Network Working Group
Request for Comments: 4060
Category: Standards Track

Q. Xie
D. Pearce
Motorola

May 2005

RTP Payload Formats for European Telecommunications Standards Institute (ETSI) European Standard ES 202 050, ES 202 211, and ES 202 212 Distributed Speech Recognition Encoding

Status of This Memo

This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Copyright Notice

Abstract

This document specifies RTP payload formats for encapsulating European Telecommunications Standards Institute (ETSI) European Standard ES 202 050 DSR Advanced Front-end (AFE), ES 202 211 DSR Extended Front-end (XFE), and ES 202 212 DSR Extended Advanced Front-end (XAFE) signal processing feature streams for distributed speech recognition (DSR) systems.

Table of Contents

1. Introduction

Distributed speech recognition (DSR) technology is intended for a remote device acting as a thin client (a.k.a. the front-end) to communicate with a speech recognition server (a.k.a. a speech engine), over a network connection to obtain speech recognition services. More details on DSR over Internet can be found in RFC 3557 [10].

To achieve interoperability with different client devices and speech engines, the first ETSI standard DSR front-end ES 201 108 was published in early 2000 [11]. An RTP packetization for ES 201 108 frames is defined in RFC 3557 [10] by IETF.

In ES 202 050 [1], ETSI issues another standard for an Advanced DSR front-end that provides substantially improved recognition performance when background noise is present. The codecs in ES 202

050 use a slightly different frame format from that of ES 201 108 and thus the two do not inter-operate with each other.

The RTP packetization for ES 202 050 front-end defined in this document uses the same RTP packet format layout as that defined in RFC 3557 [10]. The differences are in the DSR codec frame bit definition and the payload type MIME registration.

The two further standards, ES 202 211 and ES 202 212, provide extensions to each of the DSR front-end standards. The extensions allow the speech waveform to be reconstructed for human audition and can also be used to improve recognition performance for tonal languages. This is done by sending additional pitch and voicing information for each frame along with the recognition features.

The RTP packet format for these extended standards is also defined in this document.

It is worthwhile to note that the performance of most speech recognizers are extremely sensitive to consecutive frame losses and DSR speech recognizers are no exception. If a DSR over RTP session is expected to endure high packet loss ratio between the front-end and the speech engine, one should consider limiting the maximum number of DSR frames allowed in a packet, or employing other loss management techniques, such as FEC or interleaving, to minimize the chance of losing consecutive frames.

1.1. Conventions and Acronyms

The keywords MUST, MUST NOT, REQUIRED, SHALL, SHALL NOT, SHOULD, SHOULD NOT, RECOMMENDED, NOT RECOMMENDED, MAY, and OPTIONAL, when they appear in this document, are to be interpreted as described in RFC 2119 [4].

The following acronyms are used in this document:

```
DSR  - Distributed Speech Recognition
ETSI - the European Telecommunications Standards Institute
FP   - Frame Pair
DTX  - Discontinuous Transmission
VAD  - Voice Activity Detection
```

2. ETSI DSR Front-end Codecs

Some relevant characteristics of ES 202 050 Advanced, ES 202 211 Extended, and ES 202 212 Extended Advanced DSR front-end codecs are summarized below.

2.1. ES 202 050 Advanced DSR Front-end Codec

The front-end calculation is a frame-based scheme that produces an output vector every 10 ms. In the front-end feature extraction, noise reduction by two stages of Wiener filtering is performed first. Then, waveform processing is applied to the de-noised signal and mel-cepstral features are calculated. At the end, blind equalization is applied to the cepstral features. The front-end algorithm produces at its output a mel-cepstral representation in the same format as ES 210 108, i.e., 12 cepstral coefficients [C1 - C12], C0 and log Energy. Voice activity detection (VAD) for the classification of each frame as speech or non-speech is also implemented in Feature Extraction. The VAD information is included in the payload format for each frame pair to be sent to the remote recognition engine as part of the payload. This information may optionally be used by the receiving recognition engine to drop non-speech frames. The front-end supports three raw sampling rates: 8 kHz, 11 kHz, and 16 kHz (Note that unlike some other speech codecs, the feature frame size of DSR presented to RTP packetization is not dependent on the number of speech samples used in each 10 ms sample frame. This will become more evident in the following sections).

After calculation of the mel-cepstral representation, the representation is first quantized via split-vector quantization to reduce the data rate of the encoded stream. Then, the quantized vectors from two consecutive frames are put into a FP, as described in more detail in Section 3.2.

2.2. ES 202 211 Extended DSR Front-end Codec

Some relevant characteristics of ES 202 211 Extended DSR front-end codec are summarized below.

ES 202 211 is an extension of the mel-cepstrum DSR Front-end standard ES 201 108 [11]. The mel-cepstrum front-end provides the features for speech recognition but these are not available for human listening. The purpose of the extension is allow the reconstruction of the speech waveform from these features so that they can be replayed. The front-end feature extraction part of the processing is exactly the same as for ES 201 108. To allow speech reconstruction additional fundamental frequency (perceived as pitch) and voicing class (e.g., non-speech, voiced, unvoiced and mixed) information is needed. This extra information is provided by the extended front-end processing algorithms at the device side. It is compressed and transmitted along with the front-end features to the server. This extra information may also be useful for improved speech recognition performance with tonal languages such as Mandarin, Cantonese and Thai.

Full information about the client side signal processing algorithms used in the standard are described in the specification ES 202 211 [2].

The additional fundamental frequency and voicing class information is compressed for each frame pair. The pitch for the first frame of the FP is quantized to 7 bits and the second frame is differentially quantized to 7 bits. The voicing class is indicated with one bit for each frame. The total for the extension information for a frame pair therefore consists of 14 bits plus an additional 2 bits of CRC error protection computed over these extension bits only.

The total information for the frame pair is made up of 92 bits for the two compressed front-end feature frames (including 4 bits for their CRC) plus 16 bits for the extension (including 2 bits for their CRC) and 4 bits of null padding to give a total of 14 octets per frame pair. As for ES 201 208 the extended frame pair also corresponds to 20ms of speech. The extended front-end supports three raw sampling rates: 8 kHz, 11 kHz, and 16 kHz.

The quantized vectors from two consecutive frames are put into an FP, as described in more detail in Section 3.3 below.

The parameters received at the remote server from the RTP extended DSR payload specified here can be used to synthesize an intelligible speech waveform for replay. The algorithms to do this are described in the specification ES 202 211 [2].

2.3. ES 202 212 Extended Advanced DSR Front-end Codec

ES 202 212 is the extension for the DSR Advanced Front-end ES 202 050 [1]. It provides the same capabilities as the extended mel-cepstrum front-end described in Section 2.2 but for the DSR Advanced Front-end.

3. DSR RTP Payload Formats

3.1. Common Considerations of the Three DSR RTP Payload Formats

The three DSR RTP payload formats defined in this document share the following consideration or behaviours.

3.1.1. Number of FPs in Each RTP Packet

Any number of FPs MAY be aggregate together in an RTP payload and they MUST be consecutive in time. However, one SHOULD always keep the RTP payload size smaller than the MTU in order to avoid IP fragmentation and SHOULD follow the recommendations given in Section 3.1 in RFC 3557 [10] when determining the proper number of FPs in an RTP payload.

3.1.2. Support for Discontinuous Transmission

Same considerations described in Section 3.2 of RFC 3557 [10] apply to all the three DSR RTP payloads defined in this document.

3.1.3. RTP Header Usage

The format of the RTP header is specified in RFC 3550 [8]. The three payload formats defined here use the fields of the header in a manner consistent with that specification.

The RTP timestamp corresponds to the sampling instant of the first sample encoded for the first FP in the packet. The timestamp clock frequency is the same as the sampling frequency, so the timestamp unit is in samples.

As defined by all three front-end codecs, the duration of one FP is 20 ms, corresponding to 160, 220, or 320 encoded samples with a sampling rate of 8, 11, or 16 kHz being used at the front-end, respectively. Thus, the timestamp is increased by 160, 220, or 320 for each consecutive FP, respectively.

The DSR payload for all three front-end codecs is always an integral number of octets. If additional padding is required for some other purpose, then the P bit in the RTP header may be set and padding appended as specified in RFC 3550 [8].

The RTP header marker bit (M) MUST be set following the general rules for audio codecs, as defined in Section 4.1 in RFC 3551 [9].

This document does not specify the assignment of an RTP payload type for these three new packet formats. It is expected that the RTP profile under which any of these payload formats is being used will assign a payload type for this encoding or will specify that the payload type is to be bound dynamically.

3.2. Payload Format for ES 202 050 DSR

An ES 202 050 DSR RTP payload datagram uses exactly the same layout as defined in Section 3 of RFC 3557 [10], i.e., a standard RTP header followed by a DSR payload containing a series of DSR FPs.

The size of each ES 202 050 FP remains 96 bits or 12 octets, as defined in the following sections. This ensures that a DSR RTP payload will always end on an octet boundary.

3.2.1. Frame Pair Formats

3.2.1.1. Format of Speech and Non-speech FPs

The following mel-cepstral frame MUST be used, as defined in [1]:

Pairs of the quantized 10ms mel-cepstral frames MUST be grouped together and protected with a 4-bit CRC forming a 92-bit long FP. At the end, each FP MUST be padded with 4 zeros to the MSB 4 bits of the last octet in order to make the FP aligned to the octet boundary.

The following diagram shows a complete ES 202 050 FP:

```
  Frame #1 in FP:
  ===============
     (MSB)                                    (LSB)
       0     1     2     3     4     5     6     7
    +-----+-----+-----+-----+-----+-----+-----+-----+
    :  idx(2,3) |           idx(0,1)                |    Octet 1
    +-----+-----+-----+-----+-----+-----+-----+-----+
    :       idx(4,5)        |     idx(2,3) (cont)   :    Octet 2
    +-----+-----+-----+-----+-----+-----+-----+-----+
    |             idx(6,7)              |idx(4,5)(cont)  Octet 3
    +-----+-----+-----+-----+-----+-----+-----+-----+
idx(10,11)| VAD |             idx(8,9)              |    Octet 4
    +-----+-----+-----+-----+-----+-----+-----+-----+
    :       idx(12,13)      |   idx(10,11) (cont)   :    Octet 5
    +-----+-----+-----+-----+-----+-----+-----+-----+
                            |   idx(12,13) (cont)   :    Octet 6/1
                            +-----+-----+-----+-----+
```

```
    Frame #2 in FP:
    ===============
        (MSB)                                     (LSB)
          0     1     2     3     4     5     6     7
       +-----+-----+-----+-----+
       :        idx(0,1)       |                             Octet 6/2
       +-----+-----+-----+-----+-----+-----+-----+-----+
       |              idx(2,3)             |idx(0,1)(cont)  Octet 7
       +-----+-----+-----+-----+-----+-----+-----+-----+
       :  idx(6,7) |              idx(4,5)             |     Octet 8
       +-----+-----+-----+-----+-----+-----+-----+-----+
       :        idx(8,9)       |      idx(6,7) (cont)  :     Octet 9
       +-----+-----+-----+-----+-----+-----+-----+-----+
       |          idx(10,11)         | VAD |idx(8,9)(cont)  Octet 10
       +-----+-----+-----+-----+-----+-----+-----+-----+
       |                   idx(12,13)                  |     Octet 11
       +-----+-----+-----+-----+-----+-----+-----+-----+


    CRC for Frame #1 and Frame #2 and padding in FP:
    ================================================
        (MSB)                                     (LSB)
          0     1     2     3     4     5     6     7
       +-----+-----+-----+-----+-----+-----+-----+-----+
       |  0  |  0  |  0  |  0  |          CRC          |     Octet 12
       +-----+-----+-----+-----+-----+-----+-----+-----+
```

The 4-bit CRC in the FP MUST be calculated using the formula (including the bit-order rules) defined in 7.2 in [1].

Therefore, each FP represents 20ms of original speech. Note that each FP MUST be padded with 4 zeros to the MSB 4 bits of the last octet in order to make the FP aligned to the octet boundary, as shown above. This makes the total size of an FP 96 bits, or 12 octets. Note that this padding is separate from padding indicated by the P bit in the RTP header.

The definition of the indices and 'VAD' flag are described in [1] and their value is only set and examined by the codecs in the front-end client and the recognizer.

3.2.1.2. Format of Null FP

Null FPs are sent to mark the end of a transmission segment. Details on transmission segment and the use of Null FPs can be found in RFC 3557 [10].

A Null FP for the ES 202 050 front-end codec is defined by setting the content of the first and second frame in the FP to null (i.e., filling the first 88 bits of the FP with zeros). The 4-bit CRC MUST be calculated the same way as described in Section 7.2.4 of [1], and 4 zeros MUST be padded to the end of the Null FP in order to make it aligned to the octet boundary.

3.3. Payload Format for ES 202 211 DSR

An ES 202 211 DSR RTP payload datagram is very similar to that defined in Section 3 of RFC 3557 [10], i.e., a standard RTP header followed by a DSR payload containing a series of DSR FPs.

The size of each ES 202 211 FP is 112 bits or 14 octets, as defined in the following sections. This ensures that a DSR RTP payload will always end on an octet boundary.

3.3.1. Frame Pair Formats

3.3.1.1. Format of Speech and Non-speech FPs

The following mel-cepstral frame MUST be used, as defined in Section 6.2.4 in [2]:

Immediately following two frames (Frame #1 and Frame #2) worth of codebook indices (or 88 bits), there is a 4-bit CRC calculated on these 88 bits. The pitch indices of the first frame (Pidx1: 7 bits) and the second frame (Pidx2: 5 bits) of the frame pair then follow. The class indices of the two frames in the frame pair worth 1 bit each (Cidx1 and Cidx2) next follow. Finally, a 2-bit CRC calculated on the pitch and class bits (total: 14 bits) of the frame pair is included (PC-CRC). The total number of bits in a frame pair packet is therefore 44 + 44 + 4 + 7 + 5 + 1 + 1 + 2 = 108. At the end, each FP MUST be padded with 4 zeros to the MSB 4 bits of the last octet in order to make the FP aligned to the octet boundary.

The following diagram shows a complete ES 202 211 FP:

```
     Frame #1 in FP:
     ===============
       (MSB)                                    (LSB)
         0     1     2     3     4     5     6     7
      +-----+-----+-----+-----+-----+-----+-----+-----+
      :  idx(2,3) |            idx(0,1)               |    Octet 1
      +-----+-----+-----+-----+-----+-----+-----+-----+
      :       idx(4,5)        |     idx(2,3) (cont)   :    Octet 2
      +-----+-----+-----+-----+-----+-----+-----+-----+
      |            idx(6,7)               |idx(4,5)(cont)  Octet 3
      +-----+-----+-----+-----+-----+-----+-----+-----+
       idx(10,11) |              idx(8,9)             |    Octet 4
      +-----+-----+-----+-----+-----+-----+-----+-----+
      :       idx(12,13)      |   idx(10,11) (cont)   :    Octet 5
      +-----+-----+-----+-----+-----+-----+-----+-----+
                              |   idx(12,13) (cont)   :    Octet 6/1
                              +-----+-----+-----+-----+

    Frame #2 in FP:
    ===============
       (MSB)                                    (LSB)
         0     1     2     3     4     5     6     7
      +-----+-----+-----+-----+
      :        idx(0,1)       |                            Octet 6/2
      +-----+-----+-----+-----+-----+-----+-----+-----+
      |             idx(2,3)              |idx(0,1)(cont)  Octet 7
      +-----+-----+-----+-----+-----+-----+-----+-----+
      :  idx(6,7) |              idx(4,5)             |    Octet 8
      +-----+-----+-----+-----+-----+-----+-----+-----+
      :        idx(8,9)       |      idx(6,7) (cont)  :    Octet 9
      +-----+-----+-----+-----+-----+-----+-----+-----+
      |         idx(10,11)                |idx(8,9)(cont)  Octet 10
      +-----+-----+-----+-----+-----+-----+-----+-----+
      |                   idx(12,13)                  |    Octet 11
      +-----+-----+-----+-----+-----+-----+-----+-----+

    CRC for Frame #1 and Frame #2 in FP:
    ====================================
       (MSB)                                    (LSB)
         0     1     2     3     4     5     6     7
                              +-----+-----+-----+-----+
                              |          CRC          |    Octet 12/1
                              +-----+-----+-----+-----+
```

```
 Extension information and padding in FP:
 ========================================
    (MSB)                                     (LSB)
      0     1     2     3     4     5     6     7
   +-----+-----+-----+-----+
   :       Pidx1           |                         Octet 12/2
   +-----+-----+-----+-----+-----+-----+-----+-----+
   |            Pidx2            |  Pidx1 (cont)  :  Octet 13
   +-----+-----+-----+-----+-----+-----+-----+-----+
   |  0  |  0  |  0  |  0  |  PC-CRC   |Cidx2|Cidx1|  Octet 14
   +-----+-----+-----+-----+-----+-----+-----+-----+
```

The 4-bit CRC and the 2-bit PC-CRC in the FP MUST be calculated using the formula (including the bit-order rules) defined in 6.2.4 in [2].

Therefore, each FP represents 20ms of original speech. Note, as shown above, each FP MUST be padded with 4 zeros to the MSB 4 bits of the last octet in order to make the FP aligned to the octet boundary. This makes the total size of an FP 112 bits, or 14 octets. Note, this padding is separate from padding indicated by the P bit in the RTP header.

3.3.1.2. Format of Null FP

A Null FP for the ES 202 211 front-end codec is defined by setting all the 112 bits of the FP with zeros. Null FPs are sent to mark the end of a transmission segment. Details on transmission segment and the use of Null FPs can be found in RFC 3557 [10].

3.4. Payload Format for ES 202 212 DSR

Similar to other ETSI DSR front-end encoding schemes, the encoded DSR feature stream of ES 202 212 is transmitted in a sequence of FPs, where each FP represents two consecutive original voice frames.

An ES 202 212 DSR RTP payload datagram is very similar to that defined in Section 3 of RFC 3557 [10], i.e., a standard RTP header followed by a DSR payload containing a series of DSR FPs.

The size of each ES 202 212 FP is 112 bits or 14 octets, as defined in the following sections. This ensures that an ES 202 212 DSR RTP payload will always end on an octet boundary.

3.4.1. Frame Pair Formats

3.4.1.1. Format of Speech and Non-speech FPs

The following mel-cepstral frame MUST be used, as defined in Section 7.2.4 of [3]:

Immediately following two frames (Frame #1 and Frame #2) worth of codebook indices (or 88 bits), there is a 4-bit CRC calculated on these 88 bits. The pitch indices of the first frame (Pidx1: 7 bits) and the second frame (Pidx2: 5 bits) of the frame pair then follow. The class indices of the two frames in the frame pair worth 1 bit each next follow (Cidx1 and Cidx2). Finally, a 2-bit CRC (PC-CRC) calculated on the pitch and class bits (total: 14 bits) of the frame pair is included. The total number of bits in frame pair packet is therefore 44 + 44 + 4 + 7 + 5 + 1 + 1 + 2 = 108. At the end, each FP MUST be padded with 4 zeros to the MSB 4 bits of the last octet in order to make the FP aligned to the octet boundary. The padding brings the total size of a FP to 112 bits, or 14 octets. Note that this padding is separate from padding indicated by the P bit in the RTP header.

The following diagram shows a complete ES 202 212 FP:

```
  Frame #1 in FP:
  ===============
     (MSB)                                    (LSB)
       0     1     2     3     4     5     6     7
    +-----+-----+-----+-----+-----+-----+-----+-----+
    :  idx(2,3) |           idx(0,1)              |    Octet 1
    +-----+-----+-----+-----+-----+-----+-----+-----+
    :       idx(4,5)        |      idx(2,3) (cont)   :    Octet 2
    +-----+-----+-----+-----+-----+-----+-----+-----+
    |             idx(6,7)              |idx(4,5)(cont)  Octet 3
    +-----+-----+-----+-----+-----+-----+-----+-----+
idx(10,11)| VAD |             idx(8,9)              |    Octet 4
    +-----+-----+-----+-----+-----+-----+-----+-----+
    :       idx(12,13)      |    idx(10,11) (cont)   :    Octet 5
    +-----+-----+-----+-----+-----+-----+-----+-----+
                            |    idx(12,13) (cont)   :    Octet 6/1
                            +-----+-----+-----+-----+
```

```
   Frame #2 in FP:
   ===============
       (MSB)                                     (LSB)
         0     1     2     3     4     5     6     7
      +-----+-----+-----+-----+
      :        idx(0,1)       |                           Octet 6/2
      +-----+-----+-----+-----+-----+-----+-----+-----+
      |              idx(2,3)             |idx(0,1)(cont)  Octet 7
      +-----+-----+-----+-----+-----+-----+-----+-----+
      :  idx(6,7) |            idx(4,5)               |    Octet 8
      +-----+-----+-----+-----+-----+-----+-----+-----+
      :        idx(8,9)       |      idx(6,7) (cont)  :    Octet 9
      +-----+-----+-----+-----+-----+-----+-----+-----+
      |          idx(10,11)         | VAD |idx(8,9)(cont)  Octet 10
      +-----+-----+-----+-----+-----+-----+-----+-----+
      |                   idx(12,13)                  |    Octet 11
      +-----+-----+-----+-----+-----+-----+-----+-----+


   CRC for Frame #1 and Frame #2 in FP:
   ====================================
       (MSB)                                     (LSB)
         0     1     2     3     4     5     6     7
                              +-----+-----+-----+-----+
                              |          CRC          |    Octet 12/1
                              +-----+-----+-----+-----+

   Extension information and padding in FP:
   ========================================
       (MSB)                                     (LSB)
         0     1     2     3     4     5     6     7
      +-----+-----+-----+-----+
      :       Pidx1           |                           Octet 12/2
      +-----+-----+-----+-----+-----+-----+-----+-----+
      |             Pidx2           |  Pidx1 (cont)   :    Octet 13
      +-----+-----+-----+-----+-----+-----+-----+-----+
      |  0  |  0  |  0  |  0  |  PC-CRC   |Cidx2|Cidx1|    Octet 14
      +-----+-----+-----+-----+-----+-----+-----+-----+
```

The codebook indices, VAD flag, pitch index, and class index are specified in Section 6 of [3]. The 4-bit CRC and the 2-bit PC-CRC in the FP MUST be calculated using the formula (including the bit-order rules) defined in 7.2.4 in [3].

3.4.1.2. Format of Null FP

A Null FP for the ES 202 212 front-end codec is defined by setting all 112 bits of the FP with zeros. Null FPs are sent to mark the end of a transmission segment. Details on transmission segments and the use of Null FPs can be found in RFC 3557 [10].

4. IANA Considerations

For each of the three ETSI DSR front-end codecs covered in this document, a new MIME subtype registration has been registered by the IANA for the corresponding payload type, as described below.

Media Type name: audio

Media subtype names:

dsr-es202050 (for ES 202 050 front-end)

dsr-es202211 (for ES 202 211 front-end)

dsr-es202212 (for ES 202 212 front-end)

Required parameters: none

Optional parameters:

rate: Indicates the sample rate of the speech. Valid values include: 8000, 11000, and 16000. If this parameter is not present, 8000 sample rate is assumed.

maxptime: see RFC 3267 [7]. If this parameter is not present, maxptime is assumed to be 80ms.

Note, since the performance of most speech recognizers are extremely sensitive to consecutive FP losses, if the user of the payload format expects a high packet loss ratio for the session, it MAY consider to explicitly choose a maxptime value for the session that is shorter than the default value.

ptime: see RFC 2327 [5].

Encoding considerations: These types are defined for transfer via RTP [8] as described in Section 3 of RFC 4060.

Security considerations: See Section 5 of RFC 4060.

Person & email address to contact for further information:
Qiaobing.Xie@motorola.com

Intended usage: COMMON. It is expected that many VoIP applications (as well as mobile applications) will use this type.

Author: Qiaobing.Xie@motorola.com

Change controller: IETF Audio/Video transport working group

4.1. Mapping MIME Parameters into SDP

The information carried in the MIME media type specification has a specific mapping to fields in the Session Description Protocol (SDP) [5], which is commonly used to describe RTP sessions. When SDP is used to specify sessions employing ES 202 050, ES 202 211, or ES 202 212 DSR codec, the mapping is as follows:

- o The MIME type ("audio") goes in SDP "m=" as the media name.
- o The MIME subtype ("dsr-es202050", "dsr-es202211", or "dsr-es202212") goes in SDP "a=rtpmap" as the encoding name.
- o The optional parameter "rate" also goes in "a=rtpmap" as clock rate. If no rate is given, then the default value (i.e., 8000) is used in SDP.
- o The optional parameters "ptime" and "maxptime" go in the SDP "a=ptime" and "a=maxptime" attributes, respectively.

Example of usage of ES 202 050 DSR:

```
m=audio 49120 RTP/AVP 101
a=rtpmap:101 dsr-es202050/8000
a=maxptime:40
```

Example of usage of ES 202 211 DSR:

```
m=audio 49120 RTP/AVP 101
a=rtpmap:101 dsr-es202211/8000
a=maxptime:40
```

Example of usage of ES 202 212 DSR:

```
m=audio 49120 RTP/AVP 101
a=rtpmap:101 dsr-es202212/8000
a=maxptime:40
```

4.2. Usage in Offer/Answer

All SDP parameters in this payload format are declarative, and all reasonable values are expected to be supported. Thus, the standard usage of Offer/Answer as described in RFC 3264 [6] should be followed.

4.3. Congestion Control

Congestion control for RTP MUST be used in accordance with RFC 3550 [8], and in any applicable RTP profile, e.g., RFC 3551 [9].

5. Security Considerations

Implementations using the payload defined in this specification are subject to the security considerations discussed in the RTP specification RFC 3550 [8] and any RTP profile, e.g., RFC 3551 [9]. This payload does not specify any different security services.

6. Acknowledgments

The design presented here is based on that of RFC 3557 [10]. The authors wish to thank Magnus Westerlund and others for their reviews and comments.

7. References

7.1. Normative References

[1] European Telecommunications Standards Institute (ETSI) Standard ES 202 050, "Speech Processing, Transmission and Quality Aspects (STQ); Distributed Speech Recognition; Advanced Front-end Feature Extraction Algorithm; Compression Algorithms", http://pda.etsi.org/pda/.

[2] European Telecommunications Standards Institute (ETSI) Standard ES 202 211, "Speech Processing, Transmission and Quality Aspects (STQ); Distributed Speech Recognition; Extended front-end feature extraction algorithm; Compression algorithms; Back-end speech reconstruction algorithm", http://pda.etsi.org/pda/.

[3] European Telecommunications Standards Institute (ETSI) Standard ES 202 212, "Speech Processing, Transmission and Quality aspects (STQ); Distributed speech recognition; Extended advanced front-end feature extraction algorithm; Compression algorithms; Back-end speech reconstruction algorithm", http://pda.etsi.org/pda/.

[4] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[5] Handley, M. and V. Jacobson, "SDP: Session Description Protocol", RFC 2327, April 1998.

[6] Rosenberg, J. and H. Schulzrinne, "An Offer/Answer Model with the Session Description Protocol (SDP)", RFC 3264, June 2002.

[7] Sjoberg, J., Westerlund, M., Lakaniemi, A., and Q. Xie, "Real-Time Transport Protocol (RTP) Payload Format and File Storage Format for the Adaptive Multi-Rate (AMR) and Adaptive Multi-Rate Wideband (AMR-WB) Audio Codecs", RFC 3267, June 2002.

[8] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, July 2003.

[9] Schulzrinne, H. and S. Casner, "RTP Profile for Audio and Video Conferences with Minimal Control", STD 65, RFC 3551, July 2003.

[10] Xie, Q., "RTP Payload Format for European Telecommunications Standards Institute (ETSI) European Standard ES 201 108 Distributed Speech Recognition Encoding", RFC 3557, July 2003.

7.2. Informative References

[11] European Telecommunications Standards Institute (ETSI) Standard ES 201 108, "Speech Processing, Transmission and Quality Aspects (STQ); Distributed Speech Recognition; Front-end Feature Extraction Algorithm; Compression Algorithms", http://pda.etsi.org/pda/.

Authors' Addresses

Qiaobing Xie
Motorola, Inc.
1501 W. Shure Drive, 2-F9
Arlington Heights, IL 60004
US

Phone: +1-847-632-3028
EMail: qxie1@email.mot.com

David Pearce
Motorola Labs
UK Research Laboratory
Jays Close
Viables Industrial Estate
Basingstoke, HANTS RG22 4PD
UK

Phone: +44 (0)1256 484 436
EMail: bdp003@motorola.com

Intellectual Property

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.