Mobile system of vocal assistance for patients with surgical aphonia – SWARA

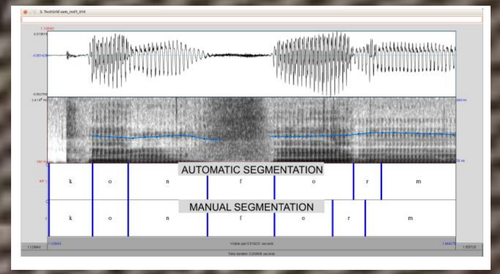
Magdalena Chirila[1], Cristina Tiple[1] , Florina Veronica Dinescu[1], Rodica Mureşan[1], Adriana Stan[2], Mircea Giurgiu[2]
[1]Department of Otorhinolaryngology, Iuliu Hatieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania; [2]Communications Department, Technical University of Cluj-Napoca, Romania.

SWARA – is a national project funded by the Romanian Ministry of Education with the main objective of enabling speech impaired persons, and especially those with surgical aphonia, to use a fast, personalised text-to-speech synthesis system.

Sample of the recording prompts along with their English translation.

RO	*Suntem una dintre cele mai vechi familii din Eforie.*
EN	We are one of the oldest families in Eforie.
RO	*Se poate face asta, dar depinde cum o faci, fiecare are modul lui de a vedea lucrurile.*
EN	This can be done, but it depends on how you do it, because everybody has his own way of looking at things.
RO	*Guvernul Britanic a comandat şaizeci de milioane de doze.*
EN	The British Government ordered sixty million doses.
RO	*De pildă, aş face un brand din brânza şi gemurile locale.*
EN	For example, I would brand the local cheese and jams.

Material and Methods: For the text-processing and prediction, we built a full stack front-end which includes text normalisation, phonetic transcription, syllabication and accent positioning, as well as a general text-predictor trained on over 5m tokens trawled from the Romanian Wikipedia website.The voice adaptation method is based on an HMM system trained eigen voice built on over 21hs of data recorded in a semi-professional studio, from 19 speakers aged 19-23. Each of the 19 speakers can then "lend" part of their speaking characteristics (such as timbre, rhythm or melody) to the resulting voice.



Results: During the SWARA project and following the procedures described above, we managed to record a number of 12 volunteer speakers with a total recording time of around 13 hours. However, the RSS corpus also contains two speakers which recorded the same set of prompts. And we also previously recorded 3 other speakers uttering the rnd1 , rnd2 and rnd3 subsets of the RSS corpus. The latter three speakers uttered the data in a TV studio using the same recording equipment as the one used for SWARA. The only difference is the presence of a slight reverberation due to the large recording room. The segmentation process for this data was the same as the one described in Section III.

Conclusion: SWARA Speech Corpus is one of the largest speech resources for the Romanian language. It contains over 21 hours of high-quality read speech data collected from 17 different speakers. 16 hours of the data is composed of 880 utterances read by each of the 17 speakers. The speaker similarity issue can be achieved by using speaker adaptation methods in statistical parametric speech synthesis frameworks.

Acknowledgement The research leading to these results has received funding from PN-II-PT-PCCA-2013-4 No6/2014 (SWARA).

UMF UNIVERSITATEA DE MEDICINĂ ŞI FARMACIE IULIU HAŢIEGANU CLUJ-NAPOCA