ANDREA CALVO, PIERO COSI, GIULIO PACI, VALENTINA PASIAN,
GIACOMO SOMMAVILLA, FABIO TESSER

Is voice personalization suitable and useful for ALS patients?

This study arises from the collaboration between MIVOQ, the ALS association CRESLA, and CNR, and it is still in an early preparatory and development phase. The aim of this experimental study is that of verifying the impact on the quality of life of patients and of his family members who use a speech synthesizer able to replicate the vocal timbre of the user's voice. The specific objectives of this work include understanding how "Your Digital Voice" is suitable and useful for ALS patients.

Key words: ALS, Voice Personalization, TTS.

1. *Introduction*

This work will be focused on voice preservation for patients with an illness that will inhibit them to speak (for ex. ALS Amyotrophic Lateral Syndrome/Lou Gehrig's disease).

Amyotrophic Lateral Sclerosis (ALS) is a neurodegenerative disease characterized by a progressive impairment of motor neurons, which causes the gradual loss of volunteer muscles function. 25-30% of the patients with ALS presents dysarthria as initial or predominant sign early in the disease phases. Dysarthria come as a first symptom in patients with ALS and it is 8-times more frequent compared to dysphagia (swallowing difficulty). Moreover, dysarthria affects up to 70% of patients with spinal onset in late phases (Chiò *et al*., 2017).

These deficit symptoms may not be evident until about 80% of the motor neurons have been damaged. Moreover, the interval between the onset of the articulators deficit and the diagnosis can vary from 33 months before the ALS diagnosis up to 60 months after the diagnosis.

Dysarthria occurs in more than 80% of the ALS patients and can produce a considerable disability, in a more precocious way in those with bulbar onset that can get an anarthria after few months. The loss of the communicative possibilities prevents these patients to participate in many activities and lead to social isolation. Dysarthria therefore significantly reduces the quality of life of people affected by ALS.

The intervention in support of such symptoms is speech therapy, but a direct intervention only to orality is not recommended as it may result in patient fatigue and prevent effective communication. In these cases, intervenes the AAC (Augmentative Alternative Communication) which is a field of clinical practice which has as its

objective the compensation of a disability (temporary or permanent) of expressive language. They try to recreate the ideal conditions so that the subjects have the opportunity to effectively and efficiently communicate, translating the content of their thoughts in a series of intelligible signs for their interlocutors. Often the use of communication aids allows to reach those goals.

2. *Personalized TTS for ALS patients*

The personalized speech synthesis (hereinafter referred as PTTS) allows to capture the timbric and prosodic characteristics of a specific user, and to make them available in the form of a voice for text-to-speech synthesis.

To generate a custom TTS voice, users must record the sentences with their own voice in the early stage of the disease, and convert it to a digital format through a TTS software. In this study, it will be used the MIVOQ technology, which allows to create a TTS voice with a hundred sentences. For reference, these sentences are recorded by users without disabilities in about 30 minutes.

MIVOQ PTTS implements two Speech Technology innovative techniques, known in the field of research as Statistical Parametric Speech Synthesis (SPSS) (Zen, Tokuda & Black, 2009) and Speaker Adaptation (Yamagishi, Kobayashi, Nakano, Ogata & Isogai, 2009). They have several advantages and among them:
– a customized synthetic voice can be created with few speech samples;
– TTS is "flexible" and "creative"[1].

This is the reason why we adopt for this aim "Your Digital Voice", a text-to-speech (TTS) service, developed by MIVOQ s.r.l., that allows fast and easy customization of synthetic voices. In order to get his/her own digital voice, the user has to record just a few dozens of sentences, without the need of a professional audio setup. Those users, who subsequently have lost the ability to speak, can still use their voice, in digital form, with the help of "Your Digital Voice" and a Speech Generating Device (SGD). The added value of this innovative technology is that it requires very few sentences to be recorded by the user, it is multilingual, it can be used in complete autonomy even by non-expert users, and it also allows for a more creative use of TTS (emotive speech, focus on specific words, etc.). The current prototype of the MIVOQ "Your Digital Voice" service consists of a web application where users can record sentences and automatically obtain a personalized TTS voice model. The voice model possesses the vocal characteristics (timbre, prosody) of the author of the recordings and can be used to convert any text into speech.

---

[1] Mivoq is developing a creative and flexible technology for TTS: it will be possible to create synthetic speech which incorporates emotions (sad, happy,...), audio effects (childish voice, monster voice,...) and vocalizations (laughter, cry, fillers like "uh" or "um",...).

3. *Goals*

The aim of this experimental study is that of verifying the impact on the quality of life of patients and of his family members who use a speech synthesizer able to replicate the vocal timbre of the user's voice. The specific objectives of this work include understanding how "Your Digital Voice" is suitable and useful for ALS patients.

Also, it is important to understand users' needs more deeply. In fact, we need to better understand also how much TTS personalization can improve the life of patients. Experiments should be set up that could clarify these issues. Therefore, the experimentation with patient is a priority and, to do so, it is necessary to integrate MIVOQ "Your Digital Voice" system into Speech Generating Devices currently used by ALS patients.

Moreover, this study aims at subjectively evaluate the TTS / PTTS. Firstly, comparing the custom synthetic voice of a user, from the point of view of quality and intelligibility, with standard synthetic voices already present in Speech Generating Devices; and, subsequently, by assessing how the custom voice resembles the patient's real voice.

4. *Methodology*

Patients diagnosed with ALS with the absence of dysarthria, but with a likely development of the symptom and patients with initial dysarthria (ALSFRS_R value of language ≥ 3) will be recruited, while patients with frontotemporal dementia diagnosis will be excluded from recruiting. A few dozens of sentences produced by these patients will be recorded and sampled.

Later, the newly created personalized TTS voice will be used as soon as the patient will require a support to communication (possible uses will be via smartphones, dynamic communicator, eye tracker etc. that is medium and high-tech aids).

After a trial period of about 2 months, the subjects will undergo a number of questionnaires that assess the perceived patients' and caregivers' quality of life:
– McGill Scale;
– Quality of Life during Serious Illness-Family Carers – QOLLTI- Fv2;
– Needs and effective communication – CETI-M;
– Psychosocial Impact of Assistive Devices Scale – PIADS;
– Quebec Evaluation of User Satisfaction with Assistive Technology – QUEST 2.0;
– an ad hoc created questionnaire.

Bibliography

Chiò, A., Mora, G., Moglia, C., Manera, U., Canosa, A., Cammarosano, S., Ilardi, A., Bertuzzo, D., Bersano, E., Cugnasco, P., Grassano, M., Pisano, F., Mazzini, L. & Calvo, A. (2017). Secular Trends of Amyotrophic Lateral Sclerosis: The Piemonte and Valle d'Aosta Register. In JAMA Neurology, 74(9), 1097-1104.

Yamagishi, J., Kobayashi, T., Nakano, Y., Ogata, K. & Isogai, J. (2009). Analysis of Speaker Adaptation Algorithms for HMM-Based Speech Synthesis and a Constrained SMAPLR Adaptation Algorithm. In *IEEE Transactions on Audio, Speech, and Language Processing*, 17(1), 66-83.

Zen, H., Tokuda, K. & Black, A.W. (2009). Statistical parametric speech synthesis. In *Speech Communication*, 51, 1039-1064.