Broadcast Language Identification System (BLIS)

James Connolly[1,2], Kevin Curran[2], Paul Mc Kevitt[1], John Macrae[3], Stephen Craig[4]
[1]School of Arts and Creative Technologies, University of Ulster, Magee
[2]Intelligent Systems Research Centre (ISRC), University of Ulster, Magee
[3]Office of Innovation, University of Ulster, Jordanstown
[4]Maiden Technology Ltd., Larne, Co. Antrim
{jp.connolly, kj.curran, p.mckevitt, j.macrae}@ulster.ac.uk
stephen@maiden.org.uk

Abstract

Virtually all TV transmission systems use highly automated file-based broadcast systems for audio and video. Content management systems automatically deliver programmes ready for broadcast by matching content to a schedule and deliver all ancillary services in their correct format and on time. Within these sub-systems is a growing need to validate that the correct language is delivered to a particular service and/or region. In many cases, a single instance of a programme exists and the automated system merely selects the correct language for a particular service. This process is currently managed manually by operators listening to the audio of each programme and confirming that the accompanying language is correct for its video broadcast. Incorrect language transmission can be caused by system faults or errors in the scheduling workflow. An error can occur at numerous points during the broadcast. The Broadcast Language Identification System (BLIS) will provide a single operator with the ability to monitor multiple services by "dash-boarding" language flags from each service and enable the operator to intervene if an error is detected. BLIS will examine streaming audio from a pre-broadcast to identify spoken language within the broadcast content and compare it with the expected language of the video for broadcast.

Keywords: Audio, Language Identification, Automatic Speech Recognition, Television Broadcast, Broadcast Language Identification System (BLIS)

1. Introduction

In a file-based modern highly-automated transmission environment, unintentional errors can produce mismatches between transmitted video and audio, resulting in a reduction in broadcast Quality of Service (QoS). The common broadcasting technical standards document [1] agreed by the BBC, BskyB, Channel 4, Channel 5, ITV and S4C includes details on video and audio production formats. All audio is encoded within the Pulse Code Modulation (PCM) standard [2] and must have an audio sampling frequency of 48 kHz, 24 bit audio depth. Although such documentation provides audio structure standards, there are several stages within audio integration and delivery processes where errors can occur. Programmes can be broadcast in standard definition (SD) or High Definition (HD) and may be delivered for broadcast in file or tape format. Audio track layout and allocations differ for both platforms and can exist in 4 or 16 track layout. Programmes that contain single language tracks, or are in SD format use 4 track audio. The first 2 tracks of SD and HD formats contain left and right final mix sound. Third and fourth tracks may contain music and effects (M&E), audio description or digital silence. With HD audio, additional tracks provide 5.1 surround sound. Remaining tracks may contain 2 or 3 additional languages. Additional audio tracks can be independently delivered in Broadcast Wave Format (BWAV) [3]. For tape broadcast format, a supplementary language can be allocated to the third and fourth track, or to track 11 and 12 for HD audio. Remaining tracks contain surround sound and M&E. Furthermore, standards for live broadcasts differ between broadcasters, although all are working towards a standardised audio layout.

Metadata contains all information relevant to a file or tape broadcast. It ensures all video and audio content is correctly reconstructed for playback or for various system conversions. Structural metadata is manually added by a broadcast producer and includes title and ID number for the programme and structure of associated audio tracks. Errors can be

introduced to the system at several stages. Audio track layout may be accidently mismatched to the definition standard of the accompanying programme. Incorrect use of file, tape or live audio format can cause misalignment between intended and expected broadcast language. Metadata information can contain incorrect reference to primary, secondary and tertiary audio language channels. Section 2 discusses our proposed solution to these problems using BLIS and section 3 concludes with future work.

2. Broadcast Language Identification System (BLIS)

An operator manually examines audio for each broadcast to ensure both are correctly matched. Typically an operator is responsible for correct identification of 8 simultaneous broadcast channels. Broadcast Language Identification System (BLIS) will provide a single operator with the ability to monitor multiple services by "dash-boarding" language flags from each service and enable the operator to intervene if an error is detected. BLIS will examine streaming audio from pre-broadcasts to identify spoken language content and compare it with the expected language of the video for broadcast. BLIS will exist as two units. The first unit will provide front-end functionality to the local broadcast operator through dash-boarding software as shown in Fig. 1(a). The dash-board will deliver error feedback to the operator if a mismatch between pre-broadcast audio and video is identified, and an instant view on the status of programmes throughout each broadcast. Training for new language dictionaries will improve language detection and accuracy. Software functionality may be integrated with existing broadcast software systems that have the capability to call third party functions that exist on the broadcast network. These functions will reside on a broadcast network server.

The second unit will provide back-office functionality to the dash-boarding software as shown in Fig. 1(b). It will reside on the broadcast network on a Windows Server system (Fig. 1(c)). This server may exist within the broadcast network, or function as a separate sub-domain. Communication using REST technology will respond to language query services of the local software system to identify spoken language of a pre-broadcast. The server-based system will interact with cloud-based language models if the local system cannot uniquely identify the spoken language of a pre-broadcast audio stream. The cloud based language model will update locally stored language models.

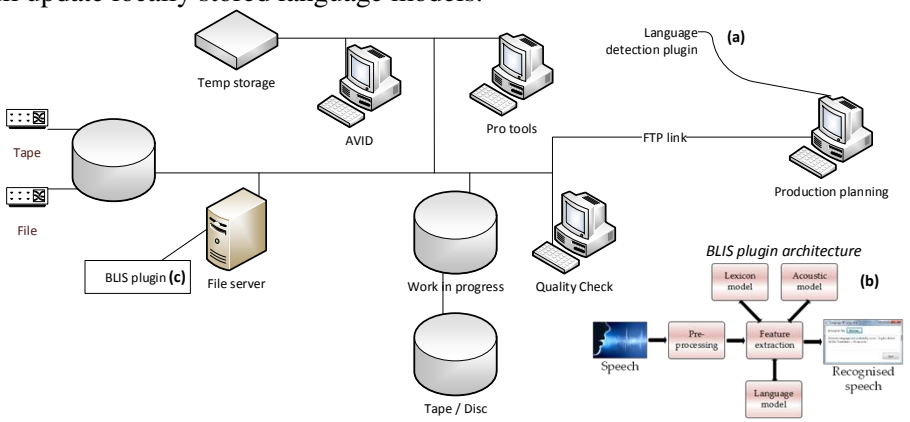

Figure 1: Architecture of typical broadcast system.

3. Conclusion & future work

TV transmission networks use highly automated file-based broadcast systems for audio and video. Human error can unintentionally introduce audio and video mismatches to the system. Manual techniques are currently used for problem identification. BLIS is an automatic language identification system that replaces manual intervention with dash-boarding software. Multiple channels can be automatically monitored and the broadcast operator can be immediately notified of potential problems.

Acknowledgements

BLIS is part financed by the European Regional Development Fund (ERDF) under the EU Sustainable Competitiveness Programme for Northern Ireland 2007-2013.

References

[1] BBC, "Technical Standards for delivery of television programmes to BBC." pp. 1–41, 2012.
[2] W. M. Waggener, *Pulse Code Modulation Techniques*, 1st ed. New York: Thompson Publishing Inc, 1995.
[3] EBU - UER, "Specification of the Broadcast Wave Format (BWF)." Geneva, pp. 1–20, 2011.