Convivo Communicator: an Interface-Adaptive VoIP System for Poor Quality Networks

Marco A. Escobedo, Michael L. Best

Abstract— **Convivo is a VoIP system designed to provide reliable voice communication for poor quality networks, especially those found in rural areas of the developing world. Convivo introduces an original approach to maintain voice communication interaction in the presence of poor network performance: an Interface-Adaptation mechanism that adjusts the user interface in order to reduce the impact of high latency and low bandwidth networks. Interface modes facilitate turn taking for high latency connections, and help to sustain voice communication even with extremely low bandwidth or high error rates. An evaluation of the system, conducted in a rural community in the Dominican Republic, found that Interface-Adaptation helped users to maintain voice communication interaction as network performance degrades. Transitions from full duplex to voice messaging were found particularly valuable. Initial results suggest that as users get more experience with the application they would like to manually control transitions based on feedback provided by the application and their own perceived voice quality.**

I. INTRODUCTION

Convivo[1] is a Voice over Internet Protocol (VOIP) computer application designed to support effective voice communication over the Internet even in the presence of poor quality networks, especially like those found in rural areas of the developing world.

A. Ethical status of communication

Article 19 of the Universal Declaration of Human Rights declares that, "[e]veryone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers" [1]. Indeed, this obligation has been further articulated internationally, expressing the right of all people to communicate freely and effectively, in instruments such as the European Convention for the Protection of Human Rights and Fundamental Freedoms [2], the UN ACC Statement on Universal Access to Basic Communication and Information Services [3], and the African Charter on Human and Peoples' Rights [4]. If free and effective communication is a human right then it is true, *ex hypothesi,* that information tools, and indeed the Internet, go to human rights concerns. Thus, as the

Internet is today's most powerful communication tool, provisioning of the Internet has become an *ethical* concern.

The deployment of the Internet has been shown to be a tool for such critical human issues as democracy and empowerment. For instance RAND researchers have shown that, controlling for economic development, the level of Internet connectivity is a strong predictor of levels of democratic attainment [5].

The specific case of telephone style communication, whether over the Internet or over traditional voice circuits, has been shown to be a fundamental tool for the economic and social development of rural communities in developing countries [6,7]. Nevertheless, telecommunication penetration is extremely low in these regions. For example, while there is more than one phone line for every two people in OECD countries, there is just one for every fifteen in developing countries—and one for every two hundred in least developed countries [8].

Happily, throughout the world people are beginning to understand the advantages of providing access to communication and information technologies. As a result, the Internet is rapidly expanding in developing countries. The expansion in these countries includes not only urban cities but also remote and isolated rural communities.

The Internet may help people to attain their development goals, but it must be used as a communication process tool [9]. For instance, rural communities with access to the Internet can benefit from new types of services such as VoIP, which has emerged as a low-cost alternative to the Public Switched Telephone Network (PSTN) voice service. Providing VoIP services in rural communities helps to satisfy one major community need: communication.

B. Technical challenges

VoIP requires a minimum quality of service (QoS) from the network to sustain a rapid delivery of voice packets and maintain voice quality. The current *best-effort* Internet model does not provide a VoIP service comparable to traditional phone service. Over the last years, significant advances have been made to achieve VoIP systems of good quality [10,11,12,13]. Thus, it is usually possible to achieve good QoS with the favorable network quality experienced in OECD countries.

However, network connections available in rural developing areas are not so favorable for VoIP service due to their frequent high latency and low bandwidth. For example, in rural India 85% of copper lines cannot sustain rates higher than 8

---

[1]Convivo's name was inspired by the Spanish verb "convivir" which means to coexist, to share.

Kbps. Wireless and satellite links are commonly used to bring Internet connectivity to such remote rural regions. These satellite connections increase the delay significantly, in the range of 250-275 ms [14]. In some countries, all national traffic goes through one satellite link. As a result, the bandwidth available per Internet connection in the country is limited. Achieving good QoS under these stringent network conditions is a challenge.

Limited bandwidth and delay have a differing level of negative impact on different forms of online communication [15,16]. For example, instant text messaging is a relatively robust mode in the presence of low bandwidth and high delay. In contrast, video and audio conferencing require far more bandwidth and are impacted by latency to a much greater degree.

This paper presents the Convivo system, a VoIP application designed to provide reliable voice communication over poor quality networks, especially those found in rural areas of the developing world. Convivo introduces an original approach to maintain voice communication interaction in the presence of poor network performance: an Interface-Adaptation mechanism that suggests adjusting the application user interface to one of three different voice communication modalities (full duplex, half duplex, and instant voice messaging).

This paper claims that changes in communication modality is an effective option to sustain voice conversation despite poor network performance. Interface-Adaptation reduces the impact of high latency and low bandwidth on voice communication interaction, facilitates turn taking for a high latency connection, and helps to sustain voice communication given extremely low bandwidth or high error rates. Additionally, Convivo dynamically adapts to the performance of the network by means of an application-level QoS control that maximizes bandwidth and minimizes the end-to-end delay.

Recognizing that a large number of people in developing countries are illiterate, Convivo presents a fully graphical interface of the voice conversation. Icons help to maintain user interaction, presence and awareness.

II. RELATED WORK

Today, a large number of commercial VoIP software applications are available. Among the most popular software is Microsoft's NetMeeting, OpenPhone, Speek Freely, Yahoo Messenger, Robust Audio Tool, Free Phone, and CUSeeME. A large list of VoIP software and their characteristics can be found in [17]. Most of these audio tools provide telephone-quality speech in the presence of good network performance. Most of them perform quality adaptation applying different techniques to achieve the best quality of service possible. Codec selection, activating silence suppression, adaptive playout adjustment, redundancy mechanism, error concealment, dynamic jitter buffers, rate control mechanisms, and echo cancellation are some of the techniques applied to improve voice quality. Nevertheless, high QoS is not always achieved.

Microsoft's and Yahoo's messengers are communication applications that support different communication modalities. Instant text messaging is their principal modality, but they also provide the option to establish voice or video communication. Users decide when and which communication modality to use.

We are particularly interested in the application of VoIP to rural and developing areas. Little work has been done in the area of software development for rural communities and in software evaluation methodologies for rural communities. A unique example is Community Knowledge Sharing, an Internet-enabled asynchronous messaging system designed for use in the developing world [18]. CKS implements a multi-literate design in which the system can be customized based on the abilities and preferences of the user. CKS provides one of the first examples of an appropriate messaging system for the developing world.

The evaluation of CKS was conducted in Bohechio, Dominican Republic. It concluded that information technologies for use in the developing world should include iconic interfaces for low literate users. Developers need to collaborate with communities and local partners to create a useful and meaningful application for the community. Moreover, the study suggests extending the multi-literate approach of CKS to a direct messaging environment to support low-cost voice and text messaging.

III. SYSTEM DESCRIPTION AND ARCHITECTURE

A. Design Principles

Convivo Communicator enables real-time voice communication between two parties, integrating a multi-literate user interface. Convivo makes use of the fundamentals of the Real Time Protocol (RTP)[19] and the Session Initiation Protocol (SIP) [20]. These protocols specify how audio and video can be transported over the Internet, and what kind of signaling is necessary to communicate between terminals and different entities on the network. Convivo uses RTP for audio encapsulation and the Real Time Control Protocol (RTCP), for QoS reports on the receiver's quality of reception. In addition, Convivo uses the call setup protocol of SIP.

The design of Convivo is based on two major principles. First, provide reliable voice communication in the presence of poor network performance. Second, provide a multi-literate graphical user interface (GUI) to support users with limited literacy and computer skills, to facilitate turn taking and to maintain presence awareness given poor network performance. Convivo employs an end-to-end (peer-to-peer) model.
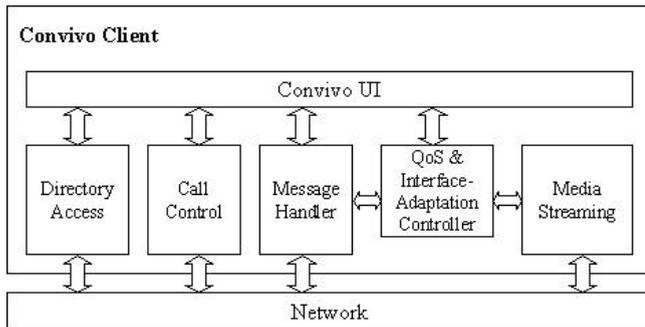
B. Overall Architecture

Figure 1 - Convivo client architecture

Convivo client is composed of six main components illustrated in Figure 1:

1. A media streaming module which handles the audio delivery, collection and playing;
2. A control module that manages the Interface-Adaptation mechanism and provides dynamic QoS adaptation;
3. A message-handler module that allows Convivo clients to exchange messages;
4. A call control module to allow for setup and shut down sessions;
5. A directory subsystem that allows users to identify and address people for their communications (not described in this paper);
6. A graphical user interface.

These modules were implemented in the Java programming language. The media streaming module and the call control module are based on the Java Media Framework API package and the Java SIP implementation respectively. The other four modules are original implementations written exclusively for Convivo.

IV. GRAPHICAL USER INTERFACE AND COMMUNICATION MODES

The graphical user interface (GUI) was designed to handle the three different voice communication modalities: full duplex, half duplex, and voice messages. The challenge was to design an interface that is simple to use to support users with a range of literacy levels and computers skills. Moreover, due to the fact that voice communication interaction is adversely affected by poor network performance, the interface has to maintain presence awareness, facilitate turn taking, and provide feedback on network status and user actions.

The user interface of Convivo is divided in two main areas: the graphical conversation area, and the controls area (Figure 2).

Figure 2 - Convivo User Interface

The graphical conversation area displays a visual representation of what is happening during the call. For example, the area shows users when the other conversant is talking, listening, recording, or playing a voice message. Users are represented with an icon and a self-portrait appears on top of this icon. From the point of view of a user, the left hand side of the graphical conversation area is related to the local user (him/herself) while the right area corresponds to the person with whom the user is communicating. The controls area holds the buttons and controls that are required for each communication modality. Also, the volume control and the network status indicator are located in this area. The indicator gives the user a graphical feedback that reflects current network performance. It represents the network link between the Convivo clients participating in the call. Yellow squares, representing IP packets, traverse the bar constantly. The speed with which these squares move is relative to the current end-to-end delay. The number of squares traversing the indicator is related to the current bandwidth available.

Convivo's user interface supports three different forms of voice communication: full duplex, half duplex, and instant voice messaging. The full duplex modality is similar to a telephone call where two persons can speak simultaneously. In contrast, in the half duplex modality only one person can speak at a time, a button needs to be pushed and held down to talk and released to stop talking. Half duplex modality is similar to communicating using a "walkie-talkie" radio. The instant voice messaging modality is similar to text-based instant messaging, two persons are online at the same time and they converse exchanging voice messages.

1) Full duplex

The full duplex modality is simple and most similar to a telephone call. Users can speak simultaneously and do not need to use any controls during the conversation.

The controls area only has the volume control and network status indicator. No other buttons are required to use the full duplex modality.

The conversation area presents both users as if they were talking using a phone. The phones are in green. This represents that both users can speak any time. Also, when one of the participants speaks an animated icon representing voice

waves is shown besides the corresponding face.

2) Half Duplex

The half duplex modality is similar to communicating using a "walkie-talkie" (or a two-way radio). Only one person can speak at a time, a button needs to be pushed and held down to talk and released to stop talking.

When Convivo operates in this modality, a push-to-talk button appears in the controls area. When one user wants to speak this button needs to be pushed and held down to talk and released to stop talking. The push-to-talk button is enabled only on the interface of the user who holds the turn to talk.

Turn taking is controlled by an internal mechanism based on exchanging messages between Convivo clients whenever the push-to-talk button is released. When the push-to-talk button is released, a message is sent to the other client to pass the turn to talk to that user. A Convivo client will hold the turn to talk while the user keeps the push-to-talk button pressed. Once it is released, then the turn is exchanged to the other client. By default, the user who started the call has the first turn to talk whenever a transition ends in the half-duplex modality. The turn-taking scheme is reflected in the graphical conversation area. When nobody is talking, both phone icons are grayed out. If the user who holds the turn to talk presses the push-to-talk button then the phone icon associated with him turns to green color as well as animated green voice waves appear beside his face representation. Similarly, the phone icon associated with the other user turns to red color and animated red voice waves appear besides his face representation.

3) VoiceMessaging

The instant voice messaging mode is similar to a text-based instant messaging. Three buttons are included in the controls area for this modality: StartRecording button, Stop&Send button, and PlayLastMessage button. When a user wants to compose and send a message, they have to follow two simple steps. First, they press the StartRecording button and record a message. Second, when finished recording the user pushes the Stop&Send button to stop recording and send the message to the other user. When a user receives a voice message it is played automatically. If the user wants to hear it again they can press the PlayLastMessage button.

The sequence of recording, sending, receiving and playing a message is represented in the conversation graphical area. Animated icons help users know when these actions occur.

V. QoS & INTERFACE-ADAPTATION

Quality of Service (QoS) is the main issue for VoIP. The current best-effort service available on IP networks is not sufficient to provide VoIP service comparable to the traditional Public Switch Telephony Network (PSTN) service in reliability and voice quality.

Three main influences define the QoS of VoIP: the performance of the VoIP device, the performance of the underlying network, and the end-to-end delay [7]. A particular quality level can be achieved by trading off one against the other. However, for high-quality systems the scope for trade-offs is greatly reduced. An in depth overview of VoIP QoS issues can be found in [11].

The QoS controller is responsible to achieve the main goal of Convivo: provide reliable voice communication and sustain voice communication interactions in the presence of poor network performance. To achieve this goal, Convivo uses a dynamic QoS adaptation control mechanism based on two sub-controls. The first control provides "backend" adjustments, such as codec selection, in an attempt to minimize delay and maximize speech quality. We call this the QoS Adaptation controller. The second control sustains voice communication interactions through "frontend" adjustments to the user interface presenting one of three modalities: full duplex, half duplex, and instant voice messaging. The latter control is referred as the Interface-Adaptation mechanism.

A. Dynamic QoS Mechanism

The dynamic QoS adaptation control mechanism that Convivo uses is analogous to a feedback control system. RTCP reports provide statistics about the current network transmission performance and the actions conducted by the module are based on the real-time statistical analysis of the following parameters: latency, bandwidth, jitter, and packet loss rate.

The Network Status Monitor collects the RTCP reports and obtains the value of the network parameters (latency, bandwidth, jitter, packet loss rate). It then sends this data to the QoS and Interface-Adaptation controllers.
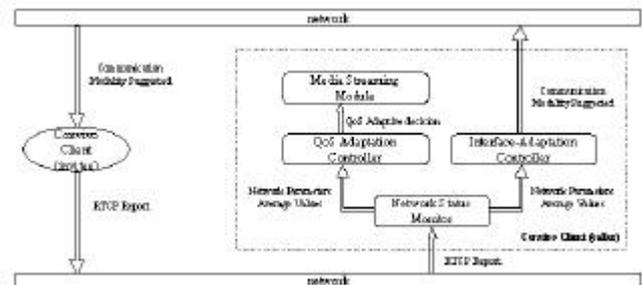

Figure 3 - Dynamic QoS and Interface-Adaptation flow chart

B. QoS Adaptation Controller

The QoS controller aims to maximize speech quality and minimize the end-to-end delay. Maximizing speech quality is related to speech coding, packetization efficiency, silence suppression, error concealment, jitter buffer, delay, echo, and speech level. All these factors contribute to the perceived speech quality. The QoS controller is responsible for codec selection, silence suppression activation, and jitter buffer adaptation.

Principally, the choice of codec in a VoIP system is dictated by the bandwidth available on the network path between call participants. The bandwidth determines the maximum bit rate

for the codec, and in consequence the maximum speech quality that the system will achieve under ideal conditions. A secondary factor that influences codec selection is the latency to encode. Three codecs, G.723.1 [21], 6.729 [22], and G.729A [23], have been designed to work well in presence of a bandwidth-, delay-, and loss-constrained environment. Table 1 lists the main characteristics of these codecs. Bit rate refers to the output bit rate of the encoder. Frame size is the length of the voice signal compressed into each packet. The lookahead delay is the amount of the next frame the coder uses to encode the current frame in order to take advantage of correlation. Frame length is the number of bytes in the encoded frames. DSP MIPS is the minimum required processor speed, in millions of instructions per second. The required RAM for each encoder is given in 16 bit-words.

Codec	6.723.1	G.729	G.729A
Bit rate (kbps)	5.3/6.4	8	8
Frame size (ms)	30	10	10
Processing delay (ms)	30	10	10
Lookahead delay (ms)	7.5	5	5
Frame length (bytes)	20/24	10	10
DSP MIPS	16	20	10.5
RAM	2200	3000	2000

TABLE 1 – CODEC BIT RATES, DELAYS, AND COMPLEXITY [24]

From Table 1, it can be seen that while G.723.1 supports the lowest bit rate, it also adds the largest processing delay. G.729 encodes at higher bit rate and requires more CPU resources, but has a lower latency. Convivo selects from one of these three codec on a per-call basis to trade between bandwidth and quality. At the beginning of each call, the Convivo Network Status Monitor runs a test to measure the bandwidth available and then selects the codec to be used during the call.

Silence suppression is the process by which periods when a user is not speaking are not coded or transmitted. The purpose of activating silence suppression is to reduce the bandwidth requirements for a VoIP call. The QoS controller activates Silence Suppression only when Convivo clients are communicating in full duplex mode, which can produce an average bandwidth saving of some 40% [12].

In the Internet, packet delay is highly variable. Therefore, VoIP applications implement a jitter buffer to allow for the proper sequencing of packets that may arrive out of order. This buffer has an impact on the delay and speech quality. The buffer holds incoming packets for a specified amount of time before forwarding them to decompression. This has the effect of smoothing packet flow, however the buffer can add significant delay. Some systems have fixed-length jitter buffers and others dynamically adjust the buffer size.

Convivo implements a jitter buffer that is adjusted dynamically upon the reception of an RTCP report. The jitter buffer depth has to be selected to balance between delay and quality. If the jitter buffer is too small, network jitter and packet loss will cause audible effects in the received voice. If the jitter buffer is too large, voice quality will be fine, but the delay will be higher. The network status monitor obtains the network jitter value from the RTCP reports. The QoS Adaptation controller manages the jitter buffer storing only the required amount of buffered voice traffic.

C. Interface-Adaptation Controller

The Interface-Adaptation controller determines which communication modality to be used between Convivo clients according to current network performance. In order to do so, the controller analyses the data generated by the network status monitor. After the analysis is completed, the controller selects one of the three different voice communication modalities: full duplex, half duplex, or voice messaging. Each one of these forms of communication has different network requirements to provide good speech quality and the suggested modality is intended to be the best performing given the network properties. A transition from one communication modality to another may happen only if the suggestion is different from the current communication modality, according to the "transition mechanism".

1) Interface-Adaptation Principles

The concept of Interface-Adaptation is based on the premise that any communication modality can be characterized by its required bandwidth, end-to-end delay, and packet loss.

a) Mapping from Network parameters to QoS

A mapping scheme from unidirectional delay and packet loss to QoS presented in [25] is shown in Figure 3. A range of QoS can be obtained based on the ITU Recommendation G.114 and codec properties. The range goes from excellent to poor quality. Excellent quality is for round-trip delays of less than 200 ms and low packet loss rates. Good quality is for round-trip delays from 200-300ms and higher loss rates. Limited quality is for round-trip delays from 300-800ms and high loss rates. Beyond round-trip delays of 800ms and very high loss rate, quality is extremely poor, not useful for synchronous VoIP Telephony. The mapping assumes a codec with good loss concealment algorithm; otherwise loss rates 5-10 percent would be enough to result in poor quality.
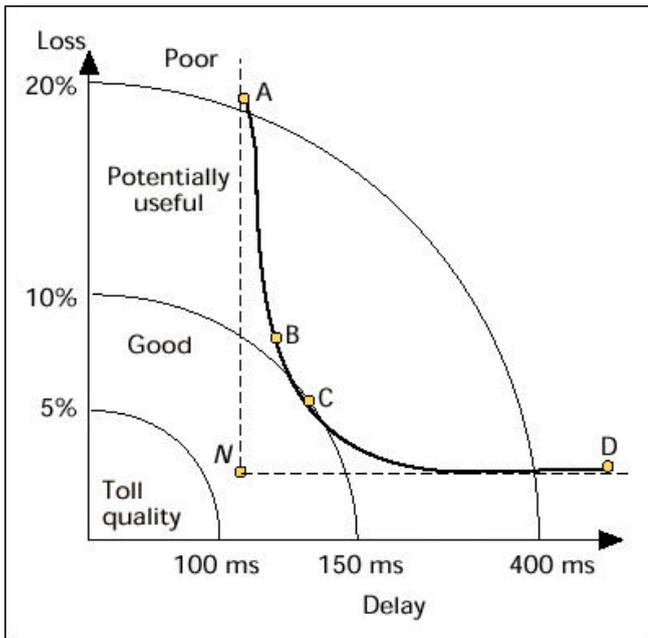
Figure 4 - Mapping unidirectional delay and packet loss to QoS

Point N in Figure 4 represents the total delay and loss that is experienced due to "network" factors, including all delays and losses up to the receiver buffer. A VoIP solution will never operate at this point, because codec and buffer delay also have to be considered. Point B represents a receiver operating under a certain amount of delay and loss. Part of this delay is due to the size of the receiver's buffer. Increasing the buffer size will bring us to point C, which increases the delay and decreases loss rate. This decrease in loss is due to the fact that by pushing back our playout point we are not dropping as many packets. If the buffer keeps increasing toward point D and on to infinity, we reach the "network" loss rate since we no longer drop additional packets. On the other hand, if we decrease the buffer size toward point A and onward to a zero buffer length, we approach the "network" delay. However, the loss rate approaches 100 percent because packets must arrive exactly at the playout point.

Available bandwidth can be added as another dimension to this mapping scheme. Bandwidth defines the threshold that determines if a synchronous real-time voice conversation is possible using VoIP. Below this threshold value (14kbps), good or poor quality of service is not the right classification; basically real-time voice interaction is not feasible. This fact opens the door for a new mapping scheme that relates bandwidth, delay, and packet loss to different forms of communication interaction.

b) Communication Modalities Mapping

Mapping communication modalities to network performance is the basic idea introduced in this paper. Convivo adapts communication modality according to the current network performance. Any communication modality can be characterized by its required bandwidth, end-to-end delay, and packet loss. For example, traditional telephony requires a large amount of bandwidth (64 kbps), low packet loss, and shows low delay. Videoconferencing demands high bandwidth allocation, low end-to-end delay, and moderate packet loss. In contrast, text instant messaging demands very low bandwidth, packet loss is not an issue because TCP protocol is used for transmission, and end-to-end delay can be high because interaction is "semi-asynchronous".

Internet audio applications, such as VoIP Telephony, lie in the region of low to moderate bandwidth and intermediate delay. A VoIP call, full duplex conversation, requires at least 14kpbs in bandwidth, end-to-end delay less than 400 ms, and packet loss rate less than 5%. The ITU-T Recommendation G.114 provides limits for one-way transmissions time, a delay of 400ms or more is unacceptable for most users [26]. Most people notice delays when they exceed 250ms. To achieve the highest quality, one-way latency should be less than 150 ms. User studies indicate that telephony users find that round-trip delays greater than 300 ms feel more like a half-duplex connection than a conversation [26]. The studies showed that user tolerance of delay varies significantly from user to user. The most tolerant users were satisfied with delays of 300-800ms, while most critical users required delays of 200 ms. Packet loss rate is a real problem when percentage of lost exceeds a certain threshold (roughly 10 %), or when loss bursts occur frequently. Even the best codecs will be unable to hide the packet loss from the user, resulting in degraded voice quality, or in worst cases making real-time voice conversation impossible. In voice communication interaction, when delay is greater than 300ms conversation conforms to a half-duplex modality. Another voice communication modality is instant voice messaging, which can be an option for voice communication when delay is extremely high (< 800 ms), when packet loss rate is notably high or when not enough bandwidth is available.

A range of communication modalities can be considered based on the previous analysis. Convivo proposes to map communication modalities to these network characteristics.

Figure 5 is similar to Figure 4 except that when we move from point to point we are also changing communication modalities. Figure 5 shows the mapping from delay and packet loss to voice communication modalities (full duplex, half duplex, and instant voice messaging).

Bandwidth available can also be added in this mapping scheme creating a modality space. The space is divided in two. The upper space corresponds to the full/half duplex modalities and the lower space corresponds to voice messaging modality.
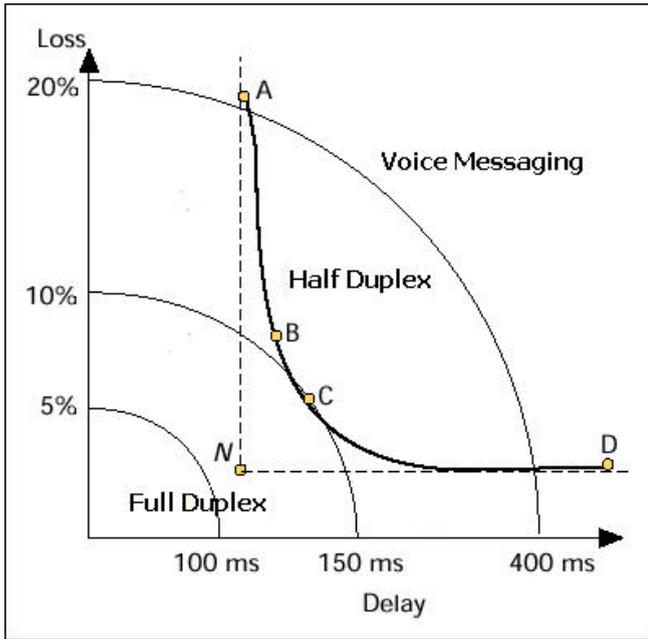
Figure 5 - Mapping communication modalities

To determine which communication modality to be used by Convivo clients, the Interface-Adaptation controller looks at the mapping shown in figures 3 and 4. The Interface-Adaptation controller suggests:

- full duplex:
    If (round-trip delay < 400 ms) and (bandwidth > 14 kbps) and (packet loss rate < 10%)
- half duplex:
    If (round-trip delay > 400 ms) or (bandwidth > 14kbps) and (packet loss rate < 20 %)
- instant voice messaging:
    If (latency >> 400ms) or (bandwidth < 14kbps) or (packet loss rate > 20%)

The thresholds were selected based on the ITU recommendations, research conducted in the area of Internet dynamics [27,28,29,30,31,32], and our own research.

D. Interface-Adaptation Mechanism

The Interface-Adaptation mechanism follows a master-slave model. The caller is the master while the other participant is the slave. In this model, the master analyses network performance and makes suggestions, i.e. only the caller's Interface-Adaptation controller is active during the session. The caller's suggestions are sent to the slave. If a change in communication modality is necessary both participants will experience a switch in their graphical user interface. This master-slave model is applied to avoid two Interface-Adaptation suggestions made almost at the same time by two different Convivo clients, generating oscillations in the communication modality being presented to the user. The Interface-Adaptation controller concludes what

communication modality to use every thirty seconds.

The transition mechanism is the scheme followed when Interface-Adaptation controller suggests a change in the user interface. This mechanism can operate in three different modes:

- Automatic: communication modality changes in both peers automatically.
- Semi-automatic: Convivo prompts both users to accept or reject the transition suggested. A transition occurs if any of the users accepts the change.
- Manual: users have the control to select the communication modality to be used.

VI. EVALUATION

Convivo was evaluated by means of a user study focused on three areas. First, the study evaluated the Interface-Adaptation mechanism assessing if the mechanism helped users to communicate when network performance was poor or varying. Second, it studied what communication modalities users preferred given specific network conditions. Thirdly, the study evaluated the system's usability and the user interface to determine if design principles and interface features were appropriate for the target communities. The study encouraged community members to collaborate on improving Convivo's user interface.

In March 2002, Convivo was tested via a user study in Bohechio, a rural agricultural village in the Dominican Republic, where a Little Intelligent Communities (LINCOS) [33] community telecenter exists. The evaluation methodology and instrument were based on recommendations obtained from previous research conducted at Bohechio [18, 34], and from existing methods of evaluating user interfaces [35]. A hands-on conversational experience with Convivo gave community members the opportunity to test the application and to express their opinions about it. A collaborative design session was conducted to obtain the community's design ideas and desired features. These ideas have helped developers improve the application's functionality and user interface to meet the realities and needs of users from rural communities like Bohechio.

During the user study participants used Convivo to communicate with members of the research team. Both the subject and the researcher had a computer linked to a third intermediate computer that emulated varying network conditions. These varying conditions triggered the Interface-Adaptation mechanism producing changes to the graphical user interface. Thus, participants had the opportunity to use all the voice communication modalities available in Convivo.

VII. RESULTS

In total, twenty-five people participated in the user study. Overall, participants were satisfied with the transition mechanism. Fourteen of them expressed satisfaction because the system helped them to communicate as the network

performance degraded. Only four participants mentioned that they do not want any transition at all. Seven of the participants did not know or did not express any level of satisfaction.

Transition Mechanism operation is:	Satisfactory	Not Satisfactory	Neutral
# of Participants	14	4	7

TABLE 2 – SATISFACTION LEVEL OF THE TRANSITION MECHANISM

Participants were asked to express whether transitions were necessary or not. Eleven participants thought that transitions were necessary due to a poor perceived quality. Only three participants thought that transitions were not necessary.

Transitions are:	Necessary	Not Necessary	Neutral
# of Participants	11	3	9

TABLE 3 – ARE TRANSITIONS NECCESARY?

Most participants agreed that Convivo should perform a change of the communication modality as quality varies. This change should happen immediately when network performance varies according to nineteen participants. Only four participants suggested that the Interface-Adaptation mechanism must wait at least 30 seconds before adapting the interface to check if network performance is still poor.

Participants had the chance to experience Convivo operating with automatic and semi-automatic transitions. A total of twelve participants were more satisfied with the transition mechanism operating in semi-automatic mode than with automatic transitions. Seven of them thought it was much better than automatic transitions and the other five thought it was better. Only one participant stated that the semi-automatic mode was worse than automatic.

Participants were asked to state what would be their preference on the transition mechanism mode. Six participants preferred automatic mode, eight particpants preferred semi-automatic mode, and nine participants preferred to manually control the Interface-Adaptation mechanism. A significant correlation was observed between computer skills and transition mechanism mode preference ($\chi2 = 13.86$ p=0.03). The data shows that users with more computer skills preferred either a semi-automatic or manual transition mechanism.

	Automatic	Semi-Automatic	Manual
# of Participants	6	8	9

TABLE 4 – FAVORITE TRANSITION MECHANISM MODE

The transition from full duplex to the voice messaging mode was mentioned by sixteen participants as most helpful as voice quality degrades. No other transition was mentioned as helpful, and nine participants did not express any preference. Eleven participants pointed out that transition from full duplex to half duplex did not help at all as the network quality

changed. The rest of the participants expressed no comments about it.

Transitions from:	Full-to-Half	Full-to-VoiceMsgs	Half-to-VoiceMsgs
# of Participants	0	16	0

TABLE 5- MOST HELPFUL TRANSITIONS

All the participants described the full-duplex modality as very easy and simple to use. In contrast, several participants had difficulties using the half-duplex modality. According to the facilitators who assisted participants during the experiment, the push-to-talk button from the half duplex modality was hard to use for almost everybody. From their observations we have the following data: thirteen had difficulties all the time, four showed minor difficulties just at the beginning, and only eight subjects used the button easily.

The main problems experienced with the half-duplex modality were: 1) participants released the button before they started talking; 2) participants thought that after they pressed it the first time, the button would remain active until they finished talking; 3) participants kept the button pressed after they finished talking when they should have released it. The facilitators' opinion was that most of the problems with the push-to-talk button were because it was the first time that participants were using the application, and to some degree, were due to a lack of control over the mouse.

Participants had less problems with the voice messages interface than with the half-duplex. They showed fewer difficulties using the interface. The only pathology detected was that some of them started to use the record button thinking it worked as the push-to-talk button.

All participants had the opportunity to experience the three communication modalities. Nineteen participants preferred the full duplex modality as their favorite form of communication. Six participants chose the voice messaging modality as their favorite. The problem with the half duplex modality was that participants found the push-to-talk button difficult to use. Therefore, given the option they would prefer to use either full duplex or voice messages.

Every participant liked and understood the graphical representation of conversation provided by Convivo. Participants welcomed with enthusiasm the portrait feature of Convivo. They were extremely excited about the possibility to see a picture of themselves and, more importantly, the portrait of the person with whom they were communicating. The portrait feature made the conversation more "familiar" and helped the conversant know more about the person with whom they were talking.

1) Participatory Design Session Results

In order to understand the qualitative requirements and desires of the users in Bohechio, a collaborative design session was conducted during the user study. This session was based on the focus group method. The focus group

consists of a group of volunteer subjects and a group leader, who is a member of the evaluation team.

The goal of the focus group in Bohechio was to collaborate with community members to get opinions and ideas about Convivo's graphical user interface (GUI). Four people took part in this one-hour session. The session was mainly focused on the GUI but it was also open to other issues around Convivo and voice communication needs in Bohechio.

The group agreed that the graphical representation of conversation presented by Convivo is very understandable. They suggested that men and women should be represented differently, but that there should be one single representation for each.

The group questioned what would happen if a user does not have a portrait. The group suggested that Convivo should allow users to select alternative representations if they do not have a portrait. They would prefer to use a toolbar for this purpose.

Another important issue for the group was the visual presence of the application. They mentioned that it was important to design Convivo's GUI with brighter and more attractive colors.

Related to the transition mechanism, the group recommended the use of audio indications prior to a communication modality change. In addition, they would like to be able to configure the level of autonomy with which the Interface-Adaptation mechanism works: automatic, semi-automatic, or manual.

Ideas were presented around the network status indicator, that portion of the interface that graphically describes the quality of the network connection. One subject suggested the use of a representation similar to that on cellular phones for signal strength. Another user suggested that a weather metaphor (sunny day for a good network connection, etc.) would be useful.

Convivo's communication modalities were also part of the discussion. The group acknowledged that for some users it would be difficult to use the push-to-talk button if they do not have any computer skills. Nevertheless, the group thought that users would get used to it. They thought that audio feedback would be useful to indicate a change in turn. No comments were made related to the other two communication modalities.

Unexpectedly, the main topic related to communication modalities was the addition of a fourth modality. The group thought that Convivo might include a text modality as the fourth form of communication available for the user. Two factors were mentioned to support this idea. First, text is an option when network performance is extremely poor because it consumes less network resources. Second, because computers in Bohechio are in a common space with other computers sitting closely, some users would like to have more privacy and would not like other people surrounding them to hear what they are saying. However, the group mentioned that the addition of text modality would have the risk of transforming Convivo into a text instant messenger. Convivo could lose its fundamental characteristic, voice communication. At the end, the consensus of the group was that text modality should be included just as a fourth option only available in the presence of extremely poor network conditions and only for literate users.

Finally, the group suggested the inclusion of multiparty conferencing functionality and stressed the importance of being able to retrieve voice messages that were sent by other users when one is offline.

VIII. CONCLUSION

This paper has presented Convivo, a VoIP application designed to provide reliable voice communication for poor quality networks, especially those found in rural areas in the developing world. The paper introduces an original approach to the maintenance of voice communication interaction in the presence of poor network performance: the Interface-Adaptation mechanism.

Convivo was evaluated by means of a user study in Bohechio, a rural agricultural community in the Dominican Republic. The study investigated the Interface-Adaptation mechanism and Convivo's user interface. The study found that the Interface-Adaptation decisions helped users to maintain voice communication interaction in the presence of poor network performance. The transition from full duplex to voice messaging modality was the most helpful change when network quality degraded, while a change from full to half duplex was perceived as unhelpful. The study found that users with more computer skills would like to manually control transitions based on feedback provided by the application and their own perceived voice quality. Full duplex modality is by far the most appreciated communication modality because is extremely simple to use and similar to the only form of communication modality that people from Bohechio are used to, the telephone. Nevertheless, a few participants mentioned voice messaging as their preferred communication modality. Convivo's user interface was also evaluated. All participants liked and understood the graphical representation of conversation that appears on the interface. In particular, the portrait feature was welcomed enthusiastically.

Convivo is designed to provide voice communication for users in rural areas and the developing world. Core communication services, such as VoIP, are critical applications for rural and developing areas. Expanding this sort of research, into new computer applications and services for the developing world, is an important and under-served research arena.

REFERENCES

[1] UN, "Universal Declaration of Human Rights". Available: http://www.un.org/Overview/rights.html.

[2] European Commission, "European Convention for the Protection of Human Rights and Fundamental Freedoms". Available: http://conventions.coe.int/Treaty/EN/CadreListeTraites.htm.

[3] UN, " ACC Statement on Universal Access to Basic Communication and Information Services". Available: http://www.unites.org/html/resource/acc1997.htm.

[4] OAU, "African Charter on Human and Peoples' Rights". Available: http://www.diplomacy.edu/africancharter/hypertext.asp.

[5] Kedzie, C.R. Communication and Democracy: Coincident Revolutions and the Emergent Dictator's Dilemma. RAND Document No: RGSD-127. Available: http://www.rand.org/publications/RGSD/RGSD127/. 1997.

[6] ITU: "Challenges to the Network: Internet for Development", Geneva, 1999

[7] Richardson, D., Ramirez, R., & Haq, M. Grameen Telecom's Village Phone Programme in Rural Bangladesh: a Multi-Media Case Study. Guelph Ontario, Canada: Canadian International Development Agency, 2000

[8] United Nations Development Programme. (2001). Human Development Report 2001. Retrieved from the World Wide Web: http://www.undp.org/hdro/report.html

[9] Richardson D., "The Internet and Rural & Agricultural Development: An integrated approach", Food and Agriculture Organizations of the United Nations, 1997

[10] Boger Y., "Fine-tuning Voice over Packet services" http://www.protocols.com/papers/voip2.htm

[11] ITU-T Recommendation G.102: 'Transmission performance objectives and Recommendations', Nov. 1988

[12] Reynolds R.J.B., Rix A.W., "Quality VoIP – an engineering challenge", *BT Technology Journal*, vol. 19, no.2; April 2001; p.23-32.

[13] Thorne D., "VoIP – the access dimension", *BT Technology Journal*, vol. 19, no.2; Apr. 2001; p.33-43

[14] Sun Z., Cruickshank H., "Analysis of Voice Conferencing over Geostationary Satellite Systems", IEE Seminar on Satellite Services and the Internet (Ref. No. 2000/017), 2000; p3/1 – 3/5

[15] Cox R. V., "Three New Speech Coders from the ITU Cover a Range of Applications," IEEE Commun. Mag., Sept. 1997

[16] IEEE Communications Magazine, Issues on Internet Telephony, vol. 38, no. 4, Apr. 2000

[17] IP xStream, Voice and Fax Over IP News and Market Research. See http://www.iptelephony.org

[18] Shakeel, H. "Community Knowledge Sharing: An Internet Application to Support Communications Across Literacy Levels" M.S. Thesis, Massachusetts Institute of Technology (2001)

[19] Schulzrinne H. et al, "RTP: a transport protocol for real-time applications', IETF RFC 1889 (1996)

[20] Handlye M: "Session initiation protocol'. IETF RFC 2543 (March 1999)

[21] ITU Rec. G.723.1 "Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s" Mar. 1996

[22] ITU Rec. G.729, "Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP)," Mar. 1996.

[23] ITU Rec. G.729 Annex A, "Reduced Complexity 8 kbit/s CS-ACELP Speech Codec," Nov. 1996.

[24] Cox R. V., "Three New Speech Coders from the ITU Cover a Range of Applications," IEEE Commun. Mag., Sept. 1997.

[25] Kostas T., Borella M., Sidhu I., Schuster M., Grabiec J., and Mahler J. Real-time voice over packet switched networks. IEEE Network, 12(1):18--27, Jan/Feb 1998.

[26] ITU-T Recommendation G.114: 'One-way transmission time', Feb. 1996

[27] Paxson V., End-to-end Internet Packet Dynamics, Proc. ACM Sigcomm '97, (1997)

[28] Mukherjee A., "On the Dynamics and Significance of Low Frequency Components of Internet Load," Internetworking: Research and Experience, vol. 5, 1994, pp. 163–205.

[29] Bolot J.C., "Characterizing End-to-End Packet Delay and Loss in the Internet," J. High Speed Networks, vol. 2, 1993, pp. 305–23.

[30] Sanghi D. et al., "Experimental Assessment of End-to-End Behavior on Internet," Proc. IEEE INFOCOM '93, Mar. 1993, pp. 867–74.

[31] Andren J., Hilding M. and Veitch D., "Understanding End-to-End Internet Traffic Dynamics," submitted to Globecom '98, Sydney, November (1998)

[32] Hagsand, O.; Hanson, K.; Marsh, I., "Measuring Internet telephony quality: where are we today?" *Global Telecommunications Conference GLOBECOM '99*, vol. 3, 1999; p.1838 –1842

[33] Little Intelligent Communities. See http://www.lincos.net

[34] Shakeel, H., Escobedo, M., & Garcia-Romero, B. (2000). LINCOS Bohechio Investigation Results. Unpublised Manuscript

[35] Hoyler A., "Methods for Evaluating User Interfaces" University of Sussex at Brighton, Cognitive Science Research Paper No. 301