

3GPP2 C.S0076-0

Version 1.0

Date: December, 2005



3RD GENERATION
PARTNERSHIP
PROJECT 2
"3GPP2"

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

Discontinuous Transmission (DTX) of Speech in cdma2000 Systems

COPYRIGHT

3GPP2 and its Organizational Partners claim copyright in this document and individual Organizational Partners may copyright and issue documents or standards publications in individual Organizational Partner's name based on this document. Requests for reproduction of this document should be directed to the 3GPP2 Secretariat at secretariat@3gpp2.org. Requests to reproduce individual Organizational Partner's documents should be directed to that Organizational Partner. See www.3gpp2.org for more information.

16

1 INTRODUCTION

In circuit-switched calls over CDMA, acoustic background noise is encoded in eighth rate packets and transmitted periodically, for example every 20 msec. Although it is necessary for circuit-switched calls to transmit a packet periodically, the same is not true for packet-switched calls. It is possible to achieve capacity savings and extend battery life for VoIP calls by using a discontinuous transmission scheme (DTX).

This specification provides requirements for source controlled variable rate vocoders to implement DTX. It is recognized that there exists a great deal of flexibility in the way DTX can be implemented. The purpose of this specification is not to specify a particular solution, but to provide a set of requirements to ensure interoperability between different implementations.

This specification uses the following verbal forms: “Shall” and “shall not” identify requirements to be followed strictly to conform to the standard and from which no deviation is permitted. “Should” and “should not” indicate that one of several possibilities is recommended as particularly suitable, without mentioning or excluding others; that a certain course of action is preferred but not necessarily required; or that (in the negative form) a certain possibility or course of action is discouraged but not prohibited. “May” and “need not” indicate a course of action permissible within the limits of the standard. “Can” and “cannot” are used for statements of possibility and capability, whether material, physical, or causal.

2 DEFINITIONS

- **Comfort Noise (CN):** Output of the decoder to represent acoustic background noise.
- **Discontinuous Transmission (DTX):** Gaps inserted in the media stream during periods of non-active speech.
- **DTX Stream:** A bit stream resulting from DTX processing, i.e., a bit stream in which silence has been suppressed.
- **DTX Update Interval:** The number of frames from one SID frame to the next. For example, an update interval of 10 frames means that a SID frame is sent once every 10 frames. This interval may be variable between updates.
- **DTX min:** The minimum acceptable DTX update interval.
- **DTX max:** The maximum acceptable DTX update interval.
- **Silence Insertion Description (SID):** Frames used to convey acoustic background information during DTX periods.
- **Eighth-Rate Hangover:** A number of eighth-rate frames transmitted at the end of active speech but before DTX.
- **DSP:** Digital Signal Processor
- **PDSN:** Packet Data Serving Node
- **RTP:** Real Time Protocol

3 APPLICABILITY

This section describes the applicability of this specification.

1 **3.1 Non-DTX-Capable Legacy Vocoders**

2 Modifying an existing vocoder that is not DTX capable to comply with this specification is optional.

3 **3.2 DTX-Capable Legacy Vocoders**

4 This specification does not supercede native DTX schemes defined in legacy vocoders, like VMR-WB [2].

5 **3.3 Future Vocoders**

6 Future vocoders have the option of adopting this specification for its DTX operation.

7 **3.4 Other DTX/CN Schemes**

8 This specification does not prohibit the use of other DTX schemes, such as [1]. In addition, other DTX
9 schemes may be negotiated simultaneously with this DTX scheme if more than one vocoder is sharing the
10 same session. The mechanism by which the endpoints negotiate this DTX scheme and its parameters is
11 beyond the scope of this specification.

12 **3.5 Intermediaries**

13 DTX streams defined by this specification can be converted to continuous streams, and vice versa, with a
14 minimum of computational complexity. For example, a continuous stream can be converted to a DTX
15 stream by discarding eighth-rate frames between SID frames. Likewise, at a minimum, a DTX stream can
16 be converted to a continuous stream by replicating eighth-rate SID frames. Since the conversion can be
17 achieved without requiring the processing horsepower of a DSP, the conversion can be performed in
18 “wrapper” code outside the DSP, in de-jitter buffers, or in network elements that handle the RTP packets
19 but have limited processing power, such as the PDSN.

20 This specification permits the use of a network element to act as an intermediary between two endpoints to
21 convert DTX streams into continuous streams, and vice versa.

22 **3.6 Network**

23 This DTX specification applies to the cdma2000^{®1} access network as well as the core network.

24 **4 ENCODER REQUIREMENTS**

25 **4.1 SID Frame Format**

26 The vocoder’s existing eighth rate packet format shall be used as the SID frame format. SID frames shall
27 not consist of full, half, or quarter rate packets, nor shall a separate packet format be defined for the
28 purpose of providing SID updates.

¹ *cdma2000[®] is the trademark for the technical nomenclature for certain specifications and standards of the Organizational Partners (OPs) of 3GPP2. Geographically (and as of the date of publication), cdma2000[®] is a registered trademark of the Telecommunications Industry Association (TIA-USA) in the United States.*

4.2 Eighth-Rate Hangover

The encoder shall implement eighth-rate hangover. The length of the hangover is negotiated in signaling, which is beyond the scope of this document.

4.3 DTX Parameters

Signaling may be used to negotiate the use of DTX, following an offer/answer model. An endpoint capable of DTX shall communicate the following parameters:

- Willingness to use DTX (i.e. on/off) [Default value: on for RFC 3558[3], off for all other RTP formats].
- Minimum DTX Update Interval [Default value: 12 frames].
- Maximum DTX Update Interval [Default value: 32 frames]. The DTX max shall be greater than or equal to the DTX min.
- Eighth-Rate Hangover [Default value: 1 frame].

4.3.1 The values for DTX max, DTX min, and eighth-rate hangover shall be represented as unsigned 8-bit integers, i.e. limited to the range 0-255.

4.3.2 For vocoders using the RTP format defined in RFC 3558 [3], the default for willingness to use DTX support defined in this specification shall be “on”, and the default values for DTX min, DTX max, and eighth-rate hangover shall be used, unless negotiated otherwise. That is to say, for those vocoders that use the RTP format in RFC 3558 [3], DTX support defined in this specification is always assumed enabled unless the negotiation specifically disables it. For all other vocoders, if all four of the DTX parameters defined above are absent from either the offer or the answer, it is assumed that DTX support defined in this specification is not available in the endpoint(s).

4.3.3 If the offer as well as the answer contains DTX parameters, the value of DTX min in the answer shall be greater than or equal to the value of DTX Min in the offer, and the value of DTX max in the answer shall be less than or equal to the value of DTX max in the offer. Furthermore, the eighth-rate hangover in the answer shall be greater than or equal to the eighth-rate hangover in the offer.

4.3.4 If both offer and answer contain DTX parameters, then both sides shall pick values for DTX parameters that are proposed in the answer.

4.3.5 If none or only one of offer or answer contains DTX parameters:

- The default values for the DTX parameters shall initially be used by both sides.
- For vocoders using the RTP format defined in RFC 3558 [3], if at any time during the flow, a vocoder detects the remote vocoder as DTX-incapable, it should stop sending SID frames. That is to say that the vocoder shall still continue to send non-eighth rate frames as well as eighth rate hangover frames, but should not send eighth rate frames as SID frames. Silence should be suppressed from the end of the eighth rate hangover until the beginning of the next talk spurt.

It may be determined that a vocoder is DTX incapable if the number of eighth rate frames with contiguous RTP timestamps exceeds the default value of the eighth rate hangover.

- For vocoders using the RTP format defined in RFC 3558 [3], if a vocoder is suppressing SID frames because it detected that the remote vocoder is not DTX capable and subsequently detects that the remote vocoder is DTX capable, the default values for the DTX parameters shall be used.

It may be determined that a vocoder is DTX capable if either of the following is true:

1. The marker bit is set in the RTP header of any packet or
2. RTP Sequence Number (SN) and Timestamp (TS) do not increase in step for any RTP header, i.e., if RTP SN increases by 'n' between any two headers, but RTP TS increases by more than 'n * STRIDE', where STRIDE is the regular increase in TS between two consecutive headers (value of STRIDE is 160 for 8kHz vocoders producing 20 msec packets).

4.3.6 If it is determined through signaling that either endpoint is unwilling to use DTX, the resulting session shall not use DTX.

The details of how the signaling is accomplished are beyond the scope of this specification.

5 DECODER REQUIREMENTS

5.1 Decoding DTX Streams

The decoder shall be capable of decoding DTX streams without a priori knowledge of the DTX parameters, i.e., without knowledge of whether DTX is on, DTX min, DTX max, or eighth-rate hangover.

5.2 Missing SID frames

The decoder shall be capable of generating comfort noise without receiving any eighth-rate frames, regardless of the values that were negotiated for the DTX parameters.

6 INFORMATIVE REFERENCES

[1] RFC 3389, *Real-Time Transport Protocol (RTP) Payload for Comfort Noise*, R. Zopf, September 2002.

[2] 3GPP2 C.S0052-A v1.0, *Source-Controlled Variable-Rate Multimode Wideband Speech Codec (VMR-WB), Service Options 62 and 63 for Spread Spectrum Systems*, April 2005.

[3] RFC 3558, *RTP Payload Format for Enhanced Variable Rate Codecs (EVRC) and Selectable Mode Vocoders (SMV)*, A. Li, July 2003.