Interspeech 2022
18-22 September 2022, Incheon, Korea

Humanizing bionic voice: interactive demonstration of aesthetic design and control factors influencing the devices assembly and waveshape engineering

Konrad Zieliński[12], *Marek Grzelec*[1], *Martin Hagmüller*[3]

[1]UHURA Technologies
[2]University of Warsaw, Faculty of Psychology, Human Interactivity and Language Lab
[3]Graz University of Technology, Signal Processing and Speech Communication Laboratory
konrad@uhura.pl, marek@uhura.pl, hagmueller@tugraz.at

Abstract

Electrolarynx is a speech aid providing voice source for people who have their larynx resected. Bionic voices (BV) extend the device capabilities by combining the biological and technical parts in a functioning communicative system. This interactive demonstration illustrates the challenges in material design and sound engineering in the domain of BV which are aimed at enriching the performance of a conversing dyad in the social context. The factors of control and aesthetics will be introduced by alaryngeal speaker presenting novel prototypes. The participants will experience the concepts in a multimodal demo, try various devices and talk with a laryngectomee.

Index Terms: electrolarynx, bionic voice, speech aids, sound engineering, human-computer interaction, laryngectomy

1. Introduction

Laryngectomy, removal of the larynx affected by cancer, profoundly changes the structure of control over social interactions. Even when a substitution voice is provided, the person is not able to produce voiced speech naturally i.e. with the adequate loudness, accompanying gestures and dynamically changing the intonation. In particular the electronic larynges provide people with the basic voice source, yet constraining arm gestures and limiting pitch control [6].

2. Rethinking electrolarynx design

This paper is geared towards demonstrating the challenges and proposed solutions for electrolarynx enriched with AI, signal processing and/or manual interfaces. We aim to show the influence of factors constraining the devices design and waveshape engineering. Some speech features responsible for human-likeness could be predicted by algorithms [1,2,3,8]. Yet when advanced signal processing is introduced it has to be taken into the consideration that the person's control over social interaction might be constrained by the computer. Such manipulation made by a device with no access to rich contextual information (e.g. gaze shifts, gestures) might limit the interaction possibilities, leaving limited space for adapting the voice to a particular situation, interlocutors and context. Enriching human-human interaction, especially in its fragile and dynamic aspects, requires the distribution of control between the given person and the utilized computer system. The interplay/changeability between the modes of control (automatic, manual) differs for a particular individual and social context constituting complex coordination problems [9]. We should choose between such modes of control as humans.

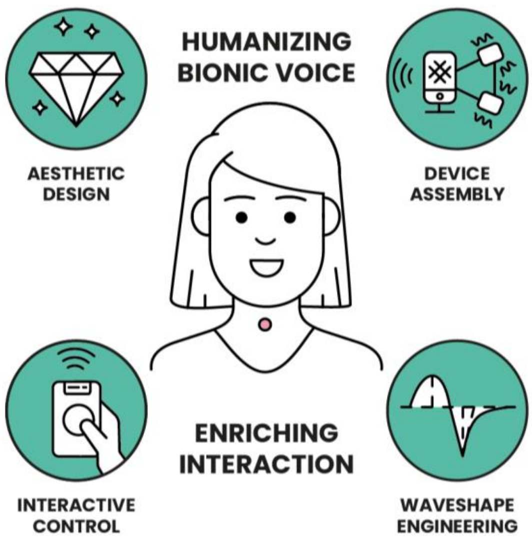

Figure 1: *The complex interaction between contributing factors: aesthetic design and interactive control influence devices assembly possibilities and waveshape engineering.*

The basic design of electrolarynx has not changed for 70 years – the majority of the devices consist of a hammer excited by a moving coil and a coupler disk [6]. Several improvements have been made in the past [2,4,7], yet we argue that to provide fluent interaction one cannot consider a single dimension only. The combination of the aesthetic design of the equipment, as well as interfaces providing relevant manual control, materials used and signal processing providing excitation signals leading to a more natural sound cannot be seen in isolation from another (Fig. 1). Failure in one dimension (e.g. not providing relevant control) might result in users not willing to adopt such a system [10]. There have been attempts to provide hands-free design of an electrolarynx [1,3,4,7,8], replacing the electro-acoustic transducer [2,3], producing a more natural voice by improving the transducer excitation signal with the use of linear prediction inverse filter or voice models such as the Liljencrants-Fant (LF) model [2,3]. Another approach is to provide vibrations directly to the mouth avoiding a difficult beck barrier [1] or using two different transducers on both sides of the neck [8]. Moreover, from the perspective of the global south affordability becomes an important parameter, e.g. see [5] for an affordable design with a vibration motor (our approach to the device based on vibration motor will be also shown during the demonstration).

Copyright © 2022 ISCA

3. On-site comparison: experiencing research directions and first prototypes

We wish to discuss with the INTERSPEECH community our current research directions within the European Electrolarynx Project. The participants will test recent prototypes during on-site demonstration and virtually during live video-stream. The interactive demonstration will be conducted by the actual laryngectomee using novel bionic voice designs who will present the importance of challenges from his perspective.

3.1. Interaction: try yourself and talk with the user

The concepts related to the consequences of laryngectomy are difficult to explain and feel via the scientific paper only. Therefore we invite you to: i) talk with alaryngeal person using several devices, ii) try to use them by yourself. We learnt that the feelings associated with the voice loss are much easier observable in the direct contact with the people who have their voice compromised. That is why the demonstration will be led by a larynx amputee himself (paper's first author).

3.2. Experience: see, hear and touch the difference

We will explore various aspects of design dimensions (devices assembly, sound engineering, aesthetic, interactivity). Specifically we will focus on comparing the control over the waveform's f0 by a person via the pressure sensor [10], contrasted to a control by a prediction model [3] – the difference presented in the example of the user operating the devices. Moreover, as the source signal shape is crucial for natural sounding voice, the speech recordings produced with varying waveforms will be presented for comparison: square, sine, LF model, inverse filter. You will touch the functional designs, so that you could compare the differences (Fig. 2).

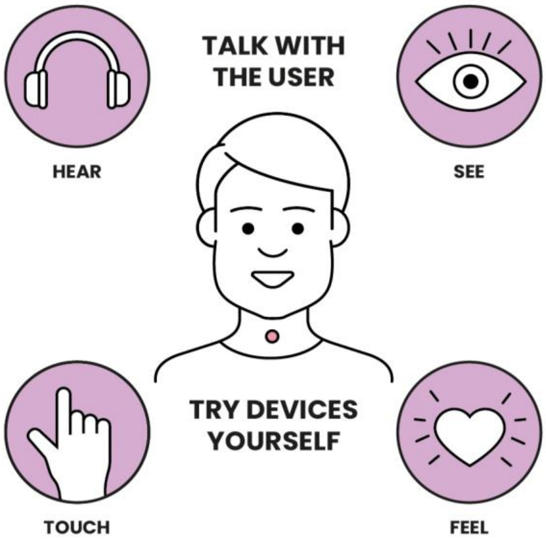

Figure 2: *On-site comparison: multimodal demonstration conducted by a laryngectomee.*

3.3. Prototypes: functional devices and design directions

The following implemented devices will be shown: i) sound generating transducers based on a) bone conduction speaker [3], b) vibration motors (novel low cost design; see also [5]) (Fig. 3a). 2) BV controlled by a) pressure-sensitive button [10], b) automatic prediction model [3]. Additionally, the further design directions will be demonstrated on 3D-printed models of a hand controller and vibrating effector (Fig. 3b). We will show how control over the speech modulation and device design influence interactions and how it is convenient (or inconvenient) to converse with the use of the proposed devices. Interaction recordings will facilitate our further work.

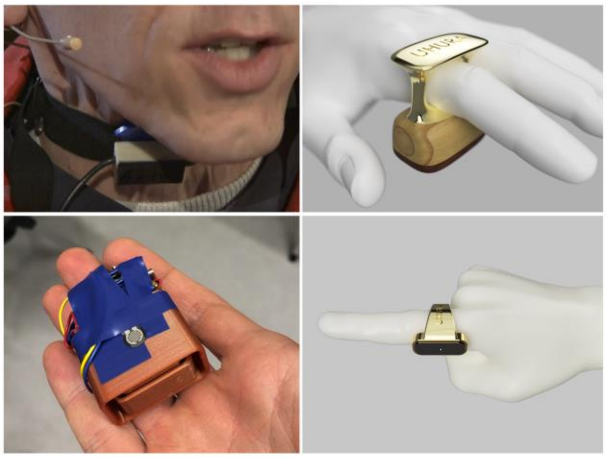

Figure 3: *Aesthetic and functional prototypes. a) left: BV based on bone conduction and vibration motors, b) right: designs of vibrating effector and hand controller.*

4. References

[1] Ahmadi, F., Kobayashi, K. and Toda, T. (2019) "Development of a Real-time Bionic Voice Generation System based on Statistical Excitation Prediction". *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. doi:10.1145/3308561.3354591

[2] Houston, K.M., Hillman, R.E., Kobler, J.B., & Meltzner, G.S. (1999). "Development of sound source components for a new electrolarynx speech prosthesis" in *Proc. ICASSP Vol4, pp.2347-2350.* doi:10.1109/ICASSP.1999.758409

[3] Fuchs, A. K., Hagmüller, M. and Kubin, G. (2016), "The New Bionic Electro-Larynx Speech System," in *IEEE Journal of Selected Topics in Signal Processing*, 10(5):952-961, doi: 10.1109/JSTSP.2016.2535970.

[4] Goldstein, Ehab A., et al. (2004), "Design and implementation of a hands-free electrolarynx device controlled by neck strap muscle electromyographic activity," *IEEE Transactions on Biomedical Engineering 51*, doi: 10.1109/CNE.2003.1196784

[5] Jayasena, K., D., N., U., D. et al., (2021), "Motor integrated Electro larynx; a novel, less expensive Electro larynx" in *Ceylon Journal of Otolaryngology*, doi: 10.4038/cjo.v9i1.5304

[6] Kaye, R., Tang, C.,, Sinclair, C. (2017), "The electrolarynx: voice restoration after total laryngectomy" in *Medical Devices: Evidence and Research*, doi: 10.2147/MDER.S133225

[7] Madden, B., Condron, J., Coyle, E. (2011). "Augmented Control of a Hands-Free Electrolarynx," *Bioengineering in Ireland*, doi:10.21427/D7T89T

[8] Takeuchi, M. Ahn, J., et al. (2022) "Hands-Free Wearable Electrolarynx using Linear Predictive Coding Residual Waves and Listening Evaluation" in *Advanced Biomedical Engineering*, doi: 10.14326/abe.11.68

[9] Zieliński, K. and Rączaszek-Leonardi, J (2022) "A Complex Human-Machine Coordination Problem: Essential Constraints on Interaction Control in Bionic Communication Systems" in *ACM CHI Conference on Human Factors in Computing Systems Extended Abstracts*, doi:10.1145/3491101.3519672

[10] Zieliński, K, Szamburski, R. and Machnacz, E (2019), "Post-laryngectomy interaction restoration system" in *Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion*, doi:10.1145/3308557.3308731