Project title:	"Multimodal multilingual human-machine speech communication"
Project Acronym:	AI-SPEAK
Milestone index:	M3.3
Version:	1.1



PROJECT MEETING REPORT

of the Project "Multimodal multilingual human-machine speech communication" (AI-SPEAK).

The meeting took place in Novi Sad, on the premises of the Speech Technology Group at the Faculty of Technical Sciences, University of Novi Sad, on **January 22nd 2026**, with participation of all project team members. The main objective of the meeting was to review the current status of preparations for the AI-SPEAK Challenge and define the remaining organizational, technical and dissemination activities required before the official launch of the competition. Particular attention was devoted to the competition structure, benchmark datasets, evaluation methodology, dissemination strategy and the timeline of upcoming project activities.

I. AI-SPEAK Challenge concept and competition structure

The meeting began with a review of the current status of the AI-SPEAK Challenge preparation.

The team confirmed that the AI-SPEAK Challenge represents one of the major dissemination activities planned within the project, intended to increase the visibility of the developed datasets while encouraging the international research community to address current problems in multimodal speech processing. Following several internal meetings held during the previous project period, the overall competition concept has now reached its final form.

It was confirmed that the first edition of the AI-SPEAK Challenge will consist of two independent competition categories addressing complementary research problems.

Category A focuses on **audio and video manipulation detection**, requiring participants to develop systems capable of distinguishing authentic multimedia content from manipulated recordings. The category is designed to stimulate research on deepfake detection and audiovisual forensics under realistic conditions.

Category B addresses the task of **speech-driven facial animation**, where participants are required to predict temporal sequences of ARKit-52 blendshape coefficients from speech and/or textual information in order to animate a realistic avatar suitable for real-time applications.

The team concluded that these two categories together represent a balanced combination of multimedia analysis and multimedia generation tasks, reflecting the scientific objectives of the AI-SPEAK project.

Although several additional ideas were discussed—including audiovisual speech recognition, lip reading, voice conversion and audiovisual speech enhancement—it was agreed that these topics should be considered for future editions after larger benchmark datasets become available and additional evaluation protocols are established.

Each competition category will therefore have its own benchmark dataset, evaluation server, technical documentation and baseline implementations, allowing participants to compete independently in each task.

II. Benchmark datasets and evaluation protocols

The second part of the meeting was devoted to reviewing the benchmark datasets and the corresponding evaluation procedures.

For Category A, the project team confirmed that the benchmark will consist of carefully prepared audiovisual recordings containing both authentic and manipulated multimedia content. The dataset was designed to ensure sufficient diversity of manipulation techniques, recording conditions and speakers, allowing fair comparison between different detection approaches. The evaluation protocol will primarily rely on standard classification measures, while maintaining hidden evaluation subsets that prevent overfitting to publicly available data.

The benchmark dataset for Category B was considered technically complete, with only minor documentation and packaging tasks remaining before release. The released training package will contain synchronized speech recordings, blendshape coefficient sequences extracted according to the ARKit-52 standard, phoneme alignments, transcripts and an avatar application intended for visualization and verification of generated animations. Participants will be free to use text, audio, or any combination of derived features as model inputs, while the expected output will be a sequence of blendshape coefficients stored in the predefined CSV format.

The team also reviewed the evaluation methodology. Besides objective animation quality measures, additional emphasis will be placed on computational efficiency. Participants will therefore report inference time, Real-Time Factor (RTF), output frame rate and algorithmic latency, while organizers reserve the right to repeat all measurements under a standardized Google Colab environment to ensure fair comparison between submitted systems.

Following discussion, it was agreed that internal baseline implementations and evaluation scripts would be prepared exclusively for the organizers and would not be publicly released. To ensure fair comparison, the ground-truth annotations for the evaluation datasets will remain unavailable to participants throughout the competition, while the final ranking will be determined exclusively by the organizers using the hidden reference annotations.

III. Dissemination strategy and community engagement

The third part of the meeting was devoted to planning dissemination activities aimed at ensuring broad participation in the first edition of the AI-SPEAK Challenge.

Following previous discussions, the team confirmed that the two competition categories target different participant profiles and should therefore be promoted through different communication channels.

It was agreed that Category A, devoted to audio and video manipulation detection, should primarily target high school pupils, especially those attending grammar schools and technical or electrical engineering schools, as well as undergraduate students in the early years of study. The objective is to motivate younger students to become involved in artificial intelligence and multimedia technologies through a practically oriented competition.

In contrast, Category B, focusing on speech-driven facial animation, was recognized as a considerably more demanding research task. Accordingly, dissemination activities for this category will primarily target senior undergraduate, master's and doctoral students, from electrical engineering and animation engineering study programmes, who are already active in speech technologies, computer vision and machine learning.

The project team agreed that invitations to participate will be distributed through several complementary channels. Official invitations will be distributed through institutional mailing lists to grammar schools, technical high schools and electrical engineering schools throughout Novi Sad and the Autonomous Province of Vojvodina. In parallel, invitations will be distributed to faculties and higher education institutions with programmes related to electrical engineering, computer science and information technologies at the Universities of Novi Sad, Belgrade and Niš.

In addition to institutional dissemination, consortium members will use their existing academic collaborations to promote the challenge among students at partner institutions. Particular attention will be devoted to informing students at the Faculty of Electrical Engineering of the University of Banja Luka and the Faculty of Electrical Engineering and Information Technologies of Ss. Cyril and Methodius University in Skopje through established research collaborations.

To further increase the visibility of the competition, the team agreed to prepare promotional materials, including posters and flyers, which will be distributed at the Faculty of Technical Sciences and other educational institutions that are easily accessible to the project partners. The same promotional materials will also be presented during faculty open days, study programme promotion events and other outreach activities organized by the Faculty of Technical Sciences, providing direct contact with prospective participants.

The meeting concluded that combining institutional communication, personal academic networks and on-site promotional activities will ensure broad dissemination of the AI-SPEAK Challenge among its intended target groups will ensure broad dissemination of the AI-SPEAK Challenge among its intended target groups and strengthen the overall visibility of the project.

IV. Final event planning and organizational activities

The final part of the meeting focused on organizational aspects of the closing AI-SPEAK Challenge event, during which the best-performing teams will present their solutions and the competition results will be officially announced.

Following discussion, the team agreed that the final event should be organized on 25 April 2026 in the ceremonial hall of the Science and Technology Park Novi Sad. It was concluded that the reservation procedure should be initiated as soon as possible in order to secure the venue and allow sufficient time for planning the accompanying activities.

The event is planned as a half-day programme with an expected duration of approximately four hours. The organizers agreed that participation in the final event should be open not only to competing teams but also to students, teachers and other interested visitors in order to maximize the dissemination impact of the project. Besides the presentation of competition results and the award ceremony, sufficient time will be allocated for networking among participants, project partners and invited guests. The formal programme will conclude with an informal lunch intended to encourage discussions between students, researchers and industry representatives.

The meeting also addressed the awards for the most successful teams. It was agreed that the winners would receive AI-SPEAK promotional materials and project merchandise, together with certificates of achievement and commemorative gifts carrying the visual identity of the project. The exact contents of the award packages will be finalized during the following organizational meeting.

To further enrich the programme, the team agreed that the event should not be limited solely to announcing the competition results. A short presentation introducing the objectives and achievements of the AI-SPEAK project will be included, together with several invited presentations showcasing selected student projects closely related to the project's research topics. This exhibition programme is expected to provide participants with additional insight into current research activities in multimodal speech technologies while promoting further collaboration between students and researchers.

Finally, the consortium agreed that the detailed agenda, invited speakers and logistical arrangements would be finalized during the next organizational meeting, once the expected number of participants had been confirmed.