Providing Touch-Tone Services in Voice Over Packet Networks

Julian Penketh, MSc, AMIEE, Nortel Networks

Abstract: *For increased bandwidth savings in packet networks, speech compression algorithms are applied to reduce the amount of data that needs to be transferred to adequately reconstruct a telephone signal at the receiver. However, such algorithms do not process pure tones (such as dual-tone multi-frequency (DTMF)) with high fidelity, hence all signalling tones must be detected before packetization occurs. This paper presents the areas currently under or being researched related to the transfer of DTMF tone data across ATM networks.*

1 Introduction

1.1 Objectives

Under normal circumstances, that is when speech only is present at the ATM transmitter (the near end), the speech is encoded to low bit-rate data (for example from 64 to 8kb/s using the G.729 codec [1]). At the far end, the ATM cells are received and the payload data is decoded using G.729 to reproduce the 64kb/s samples. However, if DTMF tones are encoded and decoded in this way, significant distortion at the far end is observed on the DTMF signal. The objective of the work is to design and integrate a multi-channel DTMF transceiver system into an ATM (and IP) gateway switch known as Passport Voice Gateway (PVG). Upon reception of incoming DTMF tone the ATM transmitter will issue an AAL2-type 3 packet containing a digit code representing the dialled digit [2,3], and level of incoming tone [4]. At the far end, the ATM receiving station will regenerate the respective DTMF signal. The requirements include fast DTMF tone detection time, high-noise immunity, high speech/tone discrimination capability, low computational time and complexity, and low code density (whole system exists as software sub-modules). Figure 1.1 shows the digit keypad and associated frequency pairs.

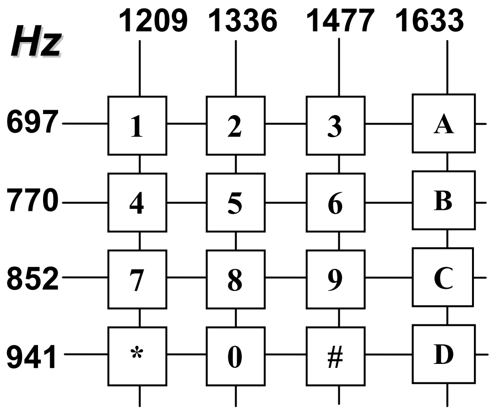


Figure 1.1 Dialled digit keypad layout and associated frequency pairs

Figure 1.2 demonstrates how the DTMF detector and regenerator blocks interact with the G729 encoder and decoder blocks; the DTMF detector effectively opens a switch when it validates an incoming tone-pair at the TDM input to replace G.729 encoding and packetisation with digit relay. According to the ITU standard, packets need to be sent triplicated with 5ms intervals [2]; this requirement is also under investigation since the quality of service (QoS) of ATM may in terms of cell loss be high enough that transmitting single packets suffices.

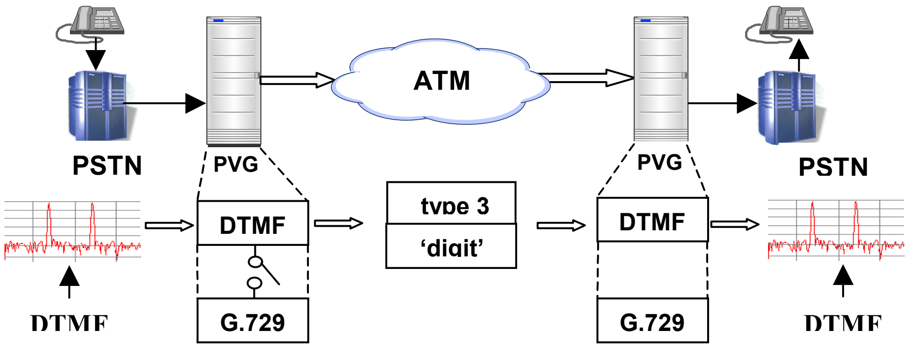


Figure 1.2 DTMF digit transmission and reproduction across an ATM network

2 Performance issues at the network layer

2.1 Scenario A: signal distortion

DTMF digit recognition is not instantaneous since a frequency scan must take place across a window of incoming samples before a result can be obtained (current scheme uses 102 samples [5]). This means that a certain portion of data which failed to be detected at the beginning of the incoming DTMF burst will be packetised as G.729 speech. See Figure 2.1.

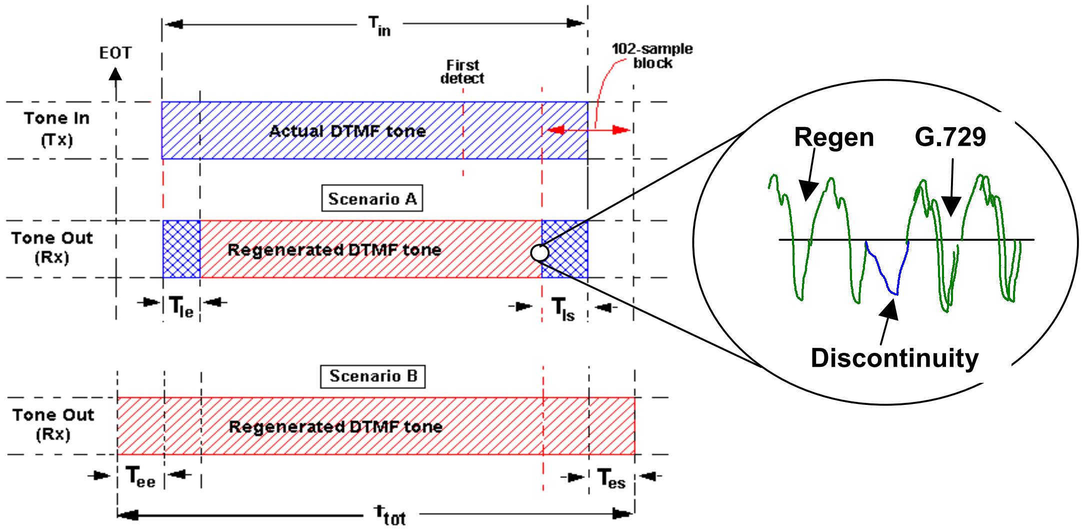


Figure 2.1 Network effects from DTMF detection processing delay

2.1.1 G.729 'leakage'

The portion arriving at the far end as G.729 speech will be slightly distorted. At the point 'first detect', the digit is relayed over ATM and the far end begins playing out the according DTMF tone on TDM. When the transmitter detects the end of the DTMF tone, it sends an end of tone (EOT) message to the far end whereupon generation of DTMF tone stops. However, an unknown number of DTMF samples (from 0 to <102) may reside in this last detection window. These samples will pass through the network as G.729 speech and be played out at the far end TDM and will also be slightly distorted. This leakage is shown as T_{ls} for the amount at the start of the tone, and T_{le} for end of tone distortion duration. In the tandem case (when the tone at the output stage becomes the input to another relay system) some of the distorted tone sections may once more be encoded as

G.729. Depending on the number of stages in the network the ‘effective’ signal could become truncated, the corruption of the start and end portions being so severe that the DTMF tone cannot be detected at these points. Ultimately digits may eventually not arrive at their final destinations in a link.

2.1.2 G.729 / Regeneration boundary discontinuities

See expanded view in Figure 2.1. At the interface between the end of the tone processed by the G.729 codec and the start of the regenerated tone there will be a phase discontinuity. This has the appearance of a glitch on the outgoing TDM. This distortion will also propagate through the various stages in the network and could result in valid DTMF digits not being recognised. There will be a discontinuity at both the start and end distorted portions.

2.2 Scenario B: signal expansion

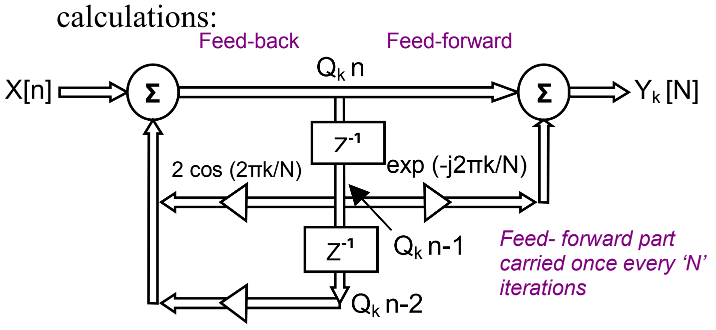
The second part of Figure 2.1 shows how the reproduced length of the DTMF tone can actually be increased, in the worst case from its beginning and its end also. If the detector is able to recognise a digit from the first window (which contains less than 102 samples), then the regenerator at the far end will begin by generating a full 102 sample buffer’s worth of DTMF tone. This effectively advances the signal in time, extending the regenerated length from the start by an amount T_{es}. Similarly, if the last window containing DTMF samples renders a valid digit a full buffer of 102 samples will be played out at the far-end TDM. The signal has thus been further extended at the end by an amount $T_{ee.}$ Total length of a DTMF tone is then:

$$T_{tot} = T_{es} + T_{ee} \tag{1}$$

Note : the periods T_{es} and T_{ee} can equal zero, and there could even be zero leakage. However, the likely scenario is that the resulting DTMF signal reproduced at the far end will exhibit a combination of distortion and expansion.

3 Mitigating adverse network effects

In order to reduce the amount of tone leakage through the speech codec, the processing delay for DTMF detection must be. A number of DTMF detection algorithms have been evaluated such as FFT, DFT, ACF, and filter bank. In-house techniques based on Neural Networks [6] are also being studied; these show promise in efficiency and speed of detection but are not yet spec-compliant [4]. Other methods currently under investigation are Chirp-Z, and DESA [7] which claims to be able to estimate a digit value using just five sample points. The chosen detector algorithm was the industry-standard and very efficient Goertzel resonator bank. Figure 3.1 shows the resonator structure and associated calculations:



Iterative feedback section (n = 0 to N) :

$Q_k[n] = 2 \cos (2\pi k/N).\ Q_k\, n\text{-}1 - Q_k\, n\text{-}2 + x[n]$

Feedforward section at n = N, get squared magnitude: $|Y_k[N]|^2 =$

$Q_k^2\, n\text{-}1 + Q_k^2\, n\text{-}2 - 2 \cos (2\pi k/N).\ Q_k n\text{-}1.\ Q_k n\text{-}2$

Figure 3.1 Goertzel resonator block

Each two-pole resonator is tuned to one of the DTMF frequencies (16 resonators used for eight tones since strength of second harmonics is also analysed). Design requires only 16(N+2) real multiplies and 32(N+1) real additions. Using a 102-sample window digits can be decoded in 12.75ms and the current implementation uses only one DSP MIPS per channel. In order to be robust against speech, a series of other checks are also performed such as relative strength of second harmonics, and row and column tone levels (twist) etc. For added robustness, a decision is not made about the validity of a DTMF digit until two successive digits have been recognised. However, this increases the digit processing delay to ~26ms, with obvious ramifications in the network scenario.

4 Current solution

4.1 Goertzel algorithm with early warning indicator (EWI)

In order to keep the amount of tone leakage through the voice codecs to a minimum, an early warning indicator (EWI) block has been added to the detector. The current implementation uses a comparative energy template to discriminate between speech, silence/noise, and incoming DTMF tones, which it does in under 5ms. When the EWI determines that DTMF is present at the TDM input, the full digit detector is called to decode the digit. The EWI filter is tuned to never miss a true incoming digit tone-pair, but can mistake speech for DTMF. This is not detrimental since the DTMF detection algorithm will reject the speech, which is subsequently packetised via G.729. The advantage is that a maximum of 5ms of DTMF signal will be G.729 encoded.

5 Conclusions and further work

Clearly, the optimal system is one which can detect and decode a digit reliably in as little time as possible (Neural Network approach being investigated decodes in 5ms). Research by using network simulation and DSP tools will be carried out to evaluate the limits in signal distortion, number of ATM 'hops', and detection window sizes. Other ideas include pre-Goertzel filtering and two for one decimation in time to reduce the number of sampling points.

References

[1] ITU-T Recommendation G.729 Annex B

[2] "AAL2 Service Specific Convergence Sublayer for Trunking, Annex K", ITU-T Recommendation I.366.2

[3] "Technical Features of Push-Button Telephone Sets", ITU-T Recommendation Q.23

[4] "Multi-Frequency Push-Button Signal Reception", ITU-T Recommendation Q.24

[5] G.Goertzel, "An Algorithm for the Evaluation of Finite Trigonometric Series", American Math Monthly, Vol.65, Jan. 1958, pp 34-35.

[6] "Fast and Efficient DTMF Detection and Decoding Using Parallel Neural Networks for Packetized Data", M.Ahmadi, to be published in ICSPAT 2000.

[7] "Detection of multi-tone signals based on energy operators", Velez, E.F. Time-Frequency and Time-Scale Analysis, 1994., Proceedings of the IEEE-SP International Symposium on , 1994 , Page(s): 229-232.