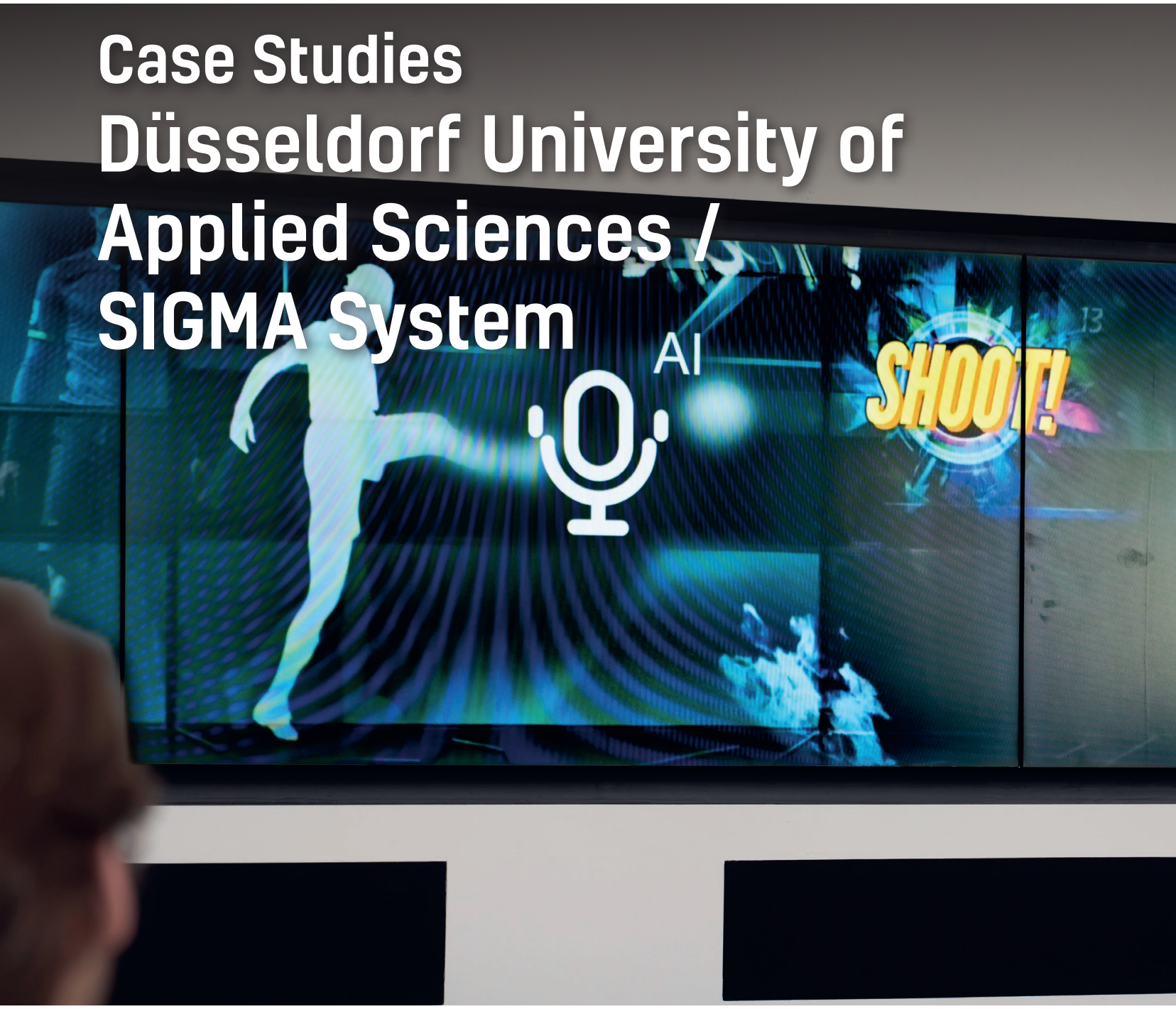
PROFESSIONAL system
Your Hub for AV & IT Integration
6 | 2026
December

# Case Studies
# Düsseldorf University of Applied Sciences / SIGMA System




Specialprint from Professional System 6-2026




# Local voice control for conference rooms

**There is usually one main objection to voice assistants in conference rooms: the cloud. SIGMA System Audio-Visuell and Düsseldorf University of Applied Sciences have therefore developed a voice control system that remains entirely within the room – without an internet connection, without stored recordings and as a complement to the existing media control system.**

Text: Johann Scheuerer | Images: SIGMA System Audio-Visuell.

„Hey Sigma, switch off the light above the table, raise the blinds and set the input to USB-C." Two to three seconds later, the room is ready without anyone having touched the touch panel. In the reference room of the Düsseldorf-based integrator Sigma System AV, the voice control system – developed in collaboration with Düsseldorf University of Applied Sciences – has been running in test mode for half a year.

The difference compared to standard assistant systems on the market lies less in the recognition itself than in the data path: a compact computer in the room handles everything, from the wake word to the handover to the control system. Wake-word recognition, transcription and command interpretation run without an internet connection; voice recordings are not stored.

The voice control system operates entirely on local

hardware. On the software side, it builds on existing room control systems and is set to be launched as a standalone product.

### The university project

SIGMA AV approached Düsseldorf University of Applied Sciences with the idea last year, where Emil Ötting, a Bachelor's student on the Data Science, AI and Intelligent Systems (DAISY) programme, spent six months further developing the system.

Stephan Nabbefeld, Managing Director at SIGMA System Audio-Visuell and himself an alumnus of the university, sees such a collaboration as mutually beneficial: companies can contribute development topics and get to know potential employees beyond the standard recruitment process – given the staffing needs in the AV sector, this provides access to profiles such as data science, which traditional media technology training programmes do not cover.

For Ötting, this involved gaining practical knowledge and the requirement to make assessments in a real-world business environment and stand by them. And a developer without industry experience asks different questions about a conference room than an experienced integrator would.

### Every room leaves its mark

Media technology in conference rooms is operated via touch panels, buttons or control apps. These interfaces assume that users are familiar with the system's logic – an assumption that regularly proves unfounded in everyday use. Stephan Nabbefeld attributes this to developments in recent years: projects have become increasingly complex in terms of functionality, whilst user involvement has fallen by the wayside. If there is no analysis of what users bring with them and what they want to do in the first few minutes of a meeting, a communication gap arises between the technology, the interface and the user.

In the case of shared use, there is the added issue that each group leaves the system in its own configuration. Ötting cites a typical example: the laptop is connected to the HDMI input, but the screen remains black because the source selection is still set to the previous user's configuration. Voice control addresses such situations as an additional level of operation. It accesses the same functions as the touch panel, but does not replace it. Nabbefeld classifies this approach under accessibility: relevant wherever technology must be usable without prior knowledge.

### From wake word to control command

The processing chain consists of four stages: wake-word recognition, transcription, semantic interpretation via a language model, and handover to the room control system.

In standby mode, the system analyses the microphone input exclusively for the wake word, which is 'Hey Sigma' in the reference room. To achieve this, a lightweight keyword-spotting model runs continuously on the CPU, using a ring buffer lasting a few seconds; nothing is recorded. A custom wake word can be fine-tuned using synthetic speech examples – a process taking several hours – followed by testing for false triggers caused by words that sound similar.

Once activated, a recording begins which is stored only in RAM. Its end is determined by voice activity detection via a tail-off time. This parameter is the most direct way to influence the perceived response time: a short tail-off time speeds up the system but cuts off utterances containing pauses for thought.

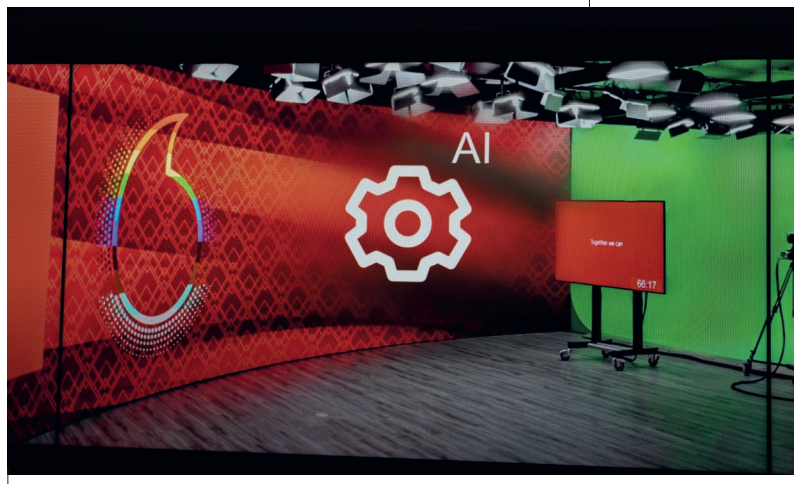
For transcription, locally executable ASR models from the Whisper family are being considered, quantised and optimised for the German language. The transcript is sent to a fine-tuned language model with a small parameter size, which searches it semantically for functions to be executed and passes the function name and parameters to the control system in a machine-readable format.

This eliminates the need for a rigid command syntax: what matters is the content of the utterance, not its phrasing. Multiple instructions can be bundled together, such as 'switch off the light above the conference table, raise the blinds and set the media input to USB-C'.

The control server responds to status enquiries – such as those regarding the volume level – with a value. ➔

## About SIGMA System Audio-Visuell

SIGMA System Audio-Visuell, based in Düsseldorf (sigma-av.com), has been planning and implementing AV integration, event technology, interactive media solutions and managed services for over 45 years. The local voice control system can be demonstrated in the Düsseldorf showroom; the company is seeking test partners to trial the system in customer environments.




The cog icon indicates that the speech model is currently analysing the utterance. Without such feedback, users would be speaking whilst the system is processing.




Demonstration room at SIGMA AV in Düsseldorf: This is where the voice control system, developed in collaboration with Düsseldorf University of Applied Sciences, is being tested.




The voice control's status indicator is superimposed over the current video content. The existing room control system, including the touch panel, remains fully functional.

A response is generated from this, voiced via a local text-to-speech (TTS) component, and output through the room's sound system. What matters here is not so much the voice quality as the latency until the first sample – A delayed confirmation is interpreted as a system error. According to the developers, it takes around two to three seconds between the end of a spoken command and its execution, depending on the hardware.
The room display shows the system status using icons for recording readiness, processing and voice output. According to Ötting, this is functionally important: without feedback, users might speak whilst the system is processing or wait in vain.

## No data path to the outside world

The speech recognition model does not require an internet connection. If the processing computer is wired directly to the control unit, a point-to-point connection is sufficient; if both are located in the same room, no data packets leave that room. Only control commands and a 'live signal' that triggers the status display are transmitted. When connected via the building network, the computer belongs to the media technology VLAN, and firewall access is restricted to the control port. For the IT department, the solution is therefore an end device on the media network, not a cloud service.

For SIGMA AV, this is the reason for developing the solution in-house rather than purchasing it. Available voice assistant systems are predominantly linked to home automation platforms and operate on a cloud-based basis. "We could simply buy a solution like that," says Nabbefeld. "But it doesn't exist in the form we need it and in the way we believe our customers need it." Sensitive data and voice recordings are at stake; one must – They can reliably ensure that users do not leave the room.

This eliminates the most common objection that AI systems face from IT departments and works councils: the transfer of data to external data centres. It is not speech recognition that sets the system apart, but the absence of an external data path. Ötting's assessment: AI does not necessarily require a data centre; small models are sufficiently powerful for room control.
During operation, the system generates log files documenting recognised and executed commands, as well as error locations – for example, during transmission to the server. From a data protection perspective, this is precisely the key point: no audio files are created; what matters is what ends up in the logs. If transcripts are recorded, personal data is generated. If logging is restricted to timestamps, function, parameters and error status, the personal reference remains minimal. Retention periods and access restrictions to IT administration must be included in the system documentation.

## For the price of a touch panel

The voice control system developed by Sigma is a purely software-based solution. The integration takes place at the same level as with a touch panel: the processing computer sends control commands to the control server, via TCP/IP socket, REST interface or serial connection, depen-

ding on the platform. The existing programming remains essentially unchanged. Depending on the existing solution, an adaptation programming may be required. If voice control is not used, the room remains fully operable via the panel.

The limitation lies in the range of functions covered: only those functions provided by the control system can be triggered. Each room function must be uniquely mapped to a function call and made known to the voice model – this is the actual configuration effort required per room type, not the installation itself. In the reference room, the range of commands comprises between 20 and around 50 functions, depending on how the parameters are counted.

Hardware sizing was the key development challenge. Processing must take place locally, the hardware must remain cost-effective, and the response time must be within a few seconds – voice control with a half-minute delay offers no advantage over simply walking over to the panel. This was resolved through additional training of the model and targeted compromises on recognition quality. As a result, a Mac mini with a current multi-core CPU and at least 16 GB of RAM is sufficient. The investment is on a par with that of an additional touch panel, not the purchase of a server.

For multiple rooms, there are two topologies to choose from. One instance per room keeps processing strictly local, but incurs hardware costs per room and requires distributed update management. One central instance per building reduces costs, but shifts processing to the building network – thereby undermining the very argument on which the solution is based. For rooms with heightened security requirements, the local instance remains the logical choice.

## The room is the deciding factor

The recognition quality of a voice control system is determined by the room, not by the model. Speech recognition operates on the same input signal as the conference transmission – with the difference that a human listener compensates for interference, whereas a recognition model does not. Ideally, therefore, ceiling-mounted microphone arrays with beamforming – as are already installed in conference rooms – are a functional prerequisite. Using noise suppression and acoustic echo cancellation (AEC), the DSP delivers the same processed signal that is also fed into the video conference.

Echo cancellation is not a convenience feature here, but a prerequisite: without AEC, the system picks up its own voice output and the audio from the other end of the line, and processes them as input. During a live conference, the quality of this signal determines whether voice control remains usable or has to be switched off.

## From the reference room to the product

SIGMA AV is developing the solution not as a bespoke product, but as a standard product for any conference room that meets the technical requirements. The internal test

The testing phase is complete. Invited customers provided feedback on the range of functions and operating logic – including how things change when the interface moves from the display to the audio channel.

Next comes testing in real customer environments. In the development environment, explains Nabbefeld, we know how the system must be addressed. Only unbiased users produce results that could not be predicted in advance.

He sees relevance wherever technology is used outside a technical environment. For the integrator, the project also serves a positioning function: a system integrator works predominantly with third-party products, but also sees itself as a consultant and developer. Voice control is a first step towards a proprietary solution that goes beyond device integration. The fact that such projects are rarely initiated is a question of attitude: the starting point is the willingness to acknowledge that a better solution exists. Citing established processes is the wrong approach. •

## Key technical specifications of the voice control system

| | |
|---|---|
| **Status** | Beta, internal testing phase completed |
| **Installations** | One demonstration room at SIGMA AV, Düsseldorf |
| **Range of commands** | 20 to around 50, depending on how they are counted |
| **Latency** | Approximately 2 to 3 seconds from the end of the utterance |
| **Language** | Standard German, expandable |
| **Activation** | Wake word, customisable |
| **Processing** | Entirely local, no internet connection required |
| **Audio data** | Stored only in RAM, no data is saved |
| **Additional hardware** | Mac mini or similar, 16 GB RAM or more |
| **Integration** | Software integrated into existing room control sys-tem; panel remains active |
| **Signal path** | Ideally, microphone via DSP with beamforming and AEC |



Emil Ötting, a Bachelor's student on the DAISY programme at Düsseldorf University of Applied Sciences, developed the voice control system as part of a practical project with SIGMA AV.