My aircraft talks to me: Current developments on voice synthesis as a modality in the cockpit of future fighter aircraft

TIM LAUDIEN and JAN-PAUL HUTTNER
German Aerospace Center (DLR)
Lilienthalplatz 7
38108 Braunschweig, Germany

ANA GARCÍA MORAL and DANIEL SÁNCHEZ ALONSO and PABLO RODRIGUEZ REY
Airbus Defence & Space
Paseo John Lennon
28906 Getafe, Spain

LEON ERICSSON
Totalförsvarets forskningsinstitut (FOI)
Gullfossgatan 6
164 40 Kista, Sweden

Within the EPIIC project – co-funded by the European Defense Fund – more than 20 partners, including major industrial companies, universities and research institutions, from twelve European countries are collaboratively researching innovative and disruptive technologies for the integration into the next generation cockpits of combat aircraft. The goal is to address the technological challenges that the future air warfare imposes. These are, among others, the collaboration with friendly uncrewed air systems (UAS) and to ensure air dominance in a multi-domain operational environment. In order to build up and maintain pilot situation awareness and, ultimately, succeed the mission, pilots need to be informed about their rapidly changing environment. Therefore, EPIIC investigates improvements for a plethora of cockpit systems. While the team "cockpit innovations consistency" creates the theoretical framework, the technical teams gain insights on the adaptive human-machine interface, virtual assistant, large area display, eyes-out technologies, crew monitoring systems and crew states identification algorithms, as well as innovative interaction modalities. The latter itself comprises research on brain-computer interfaces, multi-touch input, natural language processing, eye-tracking technologies, voice synthesis, and gesture-based interaction.

This paper presents the findings of the ongoing work from the voice synthesis task in the EPIIC project. Therefore, first an introduction is given that compiles a brief historical outline on voice synthesis, an overview of the working principle and state-of-the-art models for speech synthesis, and application of voice synthesis in everyday life as well as in aviation. Consequently, the voice synthesis models used to synthesize speech, that are Microsoft's SpeechT5, Bark by Suno, and the NVIDIA HiFiGAN are presented. Additionally, basic performance parameters of the models are investigated, compared and subjectively rated. Then, the experimental design of the first and second EPIIC technical demos are displayed and the outcomes are discussed. The demos were performed with active Spanish fighter pilots and included various interaction modalities (enumerated above). The modalities have been assessed independently from each other. Finally, a conclusion is drawn by summarizing the content and giving an outlook on possible improvements on the aural presentation of information to future fighter pilots. Furthermore, the upcoming work on interaction modalities in the EPIIC project is outlined.

Keywords: Voice Synthesis, Text-to-Speech, Voice Assistant, Interaction Modality, Future Fighter Aircraft, Cockpit Technologies, Human-Machine Interface

Co-funded by the European Union

This publication was co-funded by the European Union under the Grant Agreement 101103592, (EDF-2021-101103592-EPIIC | EDF: European Defense Fund). Its contents are the sole responsibility of the EPIIC Consortium and do not necessarily reflect the views of the European Union.