Artificial voice perception in the context of novel voice restoration technique for laryngectomees

Konrad Zieliński, Ryszard Szamburski, and Ewa Machnacz

University of Warsaw, Poland
konrad.zielinski01@gmail.com

Abstract. Laryngectomy leads to voice loss. Current voice restoration techniques are insufficient. New method is proposed based on silent speech interface and digital copy of user's voice. Analysis of artificial voice perception, particularly personality attribution, is conducted as a part of this project.

Keywords: laryngectomy · voice restoration · machine learning · silent speech · intelligent interfaces · speech synthesis · artificial voices · personality perception

1 Introduction

Laryngectomy is an invasive, radical surgical treatment of laryngeal cancer (Nowak et al. 2015). One of its consequences is voice loss. Novel voice restoration technique for laryngectomees is proposed (specified in section 4). It incorporates, among other components, artificial voice. This article presents this new possibility, focusing on artificial voice perception and how such perception influences the technologies used. The first task of the project is to review current research and answer the following questions:

1. Will users perceive a voice, even with lower quality, with their "biological trace" as more natural than the best (with respect to the voice quality) available text to speech voice?
2. Is it worthwhile to create an artificial voice with "biological trace" of a patient in the context of current project?

2 Voice and personality

How is our personality perceived by others? What factors should be taken into account while analysing this issue? Most of the studies on personality perception have focused on visual modality. Recently it has been noticed that, along with visual, other e.g. aural (voice) and haptic (touch) modalities could play a significant role in that process (Schirmer & Adolphs 2017). Authors had investigated emotion perception from those modalities on behavioral and neuronal level and

concluded that each of them engage different processing system and attributed information does not simply duplicate the visual one.

Other researchers also reach the conclusion that, among other factors, the sound of our voice plays crucial role in personality perception. Nass & Lee (2001) conducted an experiment with participants listening to an introvert or extrovert artificial voice in natural context. The subjects (introverts and extroverts) showed similarity attraction to one of the voices (also introverts and extroverts). It shows that voice is one of factors that contribute to personality perception.

Some authors even postulate a concept of *artificial personality* (Wester et al. 2015). Their experiment shows that adding disfluency (e.g. *uh, um, like, you know, I mean*) to synthesised voice can result in different personality traits being assigned to it. The authors claim that adding simple disfluencies in fact increases the level of naturalness attributed to these voices.

The subject of artificial voices is especially important in the context of emerging array of voice assistants (e.g. Alexa[1], Siri[2], Google Home[3], Cortana[4]). The perception of such systems would be an interesting research field in cognitive science. However, the same approach can be applied in order to get important information on how to help a specific group of people with the use of synthesised voice.

3 Laryngectomy

Let us consider people that have lost their voice. Laryngectomy is an invasive, radical chirurgical treatment for laryngeal cancer (Nowak et al. 2015). It is called *salvage surgery*, because it is the most efficient method for the most advanced tumors. Patients lose the ability to use their natural voice. In the light of the above research we can claim that with that, they lose opportunity to convey some of their personality aspects in interactions. Up to this moment there has been no technical possibility of saving this "biological trace" of the voice, but some new solutions appeared recently. Currently there are 3 methods of voice restoration (Tang & Sinclair 2015) listed below and rated according to following criteria:

- (Criterion 1) non-invasiveness
- (Criterion 2) naturalness of communication
- (Criterion 3) quality of the voice

- (VR 1) voice prosthesis (fulfills Criterion 2 and Criterion 3, does not meet Criterion 1)
- (VR 2) esophageal speech (fulfills Criterion 1 and Criterion 2, does not meet Criterion 3)

[1] Alexa: https://developer.amazon.com/alexa, retrieved 30.04.2017, 20:00
[2] Siri: https://www.apple.com/ios/siri/, retrieved 30.04.2017, 20:00
[3] Google Home: https://store.google.com/ca/product/google_home/, retrieved 30.04.2017, 20:00
[4] Cortana: https://www.microsoft.com/en-us/cortana, retrieved 30.04.2017, 20:00

– (VR 3) electrolarynx (fulfills Criterion 1 and Criterion 3, does not meet Criterion 2)

As can be seen none of those methods fulfill Criterion 1, Criterion 2 and Criterion 3 at the same time. Due to the drawbacks of the current methods of voice restoration after laryngectomy, our research group has proposed an alternative.

– (VR 4) Intelligent interface based on neuromuscular input and digitized biological voice

4 Novel voice restoration technique

Application of novel technologies from the field of artificial intelligence could help the group of patients following laryngectomy. The main project goal is to create a complete system of communication for laryngectomees, combining two approaches:

– (A1) AlterEgo system (Kapur et al. 2018). AlterEgo allows control of electronic devices without the need for any visible movement. The control relies on silent speech, like counting under one's breath. The system catches neuromuscular signals from selected face and neck areas responsible for speech production and then on the basis of previously learnt model, predicts words that have been "silently spoken". In its current stage, the project allows the user to count big numbers, play chess and Go with the help of a computer and to make simple queries (e.g. *What time is it it?*). Ultimately it is aimed to work with natural language, e.g. to allow write Google queries without the need of mouth movement or taking out a smartphone.
– (A2) Development in the field of biological voice digitalization. In the last years at least two commercially used models of English speech digitalization appeared, developed by companies Lyrebird AI[5] and CereProc[6]. Those companies offer a service of creating a "biological trace" of a voice - i.e. personal characteristics, or "fingerprint" of one's speech. After a few short recording sessions a complete text-to-speech system is created, with a voice easily recognizable by an user and relatives as "their voice".

System that we propose draws from both (A1) and (A2), combining and modifying them.
The training session:

1. neuromuscular signals from the face and neck of the patient are detected using EMG sensors. At the same time his/her voice is recorded;
2. recordings are manually transcribed to text;
3. once the voice is recorded it is converted into an artificial voice model, but with "biological trace" of an user (M1);

[5] Lyrebird AI: https://lyrebird.ai/, retrieved 30.04.2017, 20:00
[6] CereProc: https://cereproc.com/, retrieved 30.04.2017, 20:00

4. based on information from 1. the machine learning model is built with EMG signal-derived features and text transcriptions as a predicted values (M2).

Speech production:

1. patient tries to speak normally (that's the difference between "silent speech" interfaces and the system proposed by us). Here, mouth movements are desired increasing naturalness of speech;
2. the physiological movement signals are collected from his/her mouth and neck muscles with EMG sensors;
3. the system predicts the text that should be produced by similar movements basing on the previously built model (M2);
4. this text is converted according to previously built digital copy of user's voice (M1);
5. afterwards the sound (speech) is played from a speaker attached to the patient's body. It is aimed to be a place that the patient will not see and take little space. On the current stage of the project a JBL Go speaker is attached to the user's forehead with an elastic band.

There have been first attempts of using silent speech system to help patients following laryngectomy (Fagan et al. 2007, Meltzner et al. 2017), but none have tried to combine it with the digital copy of patients' voice using a method that could be used in clinical conditions yet. Combining (A1) and (A2) could lead to creation of a non-invasive, natural voice restoration technique with good quality of the voice which meets established criteria of success (Criteria 1-3). It will allow to build a system which will exceed all three methods for voice restoration currently used: voice prosthesis, esophageal speech and electrolarynx.

The project raised numerous research questions. Among others:

- (RQ 1) Would the system be natural for patients?
- (RQ 2) How users would perceive their own body with a speaker attached to it and how others will perceive them?
- (RQ 3) How to integrate solutions into a system for a specific task (voice restoration after laryngectomy)?
- (RQ 4) How to manage complexity of the language?
- (RQ 5) How to build suitably fast system that allow use in everyday situations?
- (RQ 6) How to provide a system that will allow to prosody control?

Investigation of naturalness of the voice will be the first step within (RQ 1). The very basic hypothesis that we posited is that users will perceive a voice, even with lower quality, with their "biological trace" as more natural than the best (on the matter of voice quality) available text to speech voice (H1).

In order to test (H1) we have established two main tasks:

- (T1) Analize current literature on the subject of artificial voices perception and psychological aspects of laryngectomy;
- (T2) Conduct empirical study that will answer this question.

5 Psychological factors of voice perception

As people attribute personality traits to artificial voice, thus output of our system could affect the way how patients will be perceived by others e.g. could be perceived as extravert, deceptive etc. Based on that assumption we argue that it is crucial to choose a better way: building entirely artificial voice (maybe with assigned artificial personality) or artificial voice with patient's "biological trace".

Issues to be considered are the psychological aspects of laryngectomy. Operation is a daunting experience for patients and their relatives. The study of Bussian et al. (2010) suggest that psychiatric disorders affects approximately one fifth of laryngectomy patients. The mail survey study with a large group of respondents after laryngectomy (Kotake et al. 2017) suggest that the most important psychological adjustment after operation is recognition of oneself as a voluntary agent. Since, as indicated above, a voice is an important quality contributing to a person's perception by others, restoring the voice as natural as possible and having full control over one's voice is one of the key factors to restoring agentivity.

Still, a study by Vilaseca & Chen (2006) could be mentioned. They showed that although patients identified speech among their most important problems, no correlation was found between speech and long-term quality of life (QOL). There is a need to establish which factors connected with voice, contribute to the self-perception. Such study would be greatly aided by the system we are going to create, where various factors can be manipulated and their effects can be measured.

6 Conclusions

After analysis conducted in previous section, we have drawn first preliminary conclusion:
(Conclusion 1) Patients may prefer a system with "biological trace" of their voice. This is based on the premise and assumption:

- (P1) Other people seem to perceive a part of people personal traits basing on their voice (Nass & Lee 2001, Wester et al. 2015)
- (As 1) Patients after laryngectomy wants to be perceived similarly as before the surgery.

Combining (Conclusion 1) with our criterion of naturalness (Criterion 2) leads to:
(Conclusion 2) The effort to create digital model of patient's voice is justified and it should be further investigated experimentally within (T2).

It seems that users could perceive a voice, even with lower quality, with their "biological trace" more natural than the best (in the respect of voice quality) available text to speech voice. As we can see is it worth building artificial voice with "biological trace" of a patient. We argue that the difference between artificial voice with or without "biological trace" of patient's is worth further

investigation, due to the fact that literature suggests that there is a difference in perception of those, but do not specify what is the character of this difference.

During the conference we have encouraged a discussion on the following questions:

- (Q1) People have most probably developed a well working system of how to attribute emotions and personality traits to others based on their voice. What are the possible cognitive biases in the context of emergence of artificial voices?
- (Q2) Is it really valuable to use patient's natural voice in our novel voice restoration technique? Do we need to try to falsify (H1) more fiercely and why?

References

1. Belin, P., Fecteau, S., Bedard, C. (2004). *Thinking the voice: neural correlates of voice perception.* Trends in cognitive sciences, 8(3), 129-135.
2. Bussian, C., Wollbrück, D., Danker, H., Herrmann, E., Thiele, A., Dietz, A., & Schwarz, R. (2010). *Mental health after laryngectomy and partial laryngectomy: a comparative study.* European Archives of Oto-Rhino-Laryngology, 267(2), 261.
3. Fagan, M., J., Ell, S., R., Gilbert, J., M., Sarrazin, E, Chapman, P., M. (2008). *Development of a (silent) speech recognition system for patients following laryngectomy.* In Medical Engineering & Physics 30 419–425
4. Meltzner, G., S., Heaton, J., T., Deng, Y., De Luca, G., Roy, S., H., Kline, J., C. (2017). *Silent Speech Recognition as an Alternative Communication Device for Persons With Laryngectomy* In IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 25, NO. 12
5. Kapur, A., Kapus, S., Maes, P. (2018). *AlterEgo: A Personalized Wearable Silent Speech Interface.* In 23rd International Conference on Intelligent User Interfaces (pp. 43-53). ACM.
6. Kotake, K., Suzukamo, Y., Kai, I., Iwanaga, K., Takahashi, A. (2017). *Social support and substitute voice acquisition on psychological adjustment among patients after laryngectomy.* European Archives of Oto-Rhino-Laryngology March 2017, Volume 274, Issue 3, Pages 1557–1565
7. Nass, C., Lee, K., M. (2001). *Does Computer-Synthesized Speech Manifest Personality? Experimental Tests of Recognition, Similarity-Attraction, and Consistency-Attraction* TJournal of Experimental Psychology: Applied, Vol. 7, No. 3, 171-181
8. Nowak, K., Szyfter, W., Wierzbicka, M. (2015). *Nowotwory w otolaryngologii, rozdział XII: Nowotwory krtani* Wydawnictwo Termedia, 279-335.
9. Schirmer, A., Adolphs, R. (2017). *Emotion Perception from Face, Voice, and Touch: Comparisons and Convergence* Trends in cognitive sciences, 21(3), 216–228. doi:10.1016/j.tics.2017.01.001.
10. Tang, C., G., Sinclair, C., F. (2015). *Voice Restoration After Total Laryngectomy.* Otolaryngologic Clinics of North America. Volume 48, Issue 4, August, Pages 687-702
11. Vilaseca, I., Chen, A. Y., & Backscheider, A. G. (2006). *Long-term quality of life after total laryngectomy.* Head & neck, 28(4), 313-320.
12. Wester, M., Aylett, M., Tomalin, M., Dall, R. (2015). *Artificial Personality and Disfluency.* INTERSPEECH 2015.