
ITU-T STUDY GROUP 16 (Porto Seguro, Brazil, May 28-June 8, 2001)

QUESTIONS: ITU-T 15/16; SG 12, SG 15; IETF, ETSI/Aurora, TIPHON & 3GPP, T1A1, TIA(3GPP2)
SOURCE: ITU-T SG 16
TITLE: New SG 16 work on Distributed Speech Recognition (DSR) and Distributed Speaker Verification (DSV)

LIAISON STATEMENT | COMMUNICATION

FROM: WP 3/16
To: LS: ITU-T SG 12, SG 15
Communication: IETF; ETSI/Aurora, TIPHON & 3GPP;
T1A1, TIA (3GPP2)
APPROVAL: SG 16
FOR: Action
DEADLINE: November 12, 2001
CONTACT:
Q.E/16 Simão Campos-Neto (acting)
LMGT
22300 Comsat Drive
Clarksburg, MD 20871
USA
Tel: +1-301-428-4516
Fax: +1-301-428-9287
Email: simao.campos@itu.int

At its May – June 2001 meeting, SG16 discussed Delayed contributions D.142 and D.143 identifying an urgent need to open a new Question on Distributed Speech Recognition (DSR). It was agreed to open a new Question and to increase the scope of the work to include Distributed Speaker Verification (DSV).

Attached, for your information and comments, please find the text of the new Question 15/16. We would like to invite your comments on the new work item, in particular before the November 2001.

Attachment:

(A) Text of new Q.15/16 approved by SG 16

Question 15/16 - Distributed Speech Recognition (DSR) and Distributed Speaker Verification (DSV)

Background and justification

Speech recognition systems are being deployed in commercial applications today, where the whole speech recognition system is typically implemented in a central place to which all speech signals are routed.

In addition to speech recognition, speaker verification plays an important role as a biometric verification mechanism, as recognized in the IP Networking and Mediacom 2004 Workshop (Geneva, April 24-27, 2001).

Speech recognition and speaker verification systems need to perform a set of operations, such as signal pre-processing, some sort of front-end extraction of features or parameters, back-end processing, and higher layer control according to the constraints of the application.

With voice communication over packet based digital networks, such as Voice-over-IP, becoming popular, elements sitting on the edge of the packet network are becoming more capable of accomplishing complex signal processing tasks, such as speech encoding and decoding. With this evolution, there is an opportunity to enhance the performance and efficiency of speech recognition and speaker verification systems by moving some of the basic speech signal processing tasks to the edge of the packet network.

Components of a speech recognition or speaker verification system can be distributed between an edge element (such as a router, gateway or IP telephone) and a remote application server in a flexible manner. For example, the front-end may be implemented on a gateway and the back-end on an application server. In this example, a gateway processor would perform pre-processing and feature-extraction for speech recognition or speaker verification purposes. The features would be compressed, packetized and sent to a speech recognition/speaker verification application server. In turn, the server would perform the back-end processing and take the appropriate action. Alternatively, a portion of the front end such as the speech end-pointer may be implemented on a gateway with the feature extraction and back end being implemented on a server.

One of the key issues to be resolved if Distributed Speech Recognition (DSR) and Distributed Speaker Verification (DSV) are to become successful is interoperability between system components at the edge of the packet network and those on the server, where the edge element and server are produced by different vendors. This is where standardization is critical.

This question will study which standards for DSR and DSV should be adopted for use over packet-based digital networks, such as IP or ATM networks.

Study items

- Develop the overall system architecture for Distributed Speech Recognition (DSR) and Distributed Speaker Verification (DSV) systems.
- Determine which sets of features are appropriate for DSR and DSV purposes, taking into consideration that the back-end processing should be left as open as possible to allow for improvements in the technologies.
- Study aspects of the front-end processing and feature extraction that should be standardized to ensure interoperability between front-end and back-end components of DSR and DSV systems.
- Define the signalling requirements for communication among front-end, back-end, and any intermediate processing elements of DSR and DSV systems, and develop a mechanism for negotiating capabilities between these elements and selecting a mode of operation.
- Define the protocol requirements for transport of the extracted information over packet based digital networks, and either identify an existing or develop a new transport protocol.
- Consider interoperability issues with existing systems (examples: ETSI AURORA and proprietary systems).

Specific tasks with expected time-frame of completion

This question will study the issues identified above and produce relevant standards for DSR and DSV systems: late 2002.

Relationships

- Other relevant Questions within Study Group 16 (including Q.B, Q.5, Q.2, and Q.3)
- ITU-T Study Group 12 on end-to-end performance issues
- ITU-T Study Group 15 on transmission equipment issues
- ETSI Aurora and TIPHON

- Committee T1
- IETF
- 3GPP, 3GPP-2

END