

Telgo323: An H.323 Bridge for Deaf Telephony

Jason B. Penton (Rhodes), William D. Tucker (UWC) and Meryl Glaser (UCT)
Department of Computer Science, Rhodes University, Grahamstown, South Africa
C: 082 376 6811 F: 046 603 1915 email: j.penton@ru.ac.za

Abstract - We have developed a prototype bridge that relays text and speech between Teldem, a text telephone for the Deaf, and a standard telephone or H.323 endpoint. Telgo323 uses modified H.323 media gateways and open source Text to Speech and Speech to Text software. The approach allows for easy integration of new tools as the technologies mature. This paper presents the design of the implementation prototype, discusses Teldem tone decoding, and suggests directions for future work. The Telgo323 provides evidence that an automated relay bridge is imminently viable for the Deaf Community, and further demonstrates an attractive approach for building bridges over the Digital Divide.

Index terms - Deaf telephony, Digital Divide bridges, H.323 gateway modification, Speech to Text, Text to Speech

I. INTRODUCTION

In South Africa, Telkom provides a device called Teldem (Figure 1) that enables Deaf people to communicate with one another over the Public Switched Telephone Network (PSTN). The Teldem is essentially a text telephone and is an asynchronous device that allows two communicating parties to type to each other in real-time. One of several major drawbacks of this system is that there is no mechanism for a Deaf person, using a Teldem, to communicate with any other party (Deaf or not) without also having a Teldem [1]. This means that if a Deaf person wishes to call a hearing person, the hearing person is required to use a Teldem.



Figure 1. Telkom's Teldem

Last year, we proposed a series of bridges to enhance synchronous telecommunications possibilities for the Deaf community by relaying between voice and text [2]. The most interesting and technically challenging bridge was between the Teldem and a standard telephone. This bridge would be

responsible for enabling the two devices to communicate by providing media conversion between the two. This bridge is the focus of this paper.

We have partially implemented a prototype consisting of two H.323/ISDN (Integrated Services Digital Network) gateways extended to provide support for Teldem communication. We have named the system Telgo323 (Teldem goes H.323) because of the H.323 IP signalling layer [3,4,5] responsible for the communication between the two gateways on the IP network.

Essentially, the requirement of this bridge is to convert voice, originating from the telephone, into a textual equivalent that can be presented to the Teldem user. Similarly, the text originating from the Teldem needs to be converted to the voice equivalent and presented to the telephone user. These two scenarios can be implemented using speech-to-text (STT) and text-to-speech (TTS) technologies, respectively. STT technologies are still not mature enough to be used in this system because:

- the users of the system are unable to train the STT software
- the audio output from the telephone may not be of a high enough quality for successful speech recognition
- there is often a considerable amount of noise apparent in telephone conversations that may affect the result of speech recognition.

As a result, we decided that we would implement the system in such a way that more improved STT systems could be integrated into the bridge as they emerge. This means that the quality of the bridging system would improve as STT systems become more mature. The same holds for TTS tools. Thus we have designed the system to allow for "plug and play" of both STT and TTS tools.

The bridge we have built at Rhodes University is still a work in progress. There is, however, enough implementation to report on the progress we have achieved so far and this paper provides technical information regarding the system's implementation to date. In addition, we will define the scope of future work and situate this work within the bigger picture of building bridges across the Digital Divide.

II. TELGO323 ARCHITECTURE AND OPERATION

The bridge is implemented using two modified H.323/ISDN gateways. The system could be implemented using a single gateway capable of determining the nature of the originating and terminating devices (Teldem or telephone), but for simplicity it was implemented using two. This simplification omits the complexity of a single gateway having to identify the nature of the originating and terminating devices on the PSTN.

Using two gateways means that there are two separate numbers to dial for the service, depending on the nature of the originating device. For example, if the caller were a Deaf person using a Teldem, s/he would dial 6038000, the telephone number of the Teldem gateway. Once connected the Teldem user is prompted for the number of the remote telephone to dial. Once the user has entered this number, the Teldem gateway establishes a call to the telephone via the second gateway, called the telephone gateway. This gateway relays the voice call to the hearing user who may use a PSTN telephone, a cell phone or an H.323 terminal such as NetMeeting.

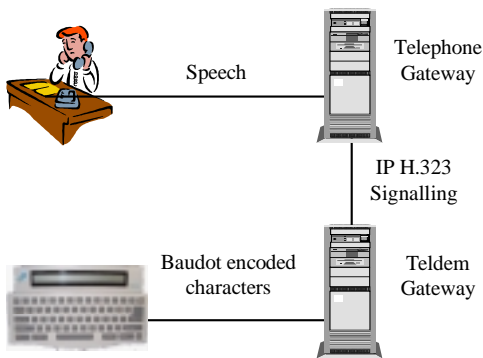


Figure 2. Telgo323 Architecture

A similar sequence of events occurs if the caller is a telephone user, with the exception that the gateway number is different (6038001) and that the hearing user is presented with a voice message requesting her/him to enter the number of the remote Teldem with which s/he wishes to communicate. This number is entered on the user's telephone and transmitted to the telephone gateway as a series of Dual Tone Multi-Frequency (DTMF) tones. On receipt of the number, the telephone gateway establishes a call with the Teldem via the Teldem gateway. From this point onwards, until the connection is dropped by either party or by the network, the gateways are working together to provide the communication link between the Teldem and the telephone. The architecture of the bridging system is illustrated in Figure 2.

A. Telephone Gateway

In addition to providing the telephone user with an interface into the IP network, the telephone gateway converts the audio it receives from the user into its textual equivalent. The text message is sent to the Teldem gateway using the H.323 *UserInput* capability. This mechanism allows the telephone gateway to send a string of characters, in this case the text message to be sent to the Teldem, to the Teldem gateway via standard H.323 messages.

In the reverse direction, the telephone gateway expects G.711 encoded audio from the Teldem gateway via an H.323 Real-Time Protocol (RTP) channel. This encoded audio is decoded into raw pulse-code modulated (PCM) audio and sent via the PSTN to the user's telephone.

B. Teldem Gateway

Similarly, the Teldem gateway provides Teldem devices with an interface into the IP network. In addition, the gateway converts the incoming tones, representing text from the Teldem, into the voice message equivalent. The voice is encoded as a G.711 audio stream and sent to the telephone gateway via an H.323 RTP audio channel. The telephone gateway relays this audio to the telephone user on the PSTN.

In the reverse direction, the Teldem gateway awaits text messages, carried within H.323 *UserInput* messages, from the telephone gateway. These text messages are finally converted to a Baudot-encoded binary format and modulated for transmission to the Teldem via the PSTN.

III. TELGO323 SOFTWARE ARCHITECTURE

We used the H.323 signalling protocol to build this bridge because H.323 is a mature standard and it integrates well with the PSTN. In addition, we already had a reliable H.323/ISDN gateway [5] that could be extended with Teldem communication functionality. Note in Figure 2 that this signalling occurs between the two gateways.

The gateways used in this system are built using the open source H.323 library provided by Equivalence PTY, called OpenH323 (www.openh323.org, www.h323forum.org). These gateways are not built from the ground up, but are instead modified H.323/ISDN gateways. Originally, the H.323/ISDN gateway enables telephones on the PSTN and H.323 terminals on the IP network to engage in voice communications with each other. By modifying two separate instances of this gateway, we wanted to create a system that enables telephones and Teldems (both on the PSTN) to communicate with each other via the IP network. The IP network is responsible for housing the logic that converts the communications into the appropriate forms compatible for each of the two devices (voice for telephones and text for Teldems). The modifications to this gateway are described later in this section.

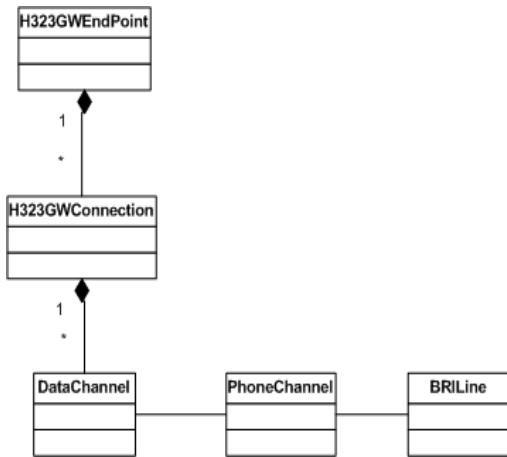


Figure 3. UML representation of unmodified standard H.323/ISDN Gateway

The Unified Modelling Language (UML) representation of the standard voice gateway without the Teldem modifications is illustrated in Figure 3¹. The root class of this gateway is the H323GWEndPoint class that represents the main thread of execution for the gateway. A single instance of this class is created when the gateway is started. Within the H323GWEndPoint object there may be any number of H323GWConnection objects, limited of course by the number of hardware interfaces into the PSTN. This H323GWConnection class encapsulates all the data and operations pertaining to a single connection to the gateway. When a Teldem connects to the gateway, an instance of this class is instantiated, including the instantiation of the audio channels used to carry the audio data. These audio channels are encapsulated within the DataChannel class that is responsible for sending and receiving audio data to or from H.323 terminals on the IP network. Two instances of the DataChannel class exist for each connection, one for outgoing or encoded audio (from the telephone to the H.323 terminal) and one for incoming or decoded audio (from the H.323 terminal to the telephone). These channels are attached to a specific codec class, e.g. G.711, that is responsible for encoding or decoding the audio at the gateway and H.323 terminal, respectively.

The BRILine class represents the actual ISDN Basic Rate Interface (BRI) line. This class embodies all the data and operations required to interact with the PSTN via an ISDN connection, for example reading and writing raw PCM audio to or from the line or detecting an incoming connection. The PhoneChannel class provides the link between the DataChannel and BRILine classes. The PhoneChannel class maintains the state of the BRILine (connected, dialing, ringing, etc.) and also provides the DataChannel class access to read or write audio to or from the ISDN line via the BRILine class.

To conclude this description, consider a small example scenario. Assume that the gateway has already established a call between a PSTN telephone and an H.323 terminal on the IP network and that all the necessary objects have been instantiated. Consider audio originating at the telephone and terminating at the H.323 terminal.

Firstly, the PhoneChannel object will know that it is connected and that audio is being transmitted between the gateway and H.323 terminal. When audio arrives on the ISDN BRI line, it is stored in a buffer within the BRILine object. The PhoneChannel object continuously removes audio from this buffer and passes it to the DataChannel object. The audio at this stage is in raw PCM format and is encoded by the codec object attached to the DataChannel object. The encoded audio is then transmitted within an RTP stream to the H.323 terminal on the IP network where it is decoded by the codec object attached to DataChannel object in the H.323 terminal. Once decoded, the audio can be presented to the user by sending it to the output port of the soundcard.

The modifications required to add Teldem support to this gateway are as follows:

- Incoming audio from the Teldem on the ISDN BRI line is not standard voice data, but rather a sequence of characters encoded and modulated as a series of tones. These tones need to be demodulated and decoded into the series of characters they represent before being sent to the telephone gateway.
- Text messages arriving from the telephone gateway need to be encoded character by character, using Baudot encoding, and modulated to produce the tones that can be carried via the PSTN to the terminating Teldem device.

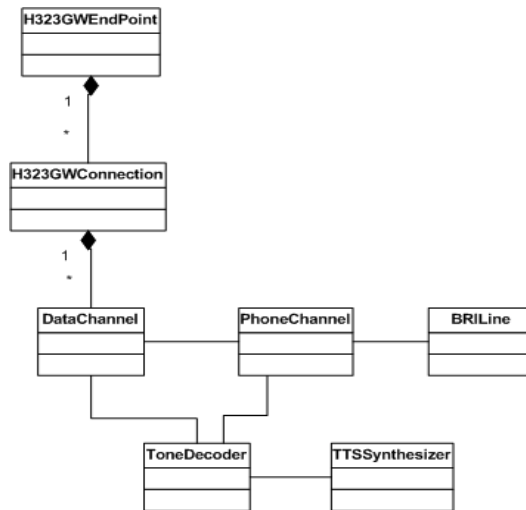


Figure 4. UML representation of modified H.323/ISDN Gateway (Teldem Gateway)

Figure 4 illustrates the decoding portion of the Teldem gateway. This portion of the gateway is implemented with the addition of two specialized classes, the ToneDecoder class and the TTSSynthesizer class. The ToneDecoder class is defined

¹ For simplicity the attributes and operations of the various classes are omitted unless relevant to the discussion.

as a thread that continuously attempts to fetch audio tones from the PhoneChannel (audio received from the Teldem via the BRILine object). The TTSSynthesizer class encapsulates the TTS system. We are currently using the Festival TTS system (www.cstr.ed.ac.uk/projects/festival). The TTSSynthesizer class wraps standard functions around the TTS system, allowing us to substitute other TTS tools in the future. The TTSSynthesizer class has methods that allow the ToneDecoder class to send it text strings that are synthesized to create voice message equivalents.

The tones received in the ToneDecoder object are demodulated and decoded to obtain the characters sent by the connected Teldem. A series of these characters is collected in the ToneDecoder object until a sentinel is received. We defined the sentinel as the uppercase character sequence "GA" for 'Go Ahead' (because this is how turn-taking is traditionally regulated during text telephony). Once the sentinel is received, the ToneDecoder object calls on the TTSSynthesizer object to synthesize a voice message equivalent of the received text string. This voice message is passed to the DataChannel object where it is ultimately sent to the telephone user via the telephone gateway.

We have implemented this part of the overall bridge. We expect that the remaining encoding portion of the Teldem gateway will be similar to the decoding portion described so far. The differences are that the TTSSynthesizer class is replaced with a analogous STTRecognizer, and the ToneDecoder class is replaced by a ToneEncoder class. The STTRecognizer class wraps STT tools just as the TTSSynthesizer class wraps TTS tools. The ToneEncoder class is defined as a thread that continuously checks for text strings arriving from the telephone gateway (the text equivalent of the speech received from the telephone user). These strings are then encoded character by character using Baudot encoding before they are modulated as a series of tones that can be sent to the Teldem via the PSTN. As stated earlier, the poor state of TTS technology to date limits the usability of this portion of the bridge. We conclude our technical discussion with a brief overview of the ToneDecoder class.

IV. DECODING TELDEM TONES

The Teldem communicates across the telephone lines using Frequency Shift Keying (FSK) modulated signals. These signals can be found in the High Frequency (HF) bands and are commonly known as Radio-Tele-TYpe (RTTY). In RTTY signals, a given frequency is used to transmit a logic '1' and a different frequency is used for a logic '0'. We call these frequencies MARK & SPACE frequencies, respectively. In the Teldem, the MARK frequency is 1388.25Hz and the SPACE frequency is 1815.4Hz. The data format is similar to those found in asynchronous serial communications: every character starts with a START bit with '0' logic value, followed by the data bits, least significant bit first, and finally one STOP bit with a logic level '1'. Inter-character time is filled with logic '1'. The character data bits are 5 bits long, and are obtained using an old encoding called BAUDOT.

The decoding process of these signals can either be done using hardware or software. Naturally, we used software decoding within the Teldem gateway. Software decoding of these signals is complex and beyond the scope of this paper.

V. FUTURE WORK

This prototype opens the way toward a fully functional automated voice/text relay for the Deaf. Future work includes building the telephone gateway for the bridge, porting the bridge to SIP and conducting user trials.

C. Telephone Gateway

We have built a pre-prototype for the telephone gateway that accepts voice from a microphone connected to a soundcard, streams the voice to an STT engine and sends the resulting text across an IP connection to a remote endpoint. The STT quality was disappointing, and we turned our attention to the Teldem gateway in anticipation of improvements in the STT quality. We expect the telephone gateway implementation to be quite straightforward, as the architecture closely resembles the Teldem gateway, without the complication of tone decoding. The eventual implementation will allow for "plug and play" of various STT tools.

D. SIP Port

While H.323 has proven to be a suitable deployment environment for this project, the Session Initiation Protocol (SIP) is quickly emerging as a rival standard. We would like to expand the connectivity in the IP world to SIP as well as H.323 users. The SIP implementation should be much simpler than the H.323 system, and can reuse much of the same code and architecture, especially the Teldem tone decoding functionality. The SIP port is currently underway at UWC.

There is an additional mechanism to enable SIP interoperability – by using a SIP/H.323 signalling gateway. Several gateways of this type are in development around the world, including one locally at UWC [6]. Because the text and voice bridging is done by the H.323 gateways, the SIP user agents are not required to even know that they are not conversing with other SIP agents.

E. Usability Trials

Despite the excitement of the technical innovation implicit in this application, the bridging system is of little use unless it is actually useful to the Deaf community. Recall that this bridge is the last and most complicated of our intended Deaf telephony bridges. [2] Telgo323 actually encompasses most of the other bridges, either by automating human relay responsibilities with TTS and STT technologies, or by providing seamless connectivity to both PSTN and Internet phones alike. Once the telephone gateway portion of the Telgo323 bridge is completed, it will be trial tested within a Deaf community setting. These trials necessitate a comparison between a traditional human relay environment and an automated relay bridge. In reality, the ultimate take-up of such a service (and the billing that can accompany it) depends on showing that the Deaf community will actually benefit from the synchronous connectivity enjoyed by the hearing world. One of the major threats to usability is the poor literacy levels

in the Deaf community. If the output of the STT is also of poor quality, it is less likely that Deaf people will reap the benefit of the expansion of their connectivity circle.

VI. BRIDGING THE DIGITAL DIVIDE

At first glance, the Telgo323 is simply a bridge that enables Deaf people to communicate synchronously with hearing people on the PSTN and Internet. This is true, but actually Telgo323 represents much more. In essence, the prototype demonstrates the ability to bridge across various types of Divides. The most obvious bridge is between Deaf and hearing people. The next obvious bridge is between voice and text. Another bridge is between the PSTN and the Internet. Yet another is between high and low bandwidth networks, because the data transport between gateways could be completely text, and not voice.

All this bridging means that anybody, hearing or deaf, with only access to the PSTN can actually participate in the ever-growing opportunities provided by the Internet. Thus millions of people without PCs can actually harness the Internet affordably for synchronous interpersonal communication today with their landline or cellular handsets.

However, issues of bandwidth, Quality of Service, endpoint capabilities (hardware and software) and billing will continue to manifest disparate access – in other words, Digital Divides. Therefore, the emphasis of building bridges over these Divides must remain on scalable design rather than design for specific technologies. For example, the Telgo323 scales in TTS and STT software without having to implement that software. Rather, the Telgo323 wraps the functionality within the overall architecture. We believe this approach will enable us to continue building bridges over Digital Divides as they continue to emerge in the future.

REFERENCES

- [1] M Glaser, "A Field Trial and Evaluation of Telkom's Teldem Terminal in a Deaf Community in the Western Cape ", Proceedings of the South African Telecommunication Networks and Applications Conference (SATNAC 2000), Stellenbosch, South Africa, 11-13 September 2000 (CD-ROM publication).
- [2] M Glaser and WD Tucker. "Web-based Telephony Bridges for the Deaf", South African Telecommunication Networks & Applications Conference (SATNAC 2001), Wild Coast Sun, South Africa. Sep 3-5, 2001, pp. 672-677.
- [3] ITU-T Recommendation H.323 (2000) "Packet Based Multimedia Communications Systems"
- [4] ITU-T Recommendation H.225.0 (03/01) "Call Signalling Protocols and Media Stream Packetization for Packet Based Multimedia Communications Systems"
- [5] JB Penton, A Terzoli and EP Wentworth. "Deploying a Feature-rich H.323 Environment in which to practice the creation of Additional Services", South African Telecommunication Networks & Applications Conference (SATNAC 2001), Wild Coast Sun, South Africa. Sep 3-5, 2001, pp. 701-705.
- [6] M Jeffries and WD Tucker. "An Interoperable Signaling Solution between SIP and H.323", South African Telecommunication Networks & Applications Conference (SATNAC 2001), Wild Coast Sun, South Africa. Sep 3-5, 2001, pp. 689-693.

BIOGRAPHIES

Jason Penton (j.penton@ru.ac.za) is a Telkom Masters student, finishing up his Masters degree in Computer Science at Rhodes University. His research addresses building H.323 application services for the Telkom/DiData/Lucent/THRIP Centre of Excellence in Distributed Multimedia.

William (Bill) Tucker (btucker@uwc.ac.za) leads the Broadband Applications and Networks Group (BANG!) at UWC's Computer Science department. He is a member of the Telkom/Cisco/THRIP Centre of Excellence in IP and Internet Computing. Bill is currently doing a PhD in Computer Science at UCT addressing Quality of Interaction in disparate IP-based bridging scenarios.

Meryl Glaser (mglaser@uctgsh1.uct.ac.za) is a lecturer in the Division of Communication Sciences and Disorders at UCT. Her research interests currently span sign language and deaf telephony. She is a member of UCT's Telkom/Siemens/THRIP Centre of Excellence in ATM & Broadband Networks and their Applications, and works closely with UWC's South African Sign Language (SASL) group.