

# 2026 IEEE International Conference on Acoustics, Speech, and Signal Processing

May 4-8, 2026

Barcelona, Spain

IEEE Signal Processing Society

# Contents

Executive Overview	03
The Data Wall and Beyond	04
Audio's New Intelligence	05
Quantum Meets Classical	06
Sensing the World Smarter	07
The Human Signal	08
Mathematical Foundations Renewed	09
Forward Look	10

## EXECUTIVE OVERVIEW

# The Signal Processing Field at an Inflection Point

**55+**

Sessions Captured

**100+**

Speakers

**20+**

Countries

ICASSP 2026 gathered the global signal processing community in Barcelona for five days of intensive technical exchange spanning tutorials, keynotes, plenaries, panels, and technical sessions. The conference revealed a field undergoing profound transformation as classical signal processing theory intersects with modern machine learning, quantum computing emerges as a practical concern, and end-to-end AI systems challenge decades of engineering practice.

**Seven major themes emerged** from the sessions. The AI scaling crisis, as data sources approach exhaustion. The shift from cascaded pipelines to unified audio models. The quantum-classical convergence in computing. Location-aware and AI-native sensing systems. The richness and variability of human signals. The mathematical bridges between classical splines and neural networks. And the workforce challenge of preparing the next generation for this hybrid future.

The following pages distill insights from plenary talks by Emmanuel Candes (Stanford), Tara Sainath (Google DeepMind), Lajos Hanzo (University of Southampton), Shrikanth Narayanan (USC), and Michael Unser (EPFL), alongside technical sessions and panels exploring the practical implications of these shifts for research, industry, and education.

# The Data Wall and Beyond

Emmanuel Candes opened his plenary "What Statistics and AI Offer Each Other" with a stark warning about the limits of current AI scaling approaches. The problem is fundamental: we have exhausted the readily available training data that has fueled the explosive growth of large language models.

*"We have one internet. Pre-training as we know it will end very soon. The fossil fuel of AI, which is data, is being depleted."*

**Emmanuel Candes**  
Stanford University

The solution, Candes argued, lies in synthetic data generation. His team's work demonstrates that synthetic data, when generated using knowledge graphs (what he calls the "antigraph method"), can match the equivalent of 8x more real training data. More remarkably, when they pre-trained models on these synthesized document corpuses, they achieved 50% of the performance gain of having 20 times more real data.

*"When we actually pre-train on these new synthesized corpus of documents, we go 50% of the way towards the performance of someone who has 20 times more data than you do."*

**Emmanuel Candes**  
Stanford University

The breakthrough extends beyond pre-training. A fine-tuned model using the antigraph method exceeded GPT-4's performance on closed-book exams. But Candes emphasized a critical bottleneck: **verification becomes expensive** as these systems scale. Human verification, the current standard at major AI companies, does not scale with the pace of synthetic data generation.

*"An enormous subject of discussion at the moment is verification. Human verification is what a lot of these companies work on. It's becoming very expensive."*

**Emmanuel Candes**  
Stanford University

# Audio's New Intelligence

Tara Sainath's keynote "Audio Processing with Large Language Models" revealed how Google DeepMind's Gemini represents a paradigm shift in audio understanding. The traditional pipeline of speech-to-text, text processing, and text-to-speech is being replaced by unified models that process audio directly.

*"Unlike other large language models that rely on speech to text conversion, Gemini can process the raw audio signal directly. This allows it to capture nuances like voice tone and emotion and pronunciation that are often lost when you first transcribe to text."*

**Tara Sainath**  
Google DeepMind

The scale of deployment is remarkable. Gemini translation now covers **70+ languages and 2,000+ language pairs**. The move to end-to-end models delivers concrete performance gains beyond just preserving acoustic information.

*"If you're able to process audio to audio in one model rather than going through a cascade of first transcribing the speech through ASR, then going to an LLM to get the text answer and then going to a third TTS model, doing this in one model drastically reduces the response time."*

**Tara Sainath**  
Google DeepMind

This shift connects to broader themes in the "Speech Enhancement in the Hearables Era" tutorial. On-device processing, pre-trained models, and personalized adaptation create new possibilities for hearing aids and assistive devices. Across signal processing, complexity is being absorbed into learned representations rather than handcrafted stages.

# Quantum Meets Classical

Lajos Hanzo's plenary "Entangling classical- and quantum-domain information/signal processing" challenged the field's quantum skepticism. His central thesis: quantum and classical computing exist in symbiosis, not competition. As semiconductor density pushes toward 3-nanometer processes, quantum effects become unavoidable even in classical systems.

*"Quantum cannot survive without classical computing, without classical communications, without classical sensing and vice versa. They live in symbiosis."*

**Lajos Hanzo**  
University of Southampton

The performance gains are tangible. Quantum neural networks achieve comparable performance with **10x fewer trainable parameters**: 6,000 versus 67,000 in classical equivalents. This efficiency matters for deployment on resource-constrained hardware.

*"We are entering the quantum regime like it or not. Unless we stop really inventing more sophisticated algorithms and systems, we have to deal with the quantum effects with 3 nanometer density."*

**Lajos Hanzo**  
University of Southampton

Hanzo's message to signal processing researchers was direct and optimistic: this is not physics reserved for quantum specialists. The field is at what he called the "innovation trigger point" on the hype cycle. The tools and frameworks exist. The entry barrier is lower than assumed.

# Sensing the World Smarter

Nitin Jonathan Myers' talk "Location-aware sensing: Can Google maps help our cars see better?" introduced a counterintuitive idea: automotive sensing can be dramatically improved by using map data to guide where sensors should look. The motivation is urgent. Nearly 1 million traffic fatalities occur annually, many preventable with better sensing.

*"What we do is very different. We take data coming from Google Street Maps and use this data to guide how a radar or LiDAR should adaptively scan the environment, very similar to the human eye example."*

**Nitin Jonathan Myers**  
TU Delft

The approach mirrors human visual attention: we look where context suggests something important might happen, not uniformly across the entire field of view. By pre-loading map geometry, the system can prioritize sensor resources on occlusions, intersections, and dynamic regions.

In parallel, Soma Velayutham's keynote on AI-Native 6G wireless systems revealed massive industry investment in programmable, software-defined physical layers. **NVIDIA invested \$1 billion in Nokia** for next-generation AI-native base stations. Chris Dick from NVIDIA described the architecture: the entire physical layer uplink shared channel runs as C or CUDA code with no fixed-function accelerators. Everything is software-defined.

*"In our physical layer, it really is all C or CUDA code for the entirety of the physical uplink shared channel. There's no notion of fixed function accelerators. It generally is all software."*

**Chris Dick**  
NVIDIA

Hamid Sheikh from Samsung provided edge-device perspective. AI models in smartphone cameras grew **16x in size over 5 years**, while NPU capability increased **8.4x**. The computational demand is outpacing hardware, forcing trade-offs between model sophistication and power consumption.

# The Human Signal

Shrikanth Narayanan's plenary "Inside Out and Outside In" reframed signal processing around a fundamental question: what makes human-generated signals different? His answer: richness and variability are features, not bugs.

*"There are a lot of signals that as people we generate, share, process, use, interact with. These signals are inside us, in people, by people, for people, from people. It's such a rich universe."*

**Shrikanth Narayanan**  
University of Southern California

Narayanan's lab uses MRI to capture speech articulation at over **100 frames per second**, revealing the dynamic coordination of tongue, lips, and vocal tract during natural speech. This multimodal approach connects acoustic output to articulatory effort, opening new paths for speech therapy, language learning, and clinical diagnosis.

*"One of the core things I want to underscore is the rich variety and variability, which is actually a core feature rather than an exception in human signals."*

**Shrikanth Narayanan**  
University of Southern California

The clinical implications are profound. In studying autism diagnosis, Narayanan's team discovered that the **clinician's prosodic characteristics are highly correlated with autism severity scores in children**. This bidirectionality, the fact that the measurer affects the measured, challenges traditional signal processing assumptions of observer independence.

His closing emphasis on trustworthiness reflects growing recognition that human-centered systems require more than accuracy. They require transparency, fairness, and accountability built into the processing pipeline.

*"Trustworthy information processing is crucial. It's not an afterthought, it has to be right in the middle."*

# Mathematical Foundations Renewed

Michael Unser's plenary "Variational splines and stochastic processes" provided a mathematical bridge between classical signal processing and modern neural networks. His central observation: the outputs of spline interpolation and ReLU neural networks are visually indistinguishable for many problems. This is not coincidence.

*"The spline and the output of the neural network, they look very much the same. That is something we're trying to explain, and we will explain."*

Michael Unser  
EPFL

Unser demonstrated that training a ReLU neural network with weight decay is mathematically equivalent to solving a variational problem with total variation regularization. The ReLU activation function is precisely the Green's function of the second derivative. This connection is not metaphorical; it is exact.

*"The Wiener estimator for the Wiener process is a piecewise linear spline. I didn't make it up. It's a theorem."*

Michael Unser  
EPFL

The implications extend beyond theoretical elegance. If neural networks are implicitly solving regularized variational problems, then decades of signal processing theory on regularization, prior modeling, and convergence apply directly.

Unser's motto captures the synthesis: **"Think analog, act digital."** The continuous mathematics guides design, while discrete neural networks execute computation. This unification provides a foundation for principled network design and hybrid architectures combining learned and handcrafted components.

# Forward Look

ICASSP 2026 surfaced tensions that will define the field's trajectory. **AI safety versus speed.** Verification bottlenecks constrain synthetic data generation, but the competitive pressure to deploy at scale is immense. **Data scarcity versus scale.** The internet is finite, but model capability increases with data volume. Synthetic generation is a workaround, not a permanent solution. **Quantum readiness versus practicality.** The tools exist, but adoption lags as researchers weigh the learning curve against immediate payoff.

The "From Labs to Learners" panel explored the workforce challenge. Preparing the next generation of signal processing engineers requires curricula that balance classical theory with modern machine learning, quantum foundations, and software engineering practice. The panel consensus: students need exposure to uncertainty, messy data, and interdisciplinary collaboration, not just clean textbook problems.

Defense versus commercial AI deployment revealed a fundamental difference in error tolerance. Panel discussions on AI for defense applications emphasized that **probabilistic systems that "usually" work are insufficient when lives depend on the output.** The tolerance for false positives and false negatives differs by orders of magnitude between consumer applications and military systems.

*"Guardrails don't guarantee safety. They're just more instructions to generative models, but these are probabilistic, they don't always follow instructions."*

**Panel Discussion**

AI for Defense Applications

The conference closed with cautious optimism. The convergence of quantum computing, end-to-end neural audio models, location-aware sensing, and mathematical unifications between classical and modern methods positions signal processing at the center of transformative applications in communication, health, transportation, and defense. The challenge is execution: translating lab results into deployed systems that work reliably, ethically, and at scale.