Internet Engineering Task Force
INTERNET-DRAFT
ietf-avt-tones-02.ps

AVT WG
Schulzrinne/Petrack
Columbia U./MetaTel

August 17, 1999
Expires: February 2000

RTP Payload for DTMF Digits, Telephony Tones and Telephony Signals

Abstract

This memo describes how to carry dual-tone multifrequency (DTMF) signaling, other tone signals and telephony events in RTP packets.

1 Introduction

This memo defines a payload type for carrying dual-tone multifrequency (DTMF) digits and other line and trunk signals in RTP [1] packets. A separate RTP payload type is desirable since low-rate voice codecs cannot be guaranteed to reproduce these tone signals accurately enough for automatic recognition. Defining a separate payload type also permits higher redundancy while maintaining a low bit rate.

The payload types described here may be useful in at least two applications: DTMF handling for gateways and end sytems, as well as "RTP trunks". In the first application, the Internet telephony gateway detects DTMF on the incoming circuits and sends the RTP payload described here instead of regular audio packets. The gateway likely has the necessary digital signal processors and algorithms, as it often needs to detect DTMF, e.g., for two-stage dialing. Having the gateway detect tones relieves the receiving Internet end system from having to do this work and also avoids that low bit-rate codecs like G.723.1 render DTMF tones unintelligible. Similarly, an Internet end system such as an "Internet phone" can emulate DTMF functionality without concerning itself with generating precise tone pairs.

In the "RTP trunk" application, RTP is used to replace a normal circuit-switched trunk between two nodes. This is particularly of interest in a telephone network that is still mostly circuit-switched. In this case, each end of the RTP trunk encodes audio channels into the appropriate encoding, such as G.723.1 or G.729. However, this encoding process destroys in-band signaling information which is carried using the least-significant bit ("robbed bit signaling") and may also interfere with in-band signaling tones, such as the MF digit tones. In addition, tone properties such as the phase reversals in the ANSam tone, will ot survive speech coding. Thus, the gateway needs to remove the in-band signaling information from the bit stream. It can now either carry it out-of-band in a signaling transport mechanism yet to be defined, or it can use the mechanism described in this memorandum. (If the two trunk end points are within reach of the same media gateway controller, the media gateway controller can also handle the signaling.) Carrying it in-band may simplify the time synchronization between audio packets and the tone or signal information. This is particularly relevant where duration and timing matter, as in the carriage of DTMF signals.

A gateway has two options for handling DTMF digits and signals. First, it can simply measure the frequency components of the voice band signals and transmit this information to the RTP receiver. All tone signals in use in the PSTN and meant for human consumption are sequences of simple combinations of sine waves, either added or modulated. (There is at least one tone, the ANSam tone [2] used for indicating data transmission over voice lines, that makes use of periodic phase reversals.)

As a second option, it can recognize the tones and translate them into a name, such as ringing or busy tone. The receiver then produces a tone signal or other indication appropriate to the signal. Generally, since the recognition of signals often depends on their on/off pattern or the sequence of several tones, this recognition can take several seconds. On the other hand, the gateway may have access to the actual signaling information that generates the tones and thus can generate the RTP packet immediately, without the detour through acoustic signals.

In the phone network, tones are generated at different places, depending on the switching technology and the nature of the tone. This determines, for example, whether a person making a call to a foreign country hears her local tones she is familiar with or the tones as used in the country called.

For analog lines, *Dial tone* is always generated by the local switch. ISDN terminals may generate dial tone locally and then send a Q.931 SETUP message containing the dialed digits. If the terminal just sends a SETUP message without any Called Party digits, then the switch does digit collection, provided by the terminal as KEYPAD messages, and provides dial tone over the B-channel. The terminal can either use the audio signal on the B-channel or can use the Q.931 messages to trigger locally generated dial tone.

Ringing tone (also called ringback tone) is generated by the local switch at the callee, with a one-way voice path opened up as soon as the callee's phone rings. (This reduces the chance of clipping of the called party's response just after answer. It also permits pre-answer announcements or in-band call-progress-indications to reach the caller before or in lieu of ringing tone.) *Congestion tone* and *special information tones* can be generated by any of the switches along the way, and may be generated by the caller's switch based on ISUP messages received. *Busy tone* is generated by the caller's switch, triggered by the appropriate ISUP message, for analog instruments, or the ISDN terminal.

Gateways which send signalling events via RTP SHOULD send both named signals (Section 2) and the tone representation (Section 4) as a single RTP session, using the redundancy mechanism defined in Section 2.4. The receiver can then choose the appropriate rendering.

If a gateway cannot present a tone representation, it SHOULD send the audio tones as regular RTP audio packets (e.g., as payload type PCMU), in addition to the named signals.

2 RTP Payload Format for Named Telephone Events

2.1 Requirements

The payload type for named telephone events must be suitable for both gateway and end-to-end scenarios. In the gateway scenario, a gateway connecting a packet voice network to the PSTN recreates the DTMF tones or other telephony events and injects them into the PSTN. Since, for example, DTMF digit recognition takes several tens of milliseconds, the first few milliseconds of a digit will arrive as regular audio packets. Thus, careful time and power (volume) alignment is needed to avoid generating spurious digits.

For interactive voice response (IVR) systems directly connected to the packet voice network, time alignment and volume levels are not important, since the unit will not perform any signal analysis to detect DTMF tones from the audio stream.

DTMF digits and named events are carried as part of the audio stream, and SHOULD use the same sequence number and time-stamp base as the regular audio channel to simplify recreation of analog audio at a gateway. The default clock frequency is 8,000 Hz, but the clock frequency can be redefined when assigning the dynamic payload type.

This format achieves a higher redundancy even in the case of sustained packet loss than the method proposed for the *Voice over Frame Relay Implementation Agreement* [3].

In circumstances where exact timing alignment between the audio stream and the DTMF digits is not important and data is sent unicast, such as the IVR example mentioned earlier, it may be preferable to use a reliable control stream such as H.245.

A source MAY send events and coded audio packets for the same time instants, using events as the redundant encoding for the audio stream, or it MAY block outgoing audio while event tones are active and only send named events as both the primary and redundant encodings.

Note that a period covered by an encoded tone may overlap in time with a period of audio encoded by other means. This is likely to occur at the onset of a tone and is necessary to avoid possible errors in the interpretation of the reproduced tone at the remote end. Implementations supporting this payload type must be prepared to handle the overlap.

This payload definition is used for five different types of signals:

- DTMF tones (Section 2.8);
- fax-related tones (Section 2.9);
- standard subscriber line tones (Section 2.10);
- for country-specific subscriber line tones (Section 3) and;
- for trunk events (Section 3.1).

A compliant implementation MUST support the events listed in Table 1. If it uses some other, out-of-band mechanism for signaling line conditions, it does not have to implement the other events.

In some cases, an implementation may simply ignore certain events, such as fax tones, that do not make sense in a particular environment. Depending on the available user interfaces, an implementation MAY render all tones in Table 5 the same or, preferably, use the tones conveyed by the concurrent "tone" payload or other RTP audio payload. Alternatively, it could provide a textual representation.

Note that end systems that emulate telephones only need to support the events described in Sections 2.8 and 2.10, while systems that receive trunk signaling need to implement those in Sections 2.8, 2.9, 2.10 and 3.1, since MF trunks also carry most of the "line" signals. Systems that do not support fax or modem functionality do not need to render fax-related events described in Section 2.9.

The payload type distinguishes between a (line) DTMF 0 tone and a (trunk) MF 0 tone. They payload type is signalled dynamically (for example, within an SDP [4] or an H.245 message), or by some other non-RTP means.

2.2 Use of RTP Header Fields

Timestamp: The RTP timestamp reflects the measurement point for the current packet. The event duration described in Section 2.3 extends forwards [NOTE: was "backwards", but that's different from all other payloads and disagrees with RFC 1889] from that time.

2.3 Payload Format

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     event     |R R| volume    |          duration             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

events: The DTMF digits and line events are encoded as shown in Section 2.10; the trunk events are shown in Section 3.1.

volume: The power level of the digit, expressed in dBm0 after dropping the sign, with range from 0 to -63 dBm0. The range of valid DTMF is from 0 to -36 dBm0 (must accept); lower than -55 dBm0 must be rejected (TR-TSY-000181, ITU-T Q.24A). Thus, larger values denote lower volume. This value is defined only for DTMF digits. For other events, it is set to zero.

Note: Since the acceptable dip is 10 dB and the minimum detectable loudness variation is 3 dB, this field could be compressed by at least a bit by reducing resolution to 2 dB, if needed.

duration: Duration of this digit, in timestamp units. Thus, the digit began at the instant identified by the RTP timestamp.

For a sampling rate of 8000 Hz, this field is sufficient to express digit durations of upto approximately 8 seconds.

R: This field is reserved for future use. The sender MUST set it to zero, the receiver MUST ignore it.

An audio source SHOULD start transmitting event packets as soon as it recognizes an event and every 50 ms thereafter or the packet interval for the audio codec used for this session, if known. (Precise spacing between event packets is not necessary.)

Q.24 [5], Table A-1, indicates that all administrations surveyed use a minimum signal duration of 40 ms, with signaling velocity (tone and pause) of no less than 93 ms.

If a digit continues for more than one period, it should send a new event packet with the RTP timestamp value corresponding to the beginning of the digit and the duration of the digit increased correspondingly. (The RTP sequence number is incremented by one for each packet.) If there has been no new digit in the last interval, the digit SHOULD be retransmitted three times (or until the next event is recognized) to ensure some measure of reliability for the last event.

DTMF digits and events are sent incrementally to avoid having the receiver wait for the completion of the digit. Since some tones are two seconds long, this would incur a substantial delay. The transmitter does not know if digit length is important and thus needs to transmit immediately and incrementally. If the receiver application does not care about digit length, the incremental transmission mechanism avoids delay. Some applications, such as gateways into the GSTN, care about both delays and digit duration.

2.4 Reliability

To achieve reliability even when the network loses packets, the audio redundancy mechanism described in RFC 2198 [6] is used. The effective data rate is r times 64 bits (32 bits for the redundancy header and 32 bits for the DTMF payload) every 50 ms or r times 1280 bits/second, where r is the number of redundant DTMF digits carried in each packet. The value of r is an implementation trade-off, with a value of 5 suggested.

> The timestamp offset in this redundancy scheme has 14 bits, so that it allows a single packet to "cover" 2.048 seconds of DTMF digits at a sampling rate of 8000 Hz. Including the starting time of previous digits allows precise reconstruction of the tone sequence at a gateway. The scheme is resilient to consecutive packet losses spanning this interval of 2.048 seconds or r digits, whichever is less. Note that for previous digits, only an average loudness can be represented.

An encoder MAY treat the event payload as a highly-compressed version of the current audio frame. In that mode, each RTP packet during a DTMF tone would contain the current audio codec rendition (say, G.723.1 or G.729) of this digit as well as the representation described in Section 2.3, plus any previous digits as before.

> This approach allows dumb gateways that do not understand this format to function. See also the discussion in Section 1.

2.5 Example

A typical RTP packet, where the user is just dialing the last digit of the DTMF sequence "911". The first digit was 200 ms long (1600 timestamp units) and started at time 0, the second digit lasted 250 ms (2000 timestamp units) and started at time 800 ms (6400 timestamp units), the third digit was pressed at time 1.4 s (11,200 timestamp units) and the packet shown was sent at 1.45 s (11,600 timestamp units). The frame duration is 50 ms. To make the parts recognizable, the figure below ignores byte alignment. Timestamp and sequence number are assumed to have been zero at the beginning of the first digit. In this example, the dynamic payload types 96 and 97 have been assigned for the redundancy mechanism and the telephone event payload, respectively.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|V=2|P|X|   CC  |M|     PT      |       sequence number         |
| 2 |0|0|   0   |0|     96      |              28               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                           timestamp                           |
|                             11200                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           synchronization source (SSRC) identifier            |
|                            0x5234a8                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|F|   block PT  |     timestamp offset      |   block length    |
|1|     97      |             11200         |         4         |
```

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|F|   block PT  |      timestamp offset       |   block length    |
|1|     97      |            4800             |         4         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|F|   Block PT  |
|0|     97      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     digit     |R R| volume    |          duration             |
|       9       |0 0|     7     |            1600               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     digit     |R R| volume    |          duration             |
|       1       |0 0|    10     |            2000               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     digit     |R R| volume    |          duration             |
|       1       |0 0|    20     |             400               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

2.6 Compact Reliability Scheme

[This section is more speculative.] A more compact representation could be achieved by measuring DTMF tones in a different sampling rate from that of the surrounding audio codec, e.g., as multiples of 1, 10, 40 or 50 ms. Each RTP payload type should have a fixed sampling rate, so choosing a value that depends on frame interval of the surrounding codec is not recommended. For a sampling interval of 50 ms, the following payload would "cover" 8 seconds of duration and offset:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    offset     |R R R|  digit  |R R| volume    |   duration    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

2.7 Indication of Receiver Capabilities using SDP

Receivers MAY indicate which named events they can handle, for example, by using the Session Description Protocol (RFC 2327 [4]). The payload types use the following fmtp format to list the event values that they can receive:

```
a=fmtp:<format> <list of values>
```

The list of values consists of comma-separated elements, which can be either a single decimal number or two decimal numbers separated by a hyphen (dash), where the second number is larger than the first. No whitespace is allowed between numbers or hyphens. The list does not have to be sorted.

For example, if the data "codec" (Section 2.9) has been assigned the payload type number 100 and the implementation can handle the ANS and /ANS signals only, the SDP message would contain

Event	encoding (decimal)
0–9	0–9
$*$	10
#	11
A–D	12–15
Flash	16

Table 1: DTMF events

```
a=fmtp:100 1,2,5-9,18
```

2.8 DTMF Events

Tables 1 summarizes the events belonging to the DTMF payload type. It uses the RTP encoding name “dtmf” and the MIME type “audio/dtmf”.

2.9 Data Modem and Fax Events

Table 2.9 summarizes the events and tones that can appear on a subscriber line serving a fax machine or modem. It uses the encoding name “data” and the MIME type “audio/data”. The tones are described below, with additional detail in Table 7.

ANS: This 2100 +/- 15 Hz tone is used to disable echo suppression for data transmission [7, 8]. For fax machines, Recommendation T.30 [8] refers to this tone as called terminal identification (CED) answer tone.

/ANS: This is the same signal as ANS, except that it reverses phase at an interval of 450 +/- 25 ms. It disables both echo cancellers and echo suppressors. (In the ITU Recommendation, this signal is rendered as ANS with a bar on top.)

ANSam: The modified answer tone (ANSam) [2] is a sinewave signal at 2100 +/- 1 Hz with phase reversals at an interval of 450 +/- 25 ms, amplitude-modulated by a sinewave at 15 +/- 0.1 Hz. This tone [9, 7] is sent by modems [10] and faxes to disable echo suppressors.

/ANSam: This is the same signal as ANSam, except that it reverses phase at an interval of 450 +/- 25 ms. It disables both echo cancellers and echo suppressors. (In the ITU Recommendation, this signal is rendered as ANSam with a bar on top.)

CNG: After dialing the called fax machine’s telephone number (and before it answers), the calling Group III fax machine (optionally) begins sending a CalliNG tone (CNG) consisting of an interrupted tone of 1100 Hz. [8]

CRd: Capabilities Request (CRd) [11] is a dual-tone signal with tones at tones at 1375 Hz and 2002 Hz for 400 ms for the initiating side and 1529 Hz and 2225 Hz for the responding side, followed by a single tone at 1900 Hz for 100 ms. “This signal requests the remote station transition from telephony mode to an information transfer mode and requests the transmission of a capabilities list message by the remote station. In particular, CRd is sent by the initiating station during the course of a call, or by the calling station at call establishment in response to a CRe or MRe.”

Procedure	indications
V.25 and V.8	ANS, ANS, ...
V.25, echo canceller disabled	ANS, /ANS, ANS, /ANS
V.8	ANSam, ANSam, ...
V.8, echo canceller disabled	ANSam, /ANSam, ANSam, ...

Table 2: Use of ANS, ANSam and /ANSam in V.x recommendations

CRe: Capabilities Request (CRe) [11] is a dual-tone signal with tones at tones at 1375 Hz and 2002 Hz for 400 ms, followed by a single tone at 400 Hz for 100 ms. "This signal requests the remote station transition from telephony mode to an information transfer mode and requests the transmission of a capabilities list message by the remote station. In particular, CRe is sent by an automatic answering station at call establishment."

ESi: Escape Signal (ESi) [11] is a dual-tone signal with tones at 1375 Hz and 2002 Hz for 400 ms, followed by a single tone at 980 Hz for 100 ms. "This signal requests the remote station transition from telephony mode to an information transfer mode. signal ESi is sent by the initiating station."

ESr: Escape Signal (ESr) [11] is a dual-tone signal with tones at 1529 Hz and 2225 Hz for 400 ms, followed by a single tone at 1650 Hz for 100 ms. Same as ESi, but sent by the responding station.

MRd: Mode Request (MRd) [11] is a dual-tone signals with tones at 1375 Hz and 2002 Hz for 400 ms for the initiating side and 1529 Hz and 2225 Hz for the responding side, followed by a single tone at 1150 Hz for 100 ms. "This signal requests the remote station transition from telephony mode to an information transfer mode and requests the transmission of a mode select message by the remote station. In particular, signal MRd is sent by the initiating station during the course of a call, or by the calling station at call establishment in response to an MRe." [11]

MRe: Mode Request (MRe) [11] is a dual-tone signal with tones at 1375 Hz and 2002 Hz for 400 ms, followed by a single tone at 650 Hz for 100 ms. "This signal requests the remote station transition from telephony mode to an information transfer mode and requests the transmission of a mode select message by the remote station. In particular, signal MRe is sent by an automatic answering station at call establishment." [11]

V.21: V.21 describes a 300 b/s full-duplex modem that employs frequency shift keying (FSK). It is now used by Group 3 fax machines to exchange T.30 information. The calling transmits on channel 1 and receives on channel 2; the answering modem transmits on channel 2 and receives on channel 1. Each bit value has a distinct tone, so that V.21 signaling comprises a total of four distinct tones.

In summary, procedures in Table 2 are used.

2.10 Line Events

Table 4 summarizes the events and tones that can appear on a subscriber line. It uses the encoding name "line" and the MIME type "audio/line".

ITU Recommendation E.182 [12] defines when certain tones should be used. It defines the following standard tones that are heard by the caller:

Event	encoding (decimal)
Answer tone (ANS)	32
/ANS	33
ANSam	34
/ANSam	35
Calling tone (CNG)	36
V.21 channel 1, "0" bit	37
V.21 channel 1, "1" bit	38
V.21 channel 2, "0" bit	39
V.21 channel 2, "1" bit	40
CRd	41
CRe	42
ESi	43
ESr	44
MRd	45
MRe	46

Table 3: Data and fax events

Dial tone: The exchange is ready to receive address information.

PABX internal dial tone: The PABX is ready to receive address information.

Special dial tone: Same as dial tone, but the caller's line is subject to a specific condition, such as call diversion or a voice mail is available (e.g., "stutter dial tone").

Second dial tone: The network has accepted the address information, but additional information is required.

Ringing tone: The call has been placed to the callee and a calling signal (ringing) is being transmitted to the callee.

Special ringing tone: A special service, such as call forwarding or call waiting, is active at the called number.

Busy tone: The called telephone number is busy.

Congestion tone: Facilities necessary for the call are temporarily unavailable.

Calling card service tone: The calling card service tone consists of 60 ms of the sum of 941 Hz and 1477 Hz tones (DTMF '#'), followed by 940 ms of 350 Hz and 440 Hz (U.S. dial tone), decaying exponentially with a time constant of 200 ms.

Special information tone: The callee cannot be reached, but the reason is neither "busy" nor "congestion". This tone should be used before all call failure announcements, for the benefit of automatic equipment.

Comfort tone: The call is being processed. This tone may be used during long post-dial delays, e.g., in international connections.

Hold tone: The caller has been placed on hold. Replaced by *Greensleeves*.

Record tone: The caller has been connected to an automatic answering device and is requested to begin speaking.

Caller waiting tone: The called station is busy, but has call waiting service.

Pay tone: The caller, at a payphone, is reminded to deposit additional coins.

Positive indication tone: The supplementary service has been activated.

Negative indication tone: The supplementary service could not be

Off-hook warning tone: The caller has left the instrument off-hook for an extended period of time. activated.

The following tones can be heard be either calling or called party during a conversation:

Call waiting tone: Another party wants to reach the subscriber.

Warning tone: The call is being recorded. This tone is not required in all jurisdictions.

Intrusion tone: The call is being monitored, e.g., by an operator. (Use by law enforcement authorities is optional.)

CPE alerting signal (CAS): A tone used to alert a device to an arriving in-band FSK data transmission. A CAS is a combined 2130 and 2750 Hz tone, both with tolerances of 0.5% and a duration of 80 to 80 ms. CAS is used with ADSI services and Call Waiting ID services, see Bellcore GR-30-CORE, Issue 2, December 1998, Section 2.5.2.

The following tones are heard by operators:

Payphone recognition tone: The person making the call or being called is using a payphone (and thus it is ill-advised to allow collect calls to such a person).

3 Extended Line Events

Table 5 summarizes country-specific events and tones that can appear on a subscriber line. It uses the encoding name "linex" and the MIME type "audio/linex".

3.1 Trunk Events

Table 6 summarizes the events and tones that can appear on a trunk. It uses the encoding name "TRUNK" and the MIME type "audio/trunk". Note that trunk can also carry line events, as MF signaling does not include backward signals [13, p. 93].

[NOTE: the list below, below wink, does not agree with the MF description in van Bosse, p. 74.]

Wink: A brief transition, typically 120-290 ms, from on-hook (unseized) to off-hook (seized) and back to onhook, used by the incoming exchange to signal that the call address signaling can proceed.

Event	encoding (decimal)
Off Hook	64
On Hook	65
Dial tone	66
PABX internal dial tone	67
Special dial tone	68
Second dial tone	69
Ringing tone	70
Special ringing tone	71
Busy tone	72
Congestion tone	73
Special information tone	74
Comfort tone	75
Hold tone	76
Record tone	77
Caller waiting tone	78
Call waiting tone	79
Pay tone	80
Positive indication tone	81
Negative indication tone	82
Warning tone	83
Intrusion tone	84
Calling card service tone	85
Payphone recognition tone	86
CPE alerting signal (CAS)	87
Off-hook warning tone	88

Table 4: E.182 line events

Event	encoding (decimal)
Acceptance tone	96
Confirmation tone	97
Dial tone, recall	98
End of three party service tone	99
Facilities tone	100
Line lockout tone	101
Number unobtainable tone	102
Offering tone	103
Permanent signal tone	104
Preemption tone	105
Queue tone	106
Refusal tone	107
Route tone	108
Valid tone	109
Waiting tone	110
Warning tone (end of period)	111
Warning Tone (PIP tone)	112

Table 5: Country-specific Line events

Event	encoding (decimal)
MF 0...9	128...137
MF K0 or KP (start-of-pulsing)	138
MF K1	139
MF K2	140
MF S0 to ST (end-of-pulsing)	141
MF S1...S3	142...143
Wink	144
Wink off	145
Incoming seizure	146
Return seizure	147
Unseize circuit	148
Continuity test	149
Default continuity tone	150
Continuity tone (single tone)	151
Continuity test send	152
Continuity verified	153
Loopback	154
Old milliwatt tone (1000 Hz)	155
New milliwatt tone (1004 Hz)	156

Table 6: Trunk events

Incoming seizure: Incoming indication of call attempt (off-hook).

Return seizure: Seizure by answering exchange, in response to outgoing seizure. [NOTE: Not clear why the difference here, but not for Unseize. Should probably be just Seizure.]

Unseize circuit: Transition of circuit from off-hook to on-hook at the end of a call.

Wink off: A brief transition, typically 100-350 ms, from off-hook (seized) to on-hook (unseized) and back to off-hook (seized). Used in operator services trunks. [CHECK!]

Continuity tone send: A tone of 2010 Hz.

Continuity tone detect: A tone of 2010 Hz.

Continuity test send: A tone of 1780 Hz is sent by the calling exchange. If received by the called exchange, it returns a "continuity verified" tone.

Continuity verified: A tone of 2010 Hz. This is a response tone, used in dual-tone procedures.

Line test: 105 [EXPLAIN!] test line progress tones (2225 Hz at -10 dbm0).

4 RTP Payload Format for Telephony Tones

4.1 Requirements

As an alternative to describing tones and events by name, it is sometimes preferable to describe them by their acoustic properties. In particular, recognition is faster than for naming signals.

There is no single international standard for telephone tones such as dial tone, ringing (ringback), busy, congestion ("fast-busy"), special announcement tones or some of the other special tones, such as payphone recognition, call waiting or record tone. However, across all countries, these tones share a number of characteristics [14]:

- Tones consist of either a single tone, the addition of two or three tones or the modulation of two tones. (Almost all tones use two frequencies; only the Hungarian "special dial tone" has three.) Tones that are mixed have the same amplitude and do not decay.
- Tones for telephony events are in the range of 25 (ringing tone in Angola) to 1800 Hz. CED is the highest used tone at 2100 Hz. The telephone frequency range is limited to 3,400 Hz.
- Modulation frequencies range between 15 (ANSam tone) to 480 Hz (Jamaica). Non-integer frequencies are used only for frequencies of 16 2/3 and 33 1/3 Hz. (These fractional frequencies appear to be derived from older AC power grid frequencies.)
- Tones that are not continuous have durations of less than four seconds.
- ITU Recommendation E.180 [15] notes that different telephone companies proscribe a tone accuracy of between 0.5 and 1.5%. The Recommendation suggests a frequency tolerance of 1%.

Tone name	frequency	on period	off period
CNG	1100	0.5	3.0
CED	2100	3.3	–
ANS	2100	3.3	–
ANSam	2100*15	3.3	–
V.21 “0” bit, ch. 1	1180	0.033	
V.21 “1” bit, ch. 1	980	0.033	
V.21 “0” bit, ch. 2	1850	0.033	
V.21 “1” bit, ch. 2	1650	0.033	
ITU dial tone	425	–	–
U.S. dial tone	350+440	–	–
ITU ringing tone	425	0.67–1.5	3–5
U.S. ringing tone	440+480	2.0	4.0
ITU busy tone	425		
U.S. busy tone	480+620	0.5	0.5
ITU congestion tone	425		
U.S. congestion tone	480+620	0.25	0.25

Table 7: Examples of telephony tones

4.2 Examples of Common Telephone Tone Signals

As an aid to the implementor, Table 7 summarizes some common tones. The rows labeled “ITU ...” refer to the general recommendation of Recommendation E.180 [15]. Note that there are no specific guidelines for these tones. In the table, the symbol “+” indicates addition of the tones, without modulation, while “*” indicates amplitude modulation. [ADD ADDITIONAL COUNTRIES, IF DESIRED.] The meaning of some of the tones is described in Section 2.10 or Section 2.9 (for V.21).

4.3 Use of RTP Header Fields

Timestamp: The RTP timestamp reflects the measurement point for the current packet. The event duration described in Section 2.3 extends forwards [NOTE: was “backwards”, but that’s different from all other payloads and disagrees with RFC 1889] from that time.

4.4 Payload Format

Based on the characteristics described above, the payload format is shown in Fig. 1.

The payload contains the following fields:

modulation: The modulation frequency, in Hz. The field is a 9-bit unsigned integer, allowing modulation frequencies up to 511 Hz. If there is no modulation, this field has a value of zero.

T: If the “T” bit is set (one), the modulation frequency is to be divided by three. Otherwise, the modulation frequency is taken as is.

```
0                   1                   2                   3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    modulation    |T|  volume   |          duration             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|R R R R|       frequency        |R R R R|       frequency       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|R R R R|       frequency        |R R R R|       frequency       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
......

+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|R R R R|       frequency        |R R R R|      frequency        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: Payload format for tones

volume: The power level of the digit, expressed in dBm0 after dropping the sign, with range from 0 to -63 dBm0. (Note: A preferred level range for digital tone generators is -8 dBm0 to -3 dBm0.)

duration: The duration of the tone, measured in timestamp units. The tone begins at the instant identified by the RTP timestamp and lasts for the duration value.

> The definition of duration corresponds to that for sample-based codecs, where the timestamp represents the sampling point for the first sample.

frequency: The frequencies of the tones to be added, measured in Hz and represented as a 12-bit unsigned integer. The field size is sufficient to represent frequencies up to 4095 Hz, which exceeds the range of telephone systems. A value of zero indicates silence.

R: This field is reserved for future use. The sender MUST set it to zero, the receiver MUST ignore it.

The RTP payload type is designated as "TONE", the MIME type as "audio/tone". The default timestamp rate is 8,000 Hz, but other rates may be used. Note that the timestamp rate does not affect the interpretation of the frequency, just the durations.

4.5 Reliability

Same as Section 2.4.

5 Combining Tones and Named Signals

The payload formats in Sections 2 and 4 can be combined into a single payload, as shown in the example depicted in Fig. 2. In the example, the RTP packet combines two TONE and one LINE payload. The payload types are chosen arbitrarily as 97 and 98, respectively, with a sample rate of 8000 Hz. Here, the redundancy format has the dynamic payload type 96.

The packet represents a snapshot of U.S. ringing tone, 1.5 seconds (12,000 timestamp units) into the second "on" part of the 2.0/4.0 second cadence, i.e., a total of 7.5 seconds (60,000 timestamp units) into the ring cycle. The 440 + 480 Hz tone of this second cadence started at RTP timestamp 48,000. Four seconds of silence preceded it, but since RFC 2198 only has a fourteen-bit offset, only 2.05 seconds (16383 timestamp units) can be represented. Even though the tone sequence is not complete, the sender was able to determine that this is indeed ringback, and thus includes the corresponding LINE event.

6 History

- This draft combines draft-ietf-avt-dtmf-00 and draft-ietf-avt-telephone-tones-01.
- From draft draft-ietf-avt-dtmf-00, the interval was changed to be uniform at 50 ms, since audio frame interval may change based on codec.
- From draft-ietf-avt-telephone-tones-01, a generic tone representation was added.

7 IANA Considerations

This document defines three new RTP payload names and associated MIME Types, TONE (audio/tone), LINE (audio/line) and TRUNK (audio/trunk). Within the TRUNK and LINE RTP payload types, additional entries for events MUST be registered with IANA. Before registration, IANA should consult the current chair of the AVT working group or its successor to avoid duplication of definitions.

8 Acknowledgements

The suggestions of the Megaco working group are gratefully acknowledged. Detailed advice and comments were provided by Fred Burg, Fatih Erdin, Mike Fox, Terry Lyons, and Steve Magnell.

9 Authors

Henning Schulzrinne
Dept. of Computer Science
Columbia University
1214 Amsterdam Avenue
New York, NY 10027
USA
electronic mail: schulzrinne@cs.columbia.edu

Scott Petrack
MetaTel
284 North Avenue
Weston, MA 02493

```
0                   1                   2                   3
0 1 2 3 4 5 6 7 8 9 0 1 2 3  4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|V=2|P|X|  CC   |M|     PT      |       sequence number         |
| 2 |0|0|   0   |0|     96      |              31               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                           timestamp                           |
|                             48000                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           synchronization source (SSRC) identifier            |
|                            0x5234a8                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|F|   block PT  |     timestamp offset      |   block length    |
|1|     98      |            16383          |         4         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|F|   block PT  |     timestamp offset      |   block length    |
|1|     97      |            16383          |         8         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|F|   Block PT  |
|0|     97      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  event=ring   |0|0| volume=0  |     duration=28383            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| modulation=0    |0| volume=63 |     duration=16383            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0 0 0 0|     frequency=0       |0 0 0 0|    frequency=0        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| modulation=0    |0| volume=5  |     duration=12000            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0 0 0 0|     frequency=440     |0 0 0 0|    frequency=480      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: Combining tones and events in a single RTP packet

USA
electronic mail: scott.petrack@metatel.com

References

[1] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: a transport protocol for real-time applications," Request for Comments (Proposed Standard) 1889, Internet Engineering Task Force, Jan. 1996.

[2] International Telecommunication Union, "Procedures for starting sessions of data transmission over the public switched telephone network," Recommendation V.8, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, Feb. 1998.

[3] R. Kocen and T. Hatala, "Voice over frame relay implementation agreement," Implementation Agreement FRF.11, Frame Relay Forum, Foster City, California, Jan. 1997.

[4] M. Handley and V. Jacobson, "SDP: session description protocol," Request for Comments (Proposed Standard) 2327, Internet Engineering Task Force, Apr. 1998.

[5] International Telecommunication Union, "Multifrequency push-button signal reception," Recommendation Q.24, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, 1988.

[6] C. Perkins, I. Kouvelas, O. Hodson, V. Hardman, M. Handley, J. C. Bolot, A. Vega-Garcia, and S. Fosse-Parisis, "RTP payload for redundant audio data," Request for Comments (Proposed Standard) 2198, Internet Engineering Task Force, Sept. 1997.

[7] International Telecommunication Union, "Automatic answering equipment and general procedures for automatic calling equipment on the general switched telephone network including procedures for disabling of echo control devices for both manually and automatically established calls," Recommendation V.25, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, Oct. 1996.

[8] International Telecommunication Union, "Procedures for document facsimile transmission in the general switched telephone network," Recommendation T.30, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, July 1996.

[9] International Telecommunication Union, "Echo cancellers," Recommendation G.165, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, Mar. 1993.

[10] International Telecommunication Union, "A modem operating at data signalling rates of up to 33 600 bit/s for use on the general switched telephone network and on leased point-to-point 2-wire telephone-type circuits," Recommendation V.34, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, Feb. 1998.

[11] International Telecommunication Union, "Procedures for the identification and selection of common modes of operation between data circuit-terminating equipments (dces) and between data terminal equipments (dtes) over the public switched telephone network and on leased point-to-point telephone-type circuits," Recommendation V.8bis, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, Sept. 1998.

[12] International Telecommunication Union, "Application of tones and recorded announcements in telephone services," Recommendation E.182, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, Mar. 1998.

[13] J. G. van Bosse, *Signaling in Telecommunications Networks*. Telecommunications and Signal Processing, New York, New York: Wiley, 1998.

[14] International Telecommunication Union, "Various tones used in national networks," Recommendation Supplement 2 to Recommendation E.180, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, Jan. 1994.

[15] International Telecommunication Union, "Technical characteristics of tones for telephone service," Recommendation Supplement 2 to Recommendation E.180, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, Jan. 1994.