AMIS: An Audiovisual Dataset for Multimodal XR Research

Abhinav Bhattacharya[1], Luís Fernando de Souza Cardoso[3], Andy Schleising[3], Gareth Rendle[4], Adrian Kreskowski[4], Felix Immohr[1,2], Wolfgang Broll[3], Rakesh Rao Ramachandra Rao[2], Alexander Raake[1]

[1]Institut für Nachrichtentechnik, RWTH Aachen University
[2]Audiovisual Technology Group, TU Ilmenau
[3]Virtual Worlds and Digital Games Group, TU Ilmenau
[4]Virtual Reality

Introduction

- Multimodal audiovisual datasets
- Consistent stimuli vs ‘live’ mediated communication
- High-quality multisensory stimuli for objective QoE in social XR [1]
- Support for auditory and cognitive research [2]
- Realistic full-body avatar render[3]

- **A**udiovisual **M**ultimodal **I**nteraction **S**uite: **AMIS**
- Synchronized A/V English monologues & dyadic conversations
- Synchronous recording
  - **Talking-head videos**
  - **Full-body videos**
  - **Personalized animated avatars**
  - **Volumetric video avatars**
- Video conference to hybrid social XR meeting
- Resource for interdisciplinary research

[1] Anna Ferrarotti and Marco Carli. 2023. Objective Metrics Definition for QoE Assessment for Extended Reality Applications. In Proc. of the 2023 ACM Int. Conf. On Interactive Media Experiences (Nantes, France) (IMX ’23). Association for Computing Machinery, New York, NY, USA, 439–442.
[2] Janina Fels, Cosima A. Ermert, Jonathan Ehret, Chinthusa Mohanathasan, Andrea Bönsch, Torsten W. Kuhlen, and Sabine J. Schlittmeier. 2021. Listening to, and Remembering Conversations between Two Talkers: Cognitive Research using Embodied Conversational Agents in Audiovisual Virtual Environments. In [Fortschritte der Akustik - DAGA 2021, Wien, Austria] (2021-08-15). 47. Jahrestagung für Akustik, Wien (Austria), 15 Aug 2021 - 18 Aug 2021, Deutsche Gesellschaft für Akustik e.V. (DEGA), Berlin, 1328–1331.
[3] Florian Weidner, Gerd Boettcher, Stephanie Arevalo Arboleda, Chenyao Diao, Luljeta Sinani, Christian Kunert, Christoph Gerhardt,Wolfgang Broll, and Alexander Raake. 2023. A Systematic Review on the Visualization of Avatars and Agents in AR & VR displayed using Head-Mounted Displays. IEEE Transactions on Visualization and Computer Graphics 29, 5 (2023), 2596–2606.

Collection Design

Captures	Hardware	Capture Settings
Talking-head videos	Canon EOS C500 Mark II Zeiss Compact Prime CP.3 85/T2.1 Lens	ISO:1000, Color: 3300K, 59.94 fps, Focal length: 85mm, Aperture: T4
Full-body videos	RED SCARLET-W 5K Camera Zeiss Compact Prime CP.3 21/T2.9 Lens	ISO:800, Color: 3322K, 59.94 fps Focal length: 21mm, Aperture: T2.9
RGB-D image streams	4x Femto Bolt Cameras	Color (RGB, 1280 × 720 px) and depth (640 × 576 px) at 30 fps
Face expressions	iPhone 12 pro	Animation using Reallusion iClone8 Software
Audio	Shure MX-150B Microphone with Shure QLX-D wireless system	48 kHz sampling rate

Fig 1: Recording setup

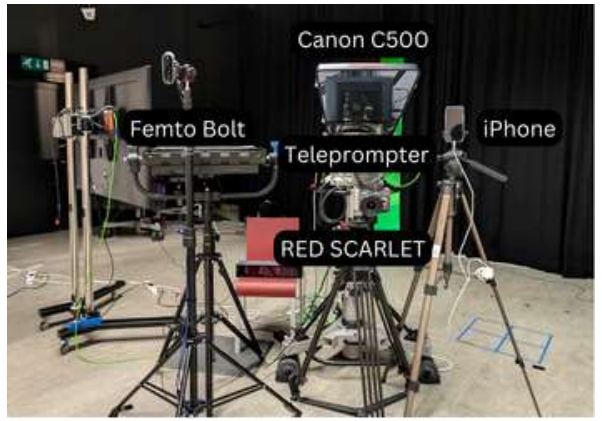

Fig 2: Multi-camera setup

Data Collection - Actors

- 2 Female (EF1, EF2)
- 1 Male (EM)
- Different ethnic backgrounds
- Language background (NS, NNS)
- 3-4 years of acting experience

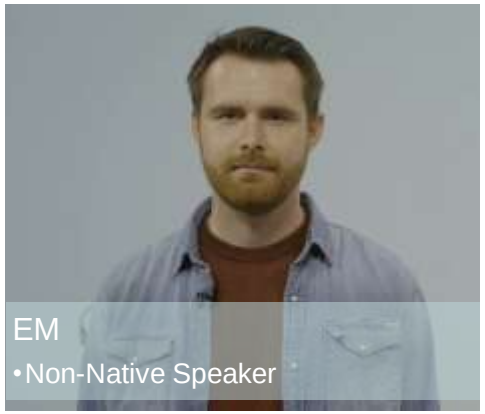

Fig 3: Actors' headshots

Data Collection – Recording scenarios

Monologues

- Replicable & comparable performances
- EUROM1 English Speech Corpus [4]
- Extended categories
  - Formal/Informal tones
  - direct/indirect speech
  - Serious emotion or humorous
- 3 actors x 9 monologues → Avg 22 sec

9x Monologues
• 3x Formal requests and complaints • 3x Personal anecdotes • 3x Narrative accounts

Dyadic Conversations

- Interpersonal interactions
- Short dyadic convo/ Small talk [5]
- Light, relatable topics [6]
- 2 conversations/actor pair
- One person recorded at a time
  - Actors prepared off-camera
  - Sync reactions and exchange

Fig 4: Word cloud of the conversation transcripts

[4] Dominic Chan,Adrian Fourcin, Dafydd Gibbon, Björn Granström, Mark Huckvale, Kokkinakis George, Knut Kvale, Lori Lamel, Børge Lindberg, Asunción Moreno, Jiannis Mouropoulos, Francesco Senia, Isabel Trancoso, Corin Veld, and Jerome Zeiliger. 1995. EUROM - a spoken language resource for the EU - the SAM projects. In 4th European Conference on Speech Communication and Technology.
[5] Isnaeni Wahab, Zul Astri, Novalia Tanasy, and Nurul Fachrunnisa. 2021. A Conversation Analysis: The Use of Small-Talk. Seltics Journal: Scope of English Language Teaching Literature and Linguistics 4, 1 (Jun. 2021), 53–62.
[6] Yanqiu Wei, Ling Zhang, Tao Lian, Ying Chen, Sihua Liao, Qi Guo, and Ping Hu. 2024. Happy storytelling promotes emotional contagion and interpersonal closeness. Current Psychology 43, 4 (2024), 2993–3001.

Suite Overview – Dataset

- 27 Monologues
- 6 Conversations (2 synced rec/convo)
- 156 visual sequences
- 39 compatible WAV audio files
- Around 2 hours of content

Fig 5: Overview of visual modalities per actor

Talking-head videos	Full-body videos	Personalized Animated avatars	Volumetric avatars
• MP4 files • 4K, 59.94 fps • 40 Mbps	• MP4 files • 4K, 59.94 fps • 40 Mbps	• FBX files • textures and meshes	• Reconstructed byte streams

Suite Overview - Additional Content

Supporting files

- Metadata
- 12 segmented reaction videos (Talking-head)
  - Non-verbal reactions/passive listening videos
- FFMPEG-based Python script
  - multiparty videoconferencing tiled videos
- Processing chain for adding a virtual background
  - DaVinci Resolve

AMIS Studio

- Unity-based demonstration tool
- Synced with corresponding audio tracks
- README and text walkthrough
- Dedicated player packages for each format

Fig 6: Screenshot of a conversation in the personalized animated avatar scene and the hybrid scene showing all visual formats synced with each other and the audio.

Video Demo

Link to demo video

Outlook

- Support for cross-lab reproducibility
- Exploratory study
  - Perceptual visual realism of avatars in VR
  - Impact on co-presence, and preference
  - Analysis of attention distribution through eye gaze behavior
- Extension in 6-DoF environment → proximity analysis

- Dataset extension
  - German version of AMIS
  - Bigger and more diverse actor pool
  - Optimized visual quality and animations

Fig 7: Perception test with virtual humans

https://github.com/Telecommunication-Telemedia-Assessment/AMIS

Thank you…