



**Imagine
the
Future**



IBC Accelerator Programme 2026

Crystal Clear

Measuring Dialogue Intelligibility in Media

Imagine the Future

Headline Sponsor:



Innovation Partner:



Associate Sponsors:





Crystal Clear

Measuring Dialogue Intelligibility in Media



#IBCAccelerators2026

Champions



Academia



Participants





Summary

#IBCAccelerators2026

The Problem

The industry lacks a practical and standardised way to measure dialogue intelligibility. Intelligibility issues often remain undetected until audience complaints are received.

The Research

300 listeners rated real broadcast clips while objective measurements were analysed and compared against human perception.

The Innovation

A hybrid quality-control workflow combining automated analysis with structured human assessment.

The PoC

Integration of intelligibility measurements into a real broadcaster workflow to identify and review only potentially problematic programme segments.

The Vision

Make intelligibility a measurable, manageable and operational part of media quality across the industry.



The Problem Statement

#IBCAccelerators2026

- **Audiences frequently struggle to understand dialogue** in television and streamed content.
- **Existing quality control focuses mainly on audibility and technical compliance**, rather than whether speech can actually be understood.
- **EBU QC catalogue defines** only a subjective “golden ears” **audibility test**
- Subjective listening checks are **difficult to scale and** can be **inconsistent**
- The industry has **no agreed intelligibility measurement** method or accepted thresholds.

Crystal Clear establishes a **hybrid assessment framework** that combines **objective measurement** with **structured human judgement**, while **respecting creative intent**.



The Vision - Accessibility

#IBCAccelerators2026

1 in 3 people in the UK are **deaf, have hearing loss or tinnitus.***

Crystal Clear aims to make **dialogue intelligibility visible, measurable, and actionable**, helping ensure that great **content is accessible to everyone**.

- We believe Crystal Clear represents a significant step towards establishing an industry-wide framework for accessible and inclusive audio quality assessment.
- By incorporating the listening experiences of hard-of-hearing audiences into objective assessment methods, Crystal Clear lays the foundation for more accessible and inclusive audio quality standards.

*source <https://rnid.org.uk/what-we-do/data-and-evidence/deafness-and-hearing-loss-prevalence/>



The Vision - Benefits Across The Industry

#IBCAccelerators2026

Producers

- Earlier visibility of intelligibility issues.
- Objective feedback during production and post-production.
- Supports creative decisions rather than replacing them.

Broadcasters & Publishers

- Identify issues before transmission/publication.
- Reduce reliance on subjective full-programme listening tests.
- More consistent QC processes across suppliers and content types.
- Reduction in dialogue-related audience complaints.



The Vision - Benefits Across The Industry

#IBCAccelerators2026

QC Process

- Faster and more efficient workflows.
- Objective evidence available alongside professional judgement.

Technology Vendors

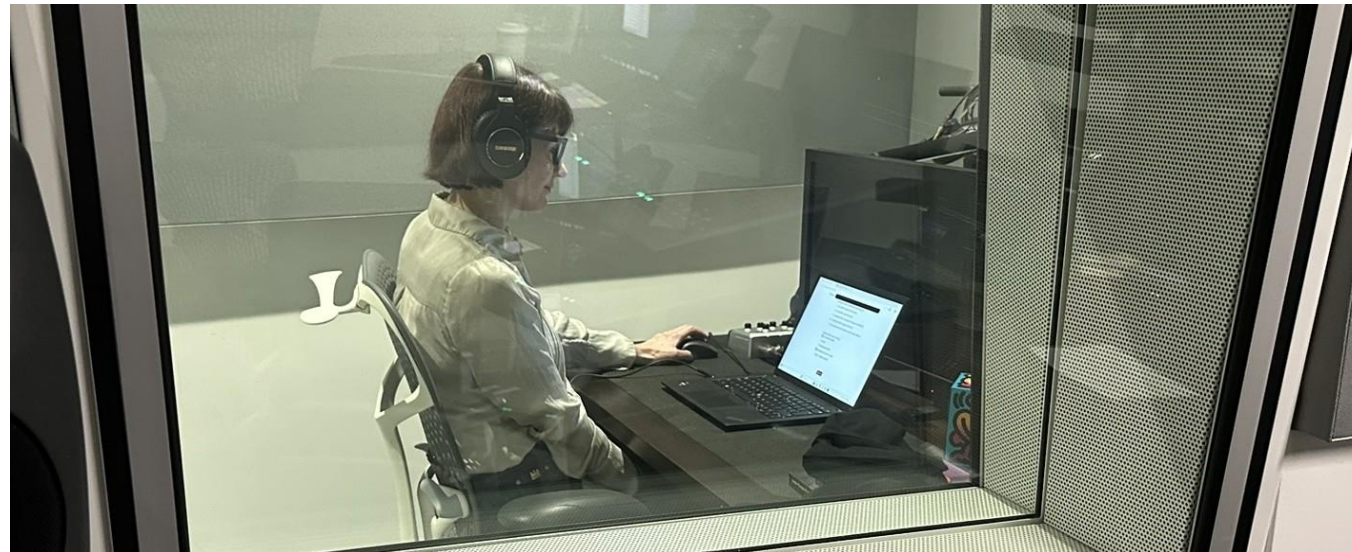
- Opportunity to integrate common intelligibility measurements into QC tools and platforms.
- Could provide interoperability between measurement systems and workflow products.



The Research

#IBCAccelerators2026

A large-scale in person and online listening study compared human perception with objective measurements using actual broadcast content. Listening effort ratings from both hard-of-hearing and normal-hearing audiences were compared against a range of machine-based measurements to identify the approaches that best reflect audience experience



The in-person study was concluded using headphones from

SHURE



The Research - Phase 1 – Preparation

#IBCAccelerators2026

Human Assessment Framework

- Definition of a simplified 5-point listening effort scale (Mappable to existing research methodologies)
- Designed for both audience studies and operational QC use
- Potential foundation for a practical human review scale within broadcaster workflows

Measurement Framework

- Review of existing intelligibility-related measurement approaches
- Evaluation of available objective measurement tools and algorithms
- Selection of candidate measures for further investigation
- Identification of measures suitable for broadcaster workflows

Research incentives were sponsored by



and





The Research - Phase 2 – Listening Experiment

#IBCAccelerators2026

Content Selection

- 60 short clips selected from actual broadcast content
- Range of intelligibility conditions (bandpass, drama, reality)
- Designed to represent realistic audience listening challenges

Human Listening Study

- 150 Normal-hearing and 150 hard-of-hearing volunteers
- Listening effort ratings collected using a simplified 5-point scale
- Volunteers used headphones or hearing devices at home or in the controlled environment

Objective Measurements

- All clips analysed using selected intelligibility-related algorithms
- Measurements compared against human listening results

Research incentives were sponsored by



and





The Research - Phase 3 – Analysis

#IBCAccelerators2026

Correlation Analysis

- Comparison of human listening effort ratings with objective measurements
- Ranked a wide range of machine models by their correlation with human listening effort ratings

Threshold Investigation

- Exploration of practical intelligibility threshold(s)
- Identification of measurement ranges associated with increased listening effort
- Assessment of differences between hard-of-hearing and normal-hearing groups

Research incentives were sponsored by

The SHURE logo, with the word 'SHURE' in green capital letters on a black rectangular background.

and

The Amazon logo, with the word 'amazon' in black lowercase letters and a curved orange arrow underneath it.



The Innovation

#IBCAccelerators2026

1. Intelligibility as a Measurable QC Parameter

- Today, broadcast QC largely measures technical compliance.
- Crystal Clear treats dialogue intelligibility as a measurable quality attribute.
- Establishes objective metrics and thresholds linked to human perception.

2. Hybrid QC model for producers and publishers QC

- Combines automated analysis with structured human assessment.
- Machines identify potentially problematic segments.
- Humans make the final editorial and quality judgement
- Does not compromise creative freedom

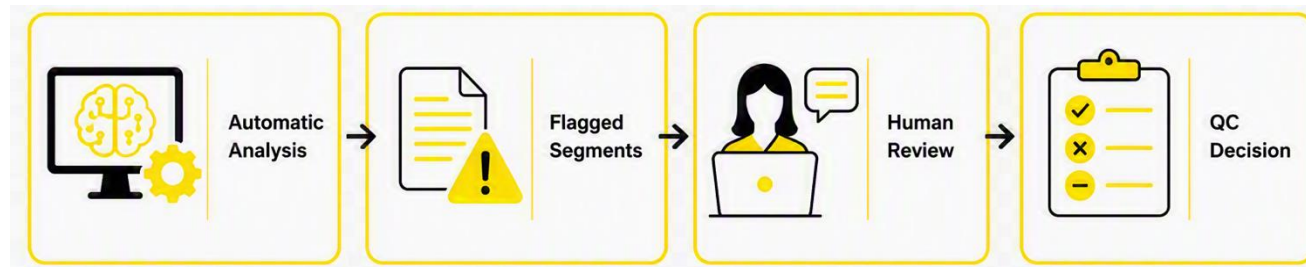


The Innovation (continued...)

#IBCAccelerators2026

3. Scalable Review Process

- Instead of reviewing an entire programme, reviewers focus only on areas identified by the system, reducing manual effort while maintaining quality.



4. Practical Workflow Integration

- Designed for integration into existing production and publishing workflows.
- QC can be applied to production process or as part of delivery.
- Supports operational use rather than remaining a research exercise.



The Proof of Concept

#IBCAccelerators2026

Integration to a real broadcaster workflow

- Lightweight integration into Channel 4's Media Services environment.
- Incoming programme content is automatically analysed after ingest
- Nugen implementation based on the Fraunhofer LEAP model.
- Measurements identify segments where dialogue intelligibility may be questionable.
- Timecoded measurement results are associated with the programme asset in Channel 4's BRAHMAAI asset management.
- Targeted review the highlighted clips



The Focus of the Project

#IBCAccelerators2026

- Crystal Clear focuses on what content producers and publishers can control: the **intelligibility of dialogue within the content itself**.
- It does not attempt to address external factors such as transmission, reproduction systems, or listening conditions
- Our research intentionally reduced the variables related to these external factors where possible -This enabled clearer conclusions to be drawn and reduced the size/duration of the required listening experiment



Crystal Clear

The Research Results



The Experiment Summary

#IBCAccelerators2026

- Dialogue "intelligibility" (widest sense) is a multi-faceted issue in broadcast media
- An improved Quality Control (QC) test that is machine-based with human-in-the-loop review is required
- 300 normal- and hard-of-hearing participants rated listening effort for 60 short clips chosen from Channel 4 and BBC dramas and documentaries
- Their ratings were correlated with dozens of potential machine-based measures of speech intelligibility (or similar, e.g. signal-to-noise ratio)
- Speech-to-text (STT) model-based measures produced the strongest correlations with listener ratings, but many other measures performed well
- Further research is required to determine a useful measurement threshold for QC, potentially using a combination of different measures



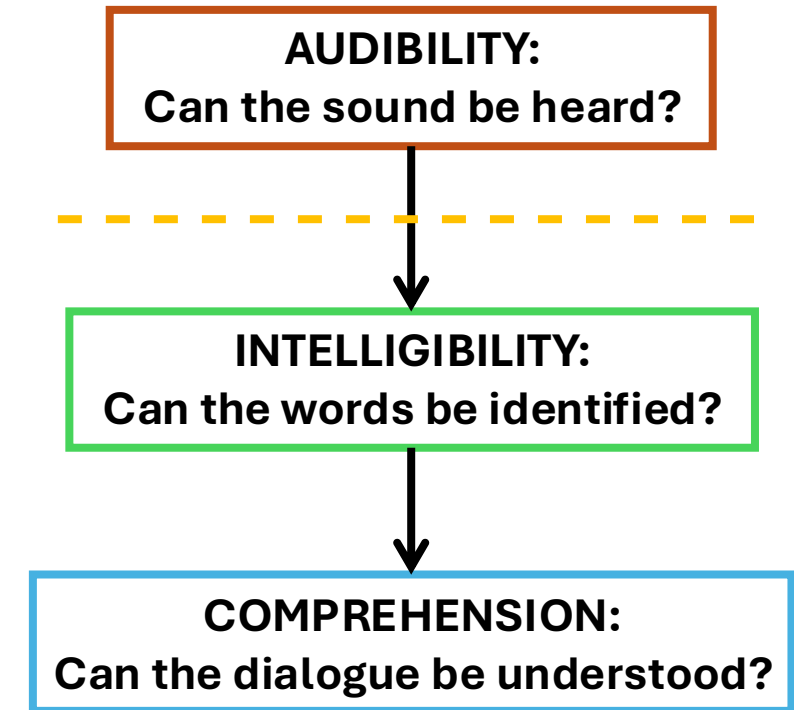
The Problem with “Intelligibility”

#IBCAccelerators2026

- **Audibility** refers to physically being able to perceive (hear) the sound.
- **Speech intelligibility** is a measure of the correct identification of words in a sentence.
- **Comprehension** refers to understanding the dialogue.

Listening effort can affect intelligibility and comprehension, but not audibility – trying harder does not make the sound louder.

“Intelligibility” has been used in broadcast media to cover all issues with dialogue. This conflates loudness (with well-established measures) with comprehension (more difficult to measure and improve).





What is Listening Effort, and Why Do We Care?

#IBCAccelerators2026

Most broadcast media content is **intelligible**, but the **effort** required to understand it can vary

Therefore, Pichora-Fuller et al. (2016) has defined **listening effort** as: "The deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a task, with listening effort applying more specifically when tasks involve listening."

Increased listening effort = more fatigue and audience complaints

Complaints = stop watching

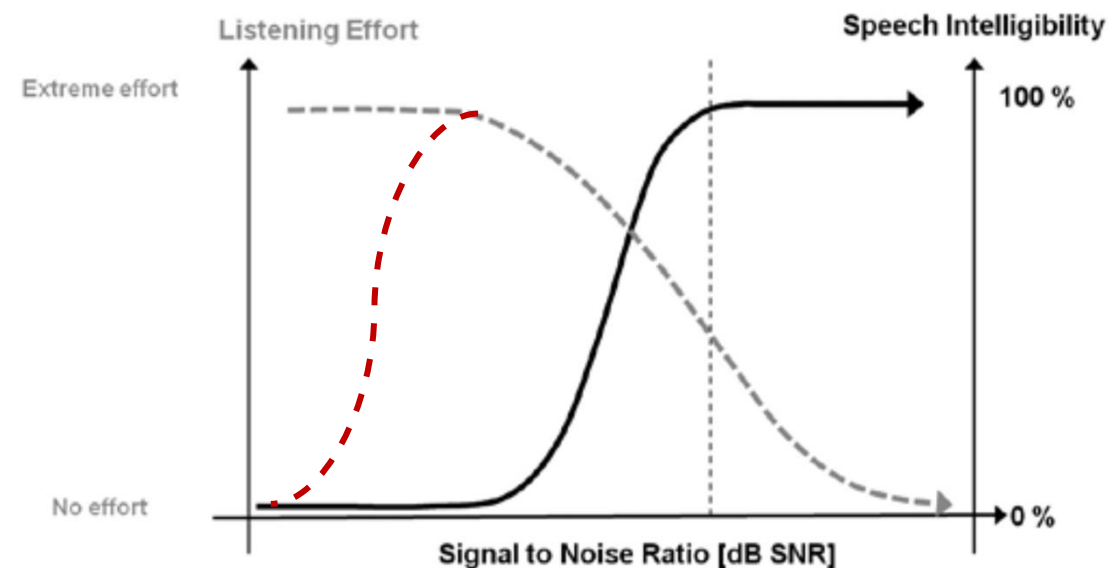


Fig. 1: Schematic illustration of the assumed relationship between speech intelligibility and listening effort.

Diagram reproduced from Klink et al. (2012) and altered, "Measuring listening effort in the field of audiology – a literature review of methods, part 1". Listeners may "give up" intelligibility drops too low (i.e. listeners "give up") (red dashed line)



Listening Effort as a Quality Control Test

#IBCAccelerators2026

- Crystal Clear aims to produce a Quality Control (QC) “traffic light” test for dialogue (Green for “good”, amber for “needs attention”, and red for “bad”)
- Intelligibility is mostly high in media
- Listening effort is considered a more sensitive listener/audience measure
- Therefore, the research focused mainly on listening effort for QC

Last broadcaster
“checkpoint” before
distribution

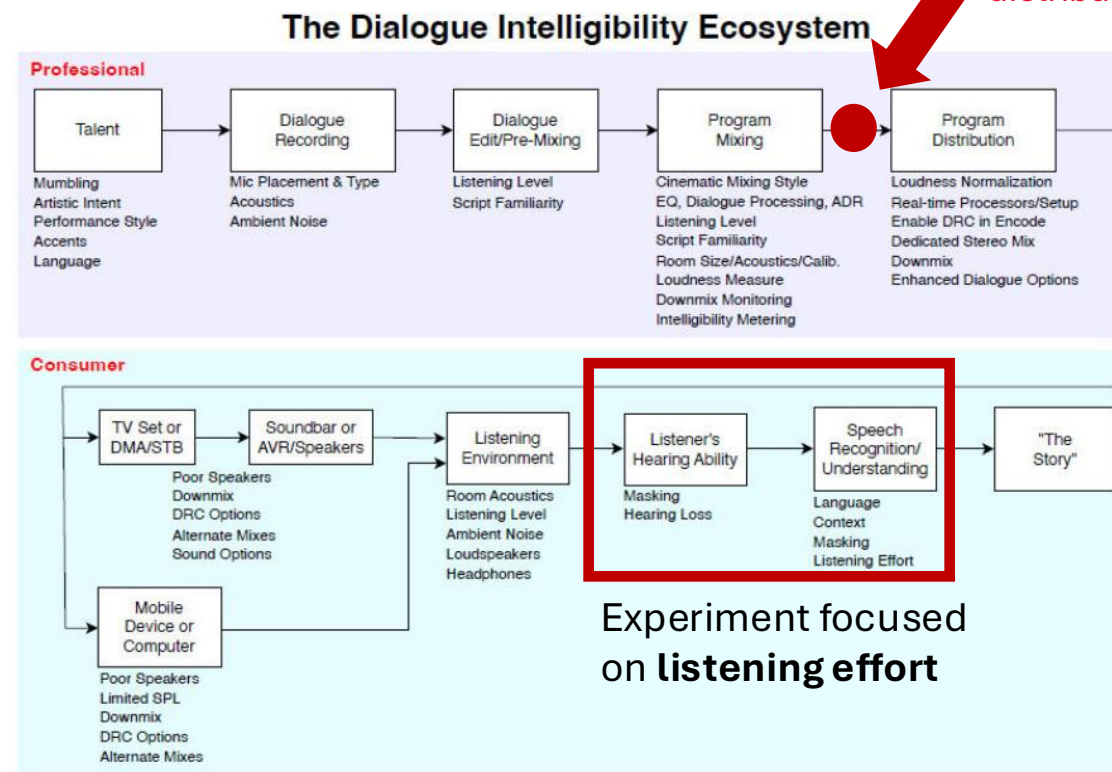


Figure 1. The Dialogue Intelligibility Ecosystem

Diagram reproduced from AES report TD1009 “Improving Dialogue Intelligibility in Media”, 2025



Requirements for a QC measure

#IBCAccelerators2026

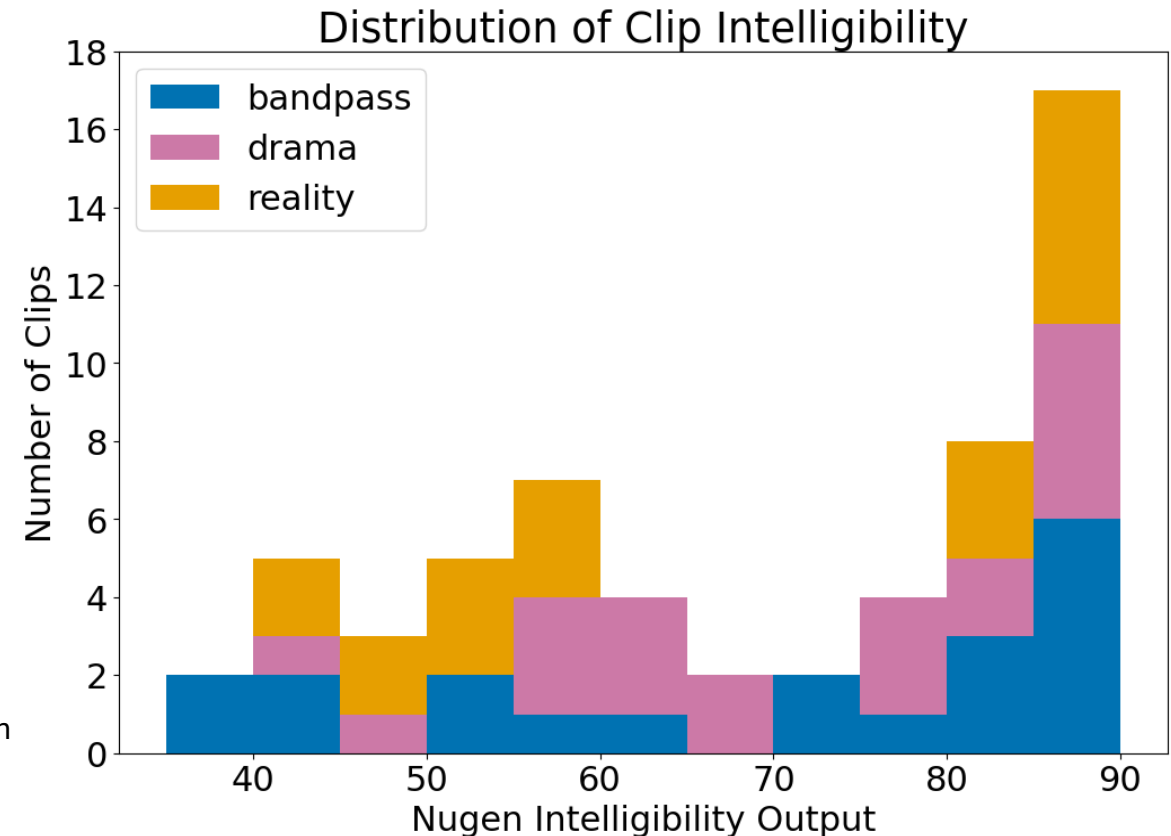
- Objective (e.g. automated, machine-based detection with human-in-the-loop review)
- Non-intrusive (e.g. no dialogue-only reference available at distribution point)
- Real-time or offline
- Consider computational costs
- Shown to correlate well with listening effort for both Hard-of-Hearing (HoH) and Normal Hearing (NH) listeners using mixes taken from broadcast media (e.g. not artificially created)



Listening Experiment – Clips

#IBCAccelerators2026

- Used the Nugen *DialogCheck* plug-in to measure hundreds of clips, mostly from Channel 4 programmes
- All but one had received complaints about the audio
- Genres like sports and feature films not included due to copyright
- Subset of 60 short (avg. 8 sec) clips selected manually:
 - 30 x clips (Nugen output < 75)
 - 30 x clips (Nugen output > 75)
 - Using custom DialogCheck settings and based on previous research
- A split distribution (at 75) was used to give the best chance of finding a clear QC threshold





Listening Experiment – Clip Selection

#IBCAccelerators2026

- All clips were normalized to -23 LUFS to reduce loudness offset.
Signal-to-noise ratios (SNRs) are unchanged
- Three loosely defined conditions (<75 only) identified by manual review:
 - Bandpass** – low-bandwidth speech (e.g., archive audio from early 60's)
 - Reality/Documentary** – difficult recording environments (e.g., hospitals)
 - Drama** – emotional or whispered speech
- The >75 clips were chosen from the same programmes as the <75 clips
This partially controlled for the effect of programme
A successful QC measure should identify issues within a programme



Listening Experiment – Machine Models

#IBCAccelerators2026

We ranked various models by their correlation to human ratings of listening effort.

- **A version of the LEAP model of listening effort (Fraunhofer IDMT)**

Incorporated into Nugen (as DialogCheck), Telos Alliance, RTW, Steinberg (Nuendo), and Junger audio products

- **Automatic speech recognition (ASR) confidence values Speech-to-text (STT) models**

e.g. Whisper (Open AI) & wav2vec2 (Meta) - Used in previous research (e.g. Clarity Challenges, claritychallenge.org)

- **Many other non-intrusive objective measures**

e.g. non-intrusive STOI & PESQ, NISQA, SQAST, SQUIM, etc)



Listening Experiment – Procedure

#IBCAccelerators2026

Progress

- 300 (150 hard-of-hearing, HoH) participants took part in an online experiment
- Normal-hearing (NH) participants used headphones
- Hard-of-hearing used their usual playback device (e.g. bluetooth-paired hearing aid, device speakers)
- Participants rated the listening effort required for each clip, and why it was difficult (optional)

How easy was it to understand the dialogue?

- ☐ 5. No effort required (Very easy)
- ☐ 4. Little effort required (Easy)
- ☐ 3. Moderate effort required (Not easy, not difficult)
- ☐ 2. Maximum effort required (Difficult)
- ☐ 1. Not understood with maximum effort (Very difficult)

(Optional) Why was it difficult?

- ☐ Low speech quality
- ☐ Accent
- ☐ Overlapping talkers
- ☐ Background sound or music
- ☐ Other - please specify:

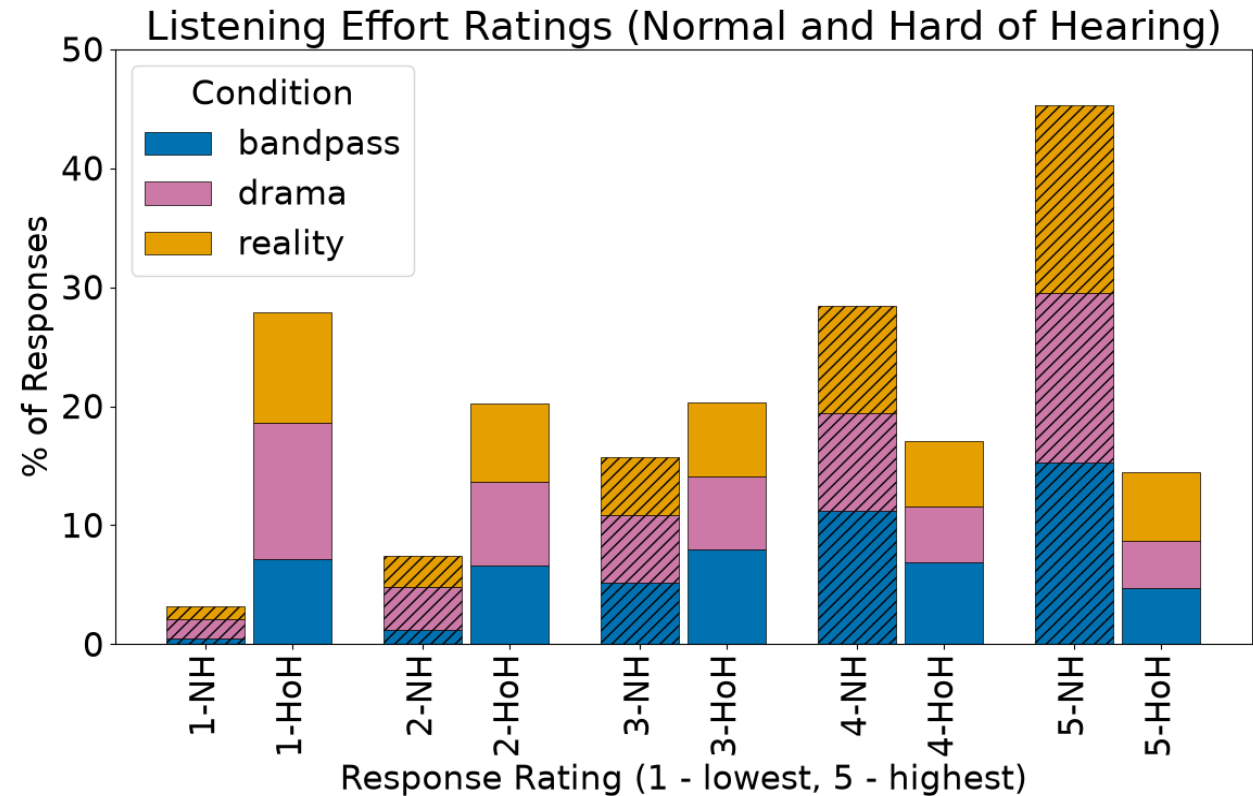
Next Clip



Listening Experiment – Results

#IBCAccelerators2026

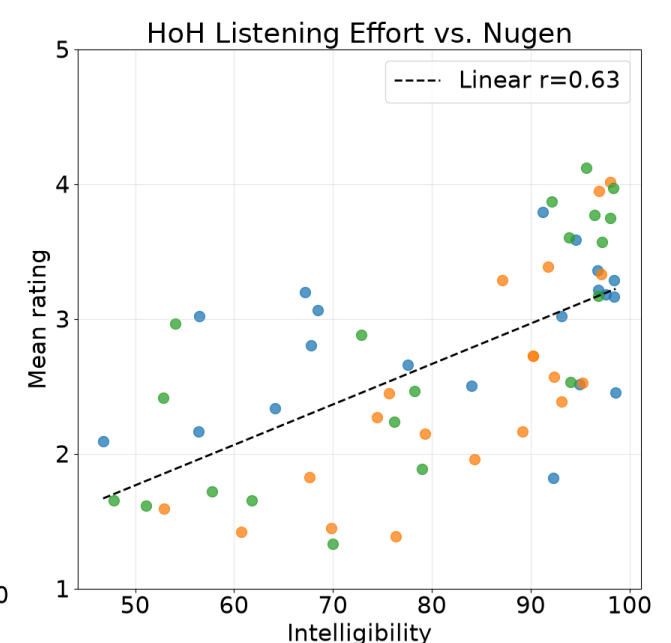
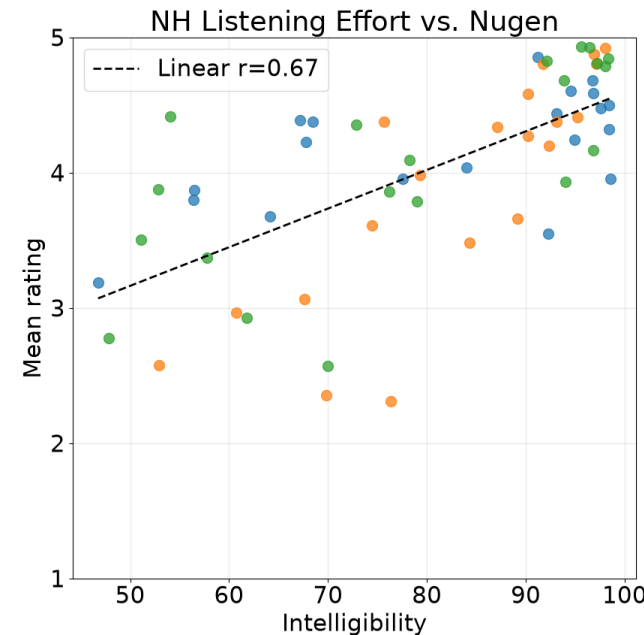
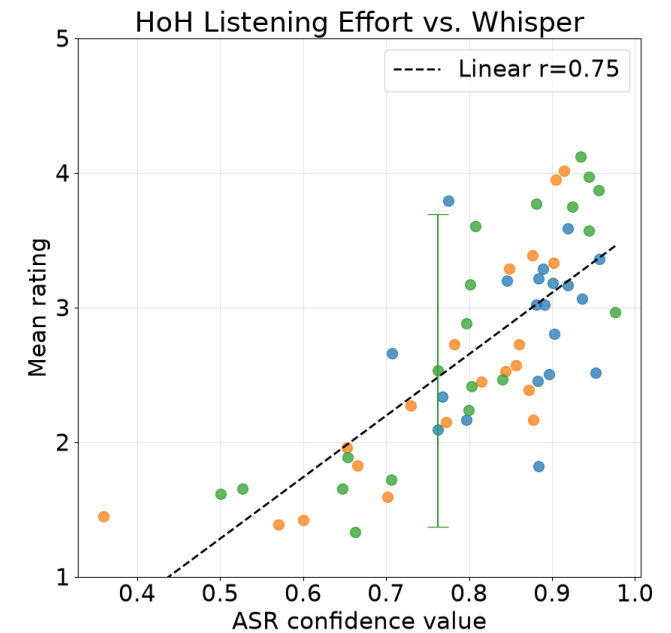
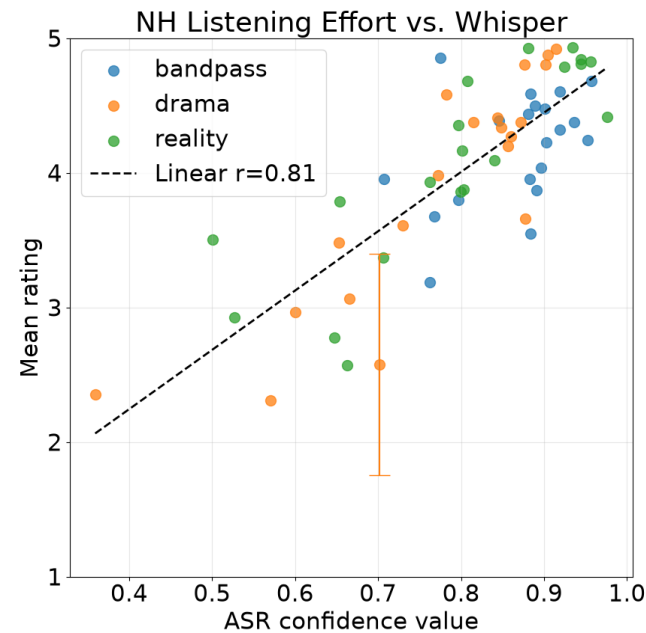
- Hard of Hearing (HoH) listeners rated clips ~1.5 pts more difficult than normal hearing (NH)
- Most NH ratings 4 or above
- Most HoH ratings 3 or below
- Ratings by condition were evenly spread





Correlations

- HoH listeners rated clips 1.5 points (on listening effort scale) lower than NH listeners
- Whisper correlated more strongly than Fraunhofer/Nugen
- It is difficult to define a specific threshold for “poor” clips using either measure
- High objective values do not guarantee low listening effort
- But low values can be used to highlight audio for additional human review



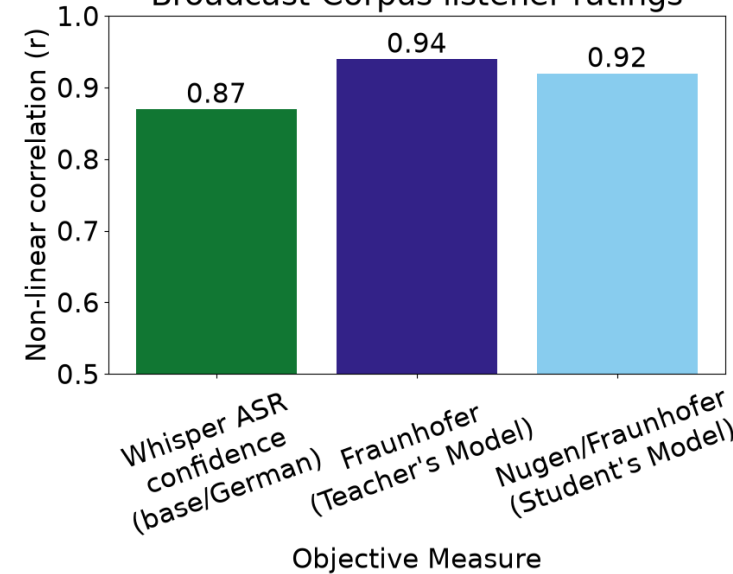
NH and HoH (columns) listener ratings (y-axis) against Nugen and Whisper (rows) ASR confidence measurements (x-axis)



Correlation Comparison

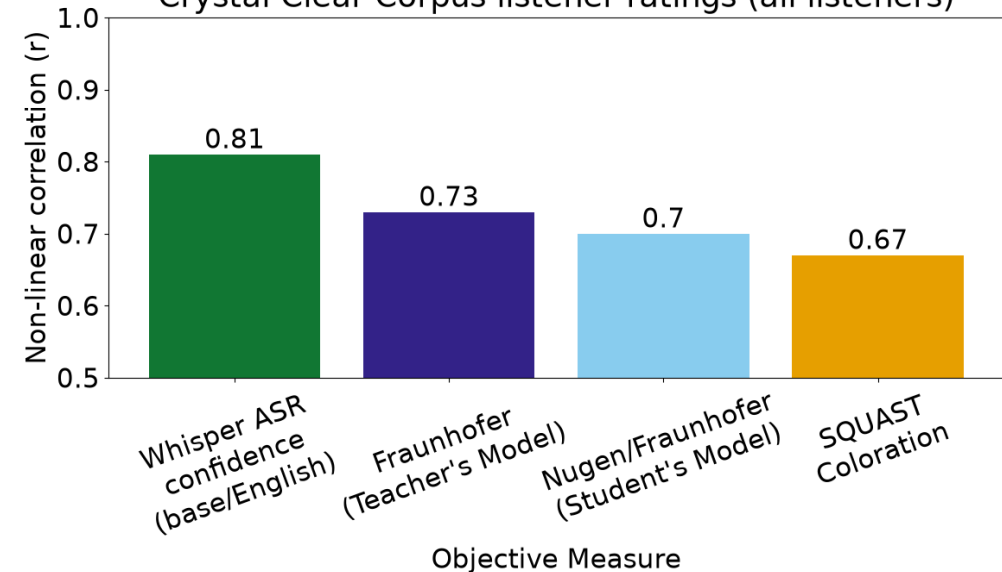
- ~100 measures, correlations = ~0.1 - 0.85
- Speech-to-text (STT) ASR confidence values correlated best with Crystal Clear corpus (**60 English clips**)
- Fraunhofer's models correlated best using a broadcast corpus (**400 German clips**)
- However, STT models are:
 - larger (~2 orders magnitude more parameters)
 - trained on much larger (often unknown) corpuses

Objective measure correlations with Fraunhofer Broadcast Corpus listener ratings



#IBCAccelerators2026

Objective measure correlations with Crystal Clear Corpus listener ratings (all listeners)





Important Considerations

#IBCAccelerators2026

- Crucially, DialogCheck is clearly modelling the perception of phonemes
- STT automatic speech recognition (ASR) confidence values may not have a perceptual basis
- The EU AI act may prevent the use of the Whisper model and others like it

Model	Training data (speech)	Data source	Model size (parameters)
DialogCheck	~8k hrs. word-labeled	Purchased dataset	458.3k
Whisper (Tiny)	~680k hours labeled	Unknown	39 million
wav2vec2 (base)	~53k hours unlabeled	Libri-Light (LibriVox)	~95 million



Conclusions

#IBCAccelerators2026

- Listening effort ratings are higher overall for hard-of-hearing than normal-hearing listeners
- The ratings correlate moderately with several objective measures, including
- Fraunhofer IDMT/Nugen and speech-to-text (STT) based models
- It is difficult to set clear “traffic light” thresholds based on these results

The best correlations were moderate-to-strong

This was partially due to a lack of very difficult audio clips in the corpus, which is representative of broadcast media more widely

- There is potential to use objective measures as an automatic “flagging” system for poor dialogue, but only with subsequent human review



Crystal Clear

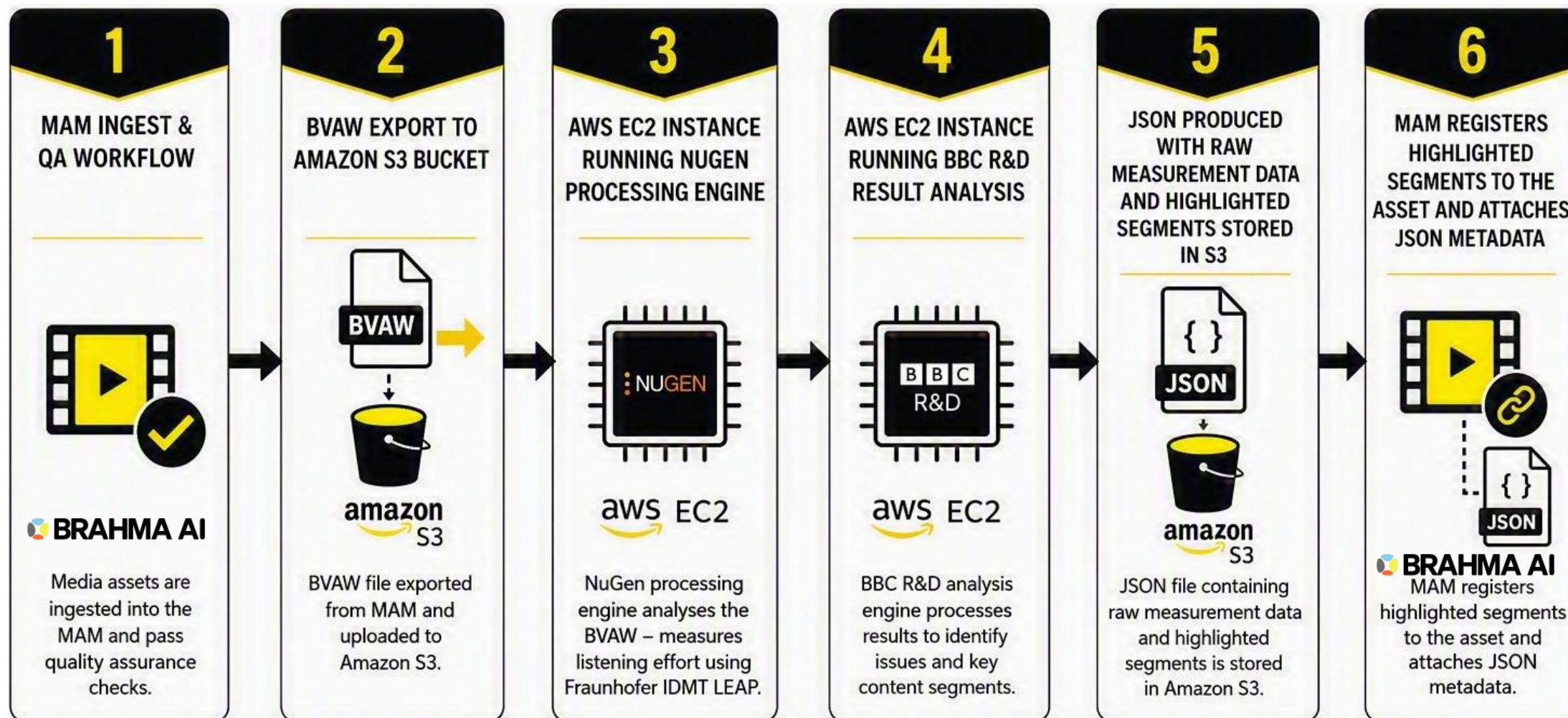
The PoC Results





The POC results - Integration Design

#IBCAccelerators2026

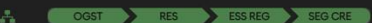
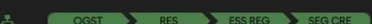
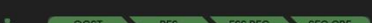




The POC results - Workflow dashboard in BRAHMA AI #IBCAccelerators2026

Process > Dialogue Intelligibility GC

Dashboard Workorders

	DUE DATE	TITLE	CREATED ON	
PB	2026 09 Sep 09:25	 Intelligibility_35765/002_5 Variant Id : 35765/002_5 RESTART ABORT	08 Sep 2026 09:25	
PB	2026 09 Sep 09:10	 Intelligibility_76930/111_1 Variant Id : 76930/111_1 RESTART ABORT	08 Sep 2026 09:10	
PB	2026 09 Sep 08:58	 Intelligibility_35765/005_5 Variant Id : 35765/005_5 RESTART ABORT	08 Sep 2026 08:58	
PB	2026 09 Sep 08:29	 Intelligibility_77002/118_1 Variant Id : 77002/118_1 RESTART ABORT	08 Sep 2026 08:29	
PB	2026 09 Sep 07:27	 Intelligibility_40591/007_2 Variant Id : 40591/007_2 RESTART ABORT	08 Sep 2026 07:27	



The POC results - Asset view in BRAHMA AI

#IBCAccelerators2026

GOGGLEBOX SERIES 28
Sep 04, 2026

Information Details Segments Essence

41-100 of 182

RecordingSegments Intelligibility

☐		low-60.93 Dur: 00:00:15:04 13:10:58:12 - 13:11:13:15
☐		marginal-72.38 Dur: 00:00:04:22 13:11:18:13 - 13:11:23:09
☐		marginal-72.69 Dur: 00:00:02:07 13:11:27:04 - 13:11:29:10
☐		marginal-87.35 Dur: 00:00:03:01 13:11:35:21 - 13:11:38:21
☐		low-64.41 Dur: 00:00:26:14 13:11:51:12 - 13:12:18:00
☐		marginal-63.46 Dur: 00:00:02:08 13:12:22:13 - 13:12:24:20
☐		marginal-72.42 Dur: 00:00:11:20 13:12:26:22 - 13:12:38:16
☐		low-60.52 Dur: 00:00:17:04 13:12:42:18 - 13:12:59:21
☐		marginal-76.56



The POC results - Areas Investigated

#IBCAccelerators2026

Segment Length

- Segment duration should be user-filterable
- Reviewers can understand flagged issues, but pre-roll and post-roll would improve context
- Approximately 5 seconds of pre-roll and post-roll is practical for operational QC

Threshold Application

- Proposed thresholds identify relevant content within the limitations of the models used
- Too many segments are currently highlighted, requiring UI-based filtering
- A single threshold may not work consistently across all programme genres and content types

Workflow Suitability

- The approach can be integrated into existing broadcaster workflows
- It has the potential to provide useful guidance without unnecessary operational overhead
- The current POC implementation requires further refinement



Crystal Clear

The Achievements of the Project



The Achievements of the Project

#IBCAccelerators2026

1. Subjective evidence collected

Collected listening-effort assessments from 300 participants, including people with normal hearing and hearing loss, using 60 broadcast clips

2. Objective measurement models assessed

Evaluated candidate measurement approaches against the subjective listening data to understand how effectively automated models reflect audience experience.

3. Broadcast content measurement pipeline developed

Created an automated process for analysing programme audio, retaining raw measurements and identifying potentially problematic time-coded segments.



The Achievements of the Project

#IBCAccelerators2026

4. Proof of concept implemented

Integrated automated intelligibility analysis into a broadcaster's media asset management and quality-assurance workflow, allowing highlighted segments and measurement metadata to be registered against the asset.

5. Hybrid human-in-the-loop workflow

Combined machine-based analysis with targeted human review, enabling reviewers to focus on system-flagged segments rather than listening to complete programmes.

6. Foundation created for industry alignment

Established research evidence, an operational PoC and shared recommendations that could support future intelligibility thresholds, delivery specifications and wider standardisation.

Crystal Clear has moved dialogue intelligibility from a largely subjective concern towards a measurable, testable and operationally manageable component of media quality.



Crystal Clear

The Team



The Team

#IBCAccelerators2026

Channel 4: Balazs Sari, Robert James, Manju Sundaram, Dr. Cara Hirst, Dr. Sue Farrell, Jesse Milverton, Steve Belford, Andy Bell, Maheen Choonara, Kirsty Walker, Jonathan Penny, Holly Fairweather, Fiona Lomax, Tamara Ballard, Roshan Alex, Ian Barrett, Lucy Antoniou, Rebecca Hollister, Andrew James

BBC: Dr. Alan Archer-Boyd, Andrew Dunne, Dr. Matt Paradis

Fraunhofer IDMT: Johanna Baumgartner-Krone, Prof. Dr. Jan Rennies-Hochmuth, Christian Rollwage, Dr. Rainer Huber

Sky: Ashley Ross

IBC: Keran Boyd, Octavia Lily

IET: Ian Nock, Ted Laverty, Sean Jones

AES: Richard Friedel, David Bialik, Robert Bleidt, Dr. Scott Norcross

Amazon: Dr. Scott Isabelle

Shure: Ian Davis, Jack Drury

RNID: Dr. Lauren Ward, Laura Marples

ITV: Matt Simpson

Nugen: Dr. Paul Tapper, Charles Blessing

University of Salford: Dr. Ben Shirley

BRAHMA AI: Madhankumar Thangavel, Atul Saxena, James Macmillan, Kavipriya, Diwakar Reddy, Sushma L, Jinoy Jairaj

Telos Alliance: Kirsten Smith, Markus Hintz, Graham Tudball, Costa Nikols



Special Thanks To:

#IBCAccelerators2026



University of
Salford
MANCHESTER





Link the AES paper: Improving Dialogue Intelligibility in Media





**Thank
You**