AMISCO: The Austrian German Multi-Sensor Corpus

HANNES PESSENTHEINER, THOMAS PICHLER, MARTIN HAGMÜLLER
{hannes.pessentheiner,hagmueller}@tugraz.at, thomas.pichler@student.tugraz.at
Signal Processing and Speech Communication Laboratory
Graz University of Technology, Austria

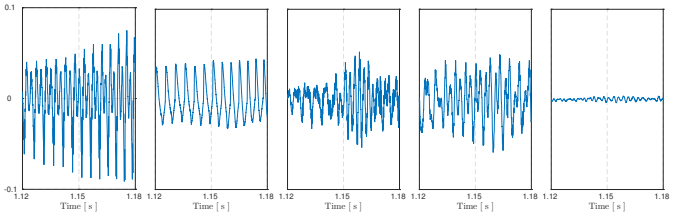
Abstract

- comprehensive, unique **speech corpus**
- **moving and non-moving speakers**
- **multi-channel audio signals** plus
  - fundamental frequencies
  - spatial positions
  - electro-glottograms
  - transcriptions
  - video data
- for joint parameter estimation, various machine learning and signal processing tasks, linguistic studies, studies related to fundamental frequency (pitch)

1. Acquisition Data

- **24 moving and non-moving speakers**
- balanced male and female
- **8.2 hours** of Austrian German read speech
- phonetically balanced sentences, commands, and digits
- **53,000 word tokens** and **2,070 word types**
- **43-channel** multi-sensor recordings in **two rooms**
  - **distant- and close-talking** speech recordings
  - laryngograph recordings representing **glottograms**
  - Kinect recordings transformed into **spatial trajectories**
  - video data provided by a camera
- 48 kHz audio files encoded with PCM S24 LE
- Kinect data points resampled with equally spaced 30 fps

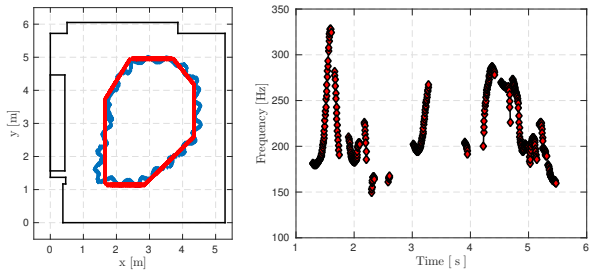
Time signals of a sentence's section:



(a) Headset. (b) Glottogram. (c) CPR 1. (d) MEMS M1/1. (e) Kitchen 15.

Time signals of a section of the (German-language) sentence.

Spatial positions and **fundamental frequencies** over time:



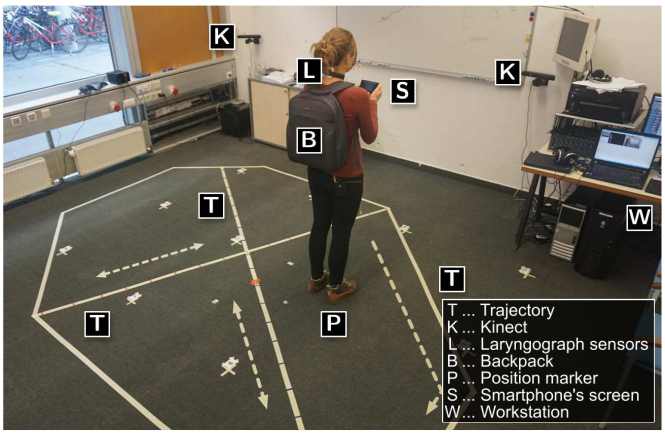
A speaker's skeleton tracks represented as trajectories (left) [red: given, blue: measured] and estimated fundamental frequencies over time (right) [marker: f_0 per time frame].

Signal-to-noise ratios (min., max., average over all speakers):

	Min-SNR [dB]	Max-SNR [dB]	Avg-SNR [dB]
Headset	23.58	52.23	38.67
CPR 1-2	18.97	36.43	24.73
CPR 3-5	19.14	35.75	24.84
CPR 6-8	19.32	36.66	25.09
CPR 9-14	19.06	37.14	24.33
Kitchen 15-17	17.27	31.54	21.35
MEMS M1	20.68	42.14	27.46
MEMS M2	21.97	45.09	29.62
MEMS M3	21.50	44.53	29.15

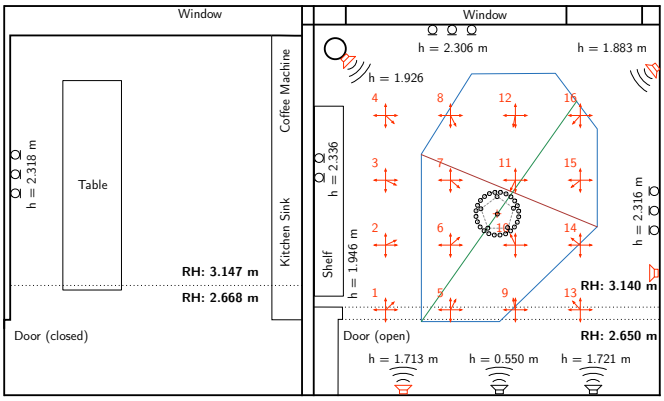
2. Data Collection

A Speaker's **Recording Situation**:



- 24 speakers, balanced male and female, aged 25-52
- 23 speakers with Austrian German and one speaker with German German to draw rough comparison between both variant's pronunciation
- body heights available to determine z-coordinates

Floorplan of the **Recording Environment**:



- $T_{60} \approx 500$ ms in meeting room and $T_{60} \approx 700$ ms in kitchen
- one 2-element linear array (30 cm)
- three 3-element linear arrays (60 cm)
- one 6-element pentagonal array (⌀ 54.44 cm)
- one 24-element MEMS circular array (⌀ 61.90 cm)
- one portable laryngograph
- one headset microphone
- one video camera
- four Kinects

Recording Procedure: Session (50-60 min.) composed of three sub-sessions (10-12 min.) where a speaker read

- 104 short items at different positions and orientations
- 40 long items while walking on marked trajectories
- 64 long items at different positions and orientations

3. Corpus Availability

- available **from summer, 2016**
- for universities, research communities and institutions
- see `www.spsc.tugraz.at/tools/amisco` for further details

Signal Processing & Speech Communication Laboratory
Graz University of Technology
Inffeldgasse 16c, 8010 Graz, Austria
http://www.spsc.tugraz.at

Acknowledgment: The K-Project ASD is supported in the context of COMET Competence Centers for Excellent Technologies by BMVIT, BMWFW, Styrian Business Promotion Agency (SFG), the Province of Styria - Government of Styria, and The Technology Agency of the City of Vienna (ZIT). The programme COMET is conducted by Austrian Research Promotion Agency (FFG). The DIRHA-Project was supported by the European Union Seventh Framework Programme for research, technological development and demonstration under grant agreement No. FP7-ICT-2011-7-288121. Furthermore, this project has received funding from the Marshall Plan Foundation. The Tesla K40 GPU-cards used for this research were donated by the NVIDIA Corporation.