transwoman there are clinical guidelines to support them to develop an appropriate female voice.

In order to overcome the shortcomings of existing approaches we planned to give the users a voice in the research process to reduce the bias that is unavoidable when only researchers make up their minds without incorporating potential users of their research. We wanted to address the specific problems women have to face, when they are forced to use a substitution voice.

There are several questions we wanted to explore together with people using a substitution voice. (1) What are the requirements of people who on a substitution voice concerning their verbal communication? (2) Do different user groups have different requirements? (3) What are specific situations that make it especially difficult to communicate with a substitution voice? (4) What is the reason so few people use an EL?

The rest of the paper is organized as follows. We first describe our methodology and the available subjects. In the results section we summarize the findings and the discussion section we reflect on the interactions with the users and we finally draw some conclusions.

II. METHODOS

We were guided by the methods of contextual design that are used for getting to know the work process of potential users of new software that should improve those processes [13].

A. Interview partners

We performed informal interviews with the potential users and most of the time, spent a longer period of time (1.5-6 hours) with them. Most of the interviews were with a single laryngectomised person, sometimes together with their partners. In addition we

had two meetings with a group of people. We are aware of one important bias in our study, as we only had contact with socially active people, who were interested in the research. When possible they went back to their workplace, are involved in social life and have learned to cope with their new situation. Others withdraw themselves from social interaction and people from that group were not interested in an interaction with us. They might have different requirements than the active group, but we don't have a possibility to assess their needs with this methodology.

We originally planned to work with regular users of an EL. We took a lot of effort to find women who use an EL, but we were not able to find any woman that uses this as her primary substitution voice. Therefore we included users of any substitution voice. For an overview on gender and means of communication see Table 1. Our small statistic reflects qualitatively what is reported in literature on the distribution of substitution voices. We only have a high proportion of EL users because we were specifically looking for them. One woman was communicating with pen and paper only. The person who whispers had only the vocal cords removed.

We complemented this first-hand information with a discussion with the team of phoniatricians and speech and language therapists (SLPs) at the phoniatric department at the ENT university clinic Graz.

We organized the interaction in several meetings that were structured as follows:

Table 1: Distribution of gender and means of communication: EL ... electronic speaking aid, ES ... esophageal voice, TE ... tracheo-esophageal voice, PN ... Pen, WH ... whisper

	EL	ES	TE	PN	WH	Total
Male	4	1	3	0	0	8
individual visit	4	1	0	0	0	5
group talk	0	0	3	0	0	3
Female	0	3	4	1	1	9
individual visit	0	2	2	1	1	6
group interview	0	1	2	0	0	3
Total	4	4	7	1	1	17

B. Structure of interaction

1) The first visit was aimed at getting to know the person and introducing ourselves. We emphasized that we visited them because they are the experts concerning their voice and we wanted to better understand their specific needs and problems. We then suggested spending up to half a day with them to get to know them better. We also ask for a specific scenario that is a challenge for their communication abilities and whether we could be take part in it and observe

them. 2) For the second visit we observed a challenging communication scenario, e.g. pub or shopping. We observed the interaction with other people and the challenges that arose because of the specific situation. 3) For the third visit we continued from the second session in a different situation and then presented our bionic voice test system.

C. Bionic Voice System

Our bionic voice test system is an improved version compared to what we presented in [5]. We use a small transducer that is attached to the neck with a neck-collar above the tracheo-stoma. The transducer is driven by a headphone amplifier that gets the signal from a notebook. We also use a head-set microphone to pick up the speech sound and use this information to calculate an F0 contour. The Matlab based system allows modifying the voice quality by means of changing the parameters of the LF-model, which is used to generate the excitation signal. The users get a wireless button to turn the signal on and off.

At the current stage, we have gone through the whole cycle with the four subjects using the EL. For the non EL users we did only complete the interviews and with some we tried to do the hands-on experiment with our bionic voice test system. We realized without a sufficient proficiency regarding speaking with an EL, the experiment didn't make much sense.

III. RESULTS

A. Interviews

a) The learnings can be summarized in three categories.

1) **Specific problems female speakers have when using a substitution voice:** Even though losing the voice is a traumatic experience for everyone, female speaker especially suffer from the quality of the substitution voices. The low pitch frequently leads to being identified as a male, which is especially critical when using telephony based services that require some form of identification. This has an impact on the feeling of self-worth and the question of attractiveness as a woman.

2) **Insights why we weren't able to find a female of an EL.** The robotic and monotonous sound of the electronic speaking device seems specifically repelling for women. A frequent comment of the female subjects on why they didn't want to use an EL was that they would rather communicate by writing than having such a strange voice.

3) **Requirements for an electronic speaking aid.** The most important shortcoming of all substitution voices seems to be the reduce loudness of their voice,

that results in not being able to take part in conversations in acoustically difficult settings. Examples we witnessed were settings such as in a restaurant, a shopping centre, an intercom at a barrier. b) Hands-free operation of the EL is another important requirement. Currently, conversation is very limited when doing something where both hands are needed, such as driving a car, cooking, or eating. People using the EL with the right hand have to change the device e.g. when shaking hands. c) Battery life. When talking a lot than the batteries drain a lot. We witnessed the use of up to four packs of battery for a period of half a day. d) The conversation over the telephone is a problem for all. We often hear that they only actively call but don't pick up the phone if they can avoid it. They report people hang up the phone when they hear the substitution voice. For EL users this seems particularly relevant. A more natural voice would reduce such situations.

B. Testing the Bionic Voice System

When testing our Bionic Voice System we got valuable feedback. All mentioned it was not loud enough and therefore could not solve one of their most important requirements. While the neck-collar was well received by some for others it seems not to be a good solution. Almost every neck was different, often due to additional problems, such as a neck dissection. A custom fit coupler disc would be necessary in some cases. One woman had issues with the pharyngeal reflex, so the collar was not an option for her.

The hands-free option, though it was not implemented in a way that would work in everyday life was confirmed as a very important feature.

The varying fundamental frequency was disturbing at first for all subjects. While some started to prefer it over the static pitch, some were not getting used to it.

In addition to voice related learning, we also learned methodological lessons. The first issue that we had to reflect was what impression our laboratory setup would leave on the users. A very complex setup with lots of cables and unfamiliar electronics might be intimidating and could create an unnecessary barrier between the scientists and the users.

IV. DISCUSSION

Once the volume of the voice was satisfied also the male speakers were concerned how they sounded. Women explicitly expressed that they were much more concerned how they sounded. We learned that woman have difficulties to accept the new voice because it doesn't sound feminine at all [11]. Some women decide to rather not speak at all than sounding like a male. In a study with 218 larygectomees (on average 6

years after surgery), 17% remain voiceless and 40% withdrew socially [14].

One older user explicitly mentioned that he didn't like technology so much. We therefore tried to reduce the visible technical complexity, while being open for those interested in the technology to explain what is going on behind the scenes. On the other hand, for younger subjects state-of-the-art technology was important, such as a connection to the smart phone.

We found it helpful to record the meetings with an audio recorder and not to rely to collect interview notes from memory in order to make the description as less subjective as possible. Since most of the times very personal issues came up, we also felt it not being appropriate when one of us was taking written notes during the conversation. Of course audio recordings were authorized by the users.

V. CONCLUSIONS

The interviews and the test of our bionic voice system showed, that there is a great need for an improved way of speaking for people without a larynx. Especially women are in need of a voice that is in line with their gender. The main problem of the current bionic

VI. ACKNOWLEDGEMENTS

This work was funded by the FEMTech project #849824 administered by the Austrian Research Promotion Agency (FFG) and the HEIMOMED Heinze GmbH & Co. KG

REFERENCES

- [1] S. R. Cox and P. C. Doyle, "The Influence of Electrolarynx Use on Postlaryngectomy Voice-Related Quality of Life," *Otolaryngology -- Head and Neck Surgery (OTO-HNS)*, vol. 150, pp. 1005-1009, 2 2014.
- [2] J. Lohscheller, "Dynamics of the Laryngectomy Substitute Voice Production," 2003.
- [3] G. S. Meltzner and R. E. Hillman, "Impact of Aberrant Acoustic Properties on the Perception of Sound Quality in Electrolarynx Speech," *Journal of Speech, Language, and Hearing Research*, vol. 48, pp. 766-779, 8 2005.
- [4] J. L. Flanagan, "Artificial larynx". US Patent 3,291,912, 1966.
- [5] A. Fuchs, M. Hagmüller and G. Kubin, "The New Bionic Electro-Larynx Speech System," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, pp. 952-961, 2016.
- [6] M. S. Weiss, G. H. Yeni-Komshian and J. M. Heinz, "Acoustical and perceptual characteristics of speech produced with an electronic artificial larynx," *The Journal of the Acoustical Society of America*, vol. 65, pp. 1298-1308, 1979.
- [7] E. A. Goldstein, J. T. Heaton, J. B. Kobler, G. B. Stanley and R. E. Hillman, "Design and implementation of a hands-free electrolarynx device controlled by neck strap muscle electromyographic activity," *IEEE Transactions on Biomedical Engineering*, vol. 51, pp. 325-332, 2004.
- [8] K. Nakamura, "Speaking-Aid Systems using Statistical Voice Conversion for Electrolaryngeal Speech," Ph.D. dissertation, Nara Institute of Science and Technology, 2010.
- [9] H. R. Sharifzadeh, "Reconstruction of Natural Sounding Speech from Whispers," Ph.D. dissertation, Nanyang Technological University, Singapore, School of Computer Engineering, 2011.
- [10] W. H. Gardner, "Adjustment problems of laryngectomized women," *Archives of Otolaryngology*, vol. 83, pp. 31-42, 1966.
- [11] S. R. Cox, J. A. Theurer, S. J. Spaulding and P. C. Doyle, "The multidimensional impact of total laryngectomy on women," *Journal of Communication Disorders*, vol. 56, pp. 59-75, 2015.
- [12] S. Davies, V. G. Papp and C. Antoni, "Voice and Communication Change for Gender Nonconforming Individuals: Giving Voice to the Person Inside," *International Journal of Transgenderism*, vol. 16, pp. 117-159, 7 2015.
- [13] H. Beyer and K. Holtzblatt, *Contextual Design: Defining Customer-centered Systems*, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1998.
- [14] H. Danker, D. Wollbrück, S. Singer, M. Fuchs, E. Brähler and A. Meyer, "Social withdrawal after laryngectomy," *European Archives of Oto-Rhino-Laryngology*, vol. 267, pp. 593-600, 9 2009.