Vianix, LLC

MASC

SPEECH PROCESSING

FEB 02, 2001

Technical Contact: Veeru Ramaswamy (757) 321-9971
veeru@vianix.com
Fax: (757) 321-9973
2696 Reliance Drive, Suite #100
Virginia Beach, VA 23452

Administration Contact: Reza Hashampour (757) 321-9971
rh@vianix.com
Fax: (757) 321-9973
2696 Reliance Drive, Suite #100
Virginia Beach, VA 23452

TABLE OF CONTENTS

INNOVATIONS

Vianix, LLC, also referred to as "the Company" in this document, has spent over three years on **internally funded** research and development. These efforts have resulted in new technologies that can significantly advance the state-of-the-art in voice recording.

The Company's proprietary speech processing technology allows higher quality voice to be recorded into less digital memory, enabling significantly increased recording times and more efficient use of available bandwidth. The clarity of the Company's voice technology is unsurpassed at its data rates. This makes the recorded voice easier to listen to and much more intelligible. ***Vianix, LLC's*** proprietary speech processing technology implements a variable rate vocoder that maintains near toll-quality speech at very low data rates. The speech is encoded using a proprietary Codebook Excited Linear Predictive (CELP) speech coding algorithm that achieves high speech quality at low data rates. In variable rate mode, speech is coded at under 3.5 kbps in typical two-way conversations. The vocoder, in variable rate mode, dynamically adjusts the data rate every 20ms based on the measured speech energy in a given 20ms window. Windows with low speech energy are encoded at low data rates and windows with high speech energy are encoded at higher data rates thus maintaining the over all high speech quality. The vocoder can also encode speech at fixed data rates.

This proprietary voice compression technology has been successfully incorporated into five of the Company's prototype products that are in the "Beta" phase of development. **The Company's proprietary speech processing technology, has been designed, built and proven.**

There are several technologies available for digitally recording voice. The most significant limitation to digital voice recording in general is the amount of memory required to store voice of acceptable quality. For example, to reproduce the human voice well you need 3.4 kHz of recording bandwidth. To achieve this you need to sample the voice at 8,000 times per second at an acceptable resolution, generally 8 bits (with companding). This means you have 64,000 bits for every second of recording time. Which translates to 480,000 Bytes per minute or 28,800,000 Bytes per hour. That is a lot of memory for relatively little voice. This may be acceptable in large desktop designs, but in extreme miniature portable designs it certainly is not acceptable. ***Vianix, LLC's*** technology drastically reduces the memory requirements for recording voice while maintaining high voice quality. No other technology today can match ***Vianix, LLC's*** voice quality at its low data rates.

A quality test called the Mean Opinion Score (MOS) is used to rate the quality of different speech codecs. On this scale, a score of 4.0 is considered "toll" quality, the quality of speech heard through a normal telephone line. The closest competition to ***Vianix, LLLC's*** proprietary speech processing technology in data rate and quality is the ITU (International Telecommunication Union) standard G.729, "Coding of speech at 8 kbps using conjugate-structure algebraic-code-excited linear-prediction". As the title states, it can code speech at 8kbps and gets a MOS score of 3.8. ***Vianix, LLC's*** proprietary vocoder, at a data rate of 6.8 kbps, achieves a MOS score of 4.2. That is a significant improvement in quality at a lower data rate. At an average data rate of 4.7 kbps the MOS score for ***Vianix, LLC's*** proprietary vocoder only drops to 4.0. That's over a 41% decrease in the data rate versus the competition while attaining a higher MOS score. This is the performance of ***Vianix, LLC's*** technology today, without the improvements currently being worked on in the R&D lab. Clearly, ***Vianix, LLC's*** proprietary vocoder achieves higher quality voice at significantly lower data rates and is therefore a superior solution.

Vianix, LLC has already developed and proven its speech processing technology. First this technology, as it exists, will be described. Descriptions of improvements that are currently being worked on will follow. These improvements are not guaranteed to be deliverable, but are possibilities.

The Technology (currently):

Vianix, LLC's proprietary speech compression technology implements a variable rate "vocoder" that maintains near toll-quality speech at very low data rates. The speech is encoded using a proprietary Codebook Excited Linear Predictive (CELP) speech coding algorithm. The vocoder can encode speech at fixed or variable data rates. In fixed rate mode it can code speech at rates of 4 kilobits per second (kbps), 4.8 kbps, 8 kbps, or 9.6 kbps. In variable rate mode, the vocoder automatically adjusts the data rate from 800 bps to 9.6 kbps every 20 milliseconds (ms), based on the amount of speech energy present. If the speech energy is high, the maximum rate will be used. If the speech energy is at a medium level, the intermediate rate will be used. If the speech energy is low, the 800 bps data rate will be used. This allows high speech quality to be maintained while significantly lowering the overall data rate. When in variable rate mode, the vocoder codes speech at under 3.5 kbps in typical two-way conversations without degrading the speech quality.

The proprietary CELP Encoder is the most complex function of the vocoder. The Encoder operates on one 20 ms frame at a time. It accepts digital speech from an industry standard 64kbps μ-law speech encoder (see codec description below). Each frame contains 160 Pulse Code Modulated (PCM) samples from the speech encoder. The encoding process includes measurement of the speech energy, data rate determination, dynamic adjustment of the rate thresholds, and encoding the speech into packets of processed data.

The proprietary CELP Decoder receives the packets of processed data needed for reconstructing the speech. The Decoder then provides a reconstructed speech output of 160 8-bit μ-law companded speech samples every 20 ms to the PCM interface. The PCM Interface of the vocoder connects to an industry standard μ-law PCM codec. This interface receives and transmits 64 kbps μ-law companded speech samples. These samples are transferred as 8 bit serial words every 125 μs. All data is transferred synchronously with externally sourced clocks and strobes.

Vianix, LLC's proprietary speech processing technology also contains a unique feature for tracking the background noise level to optimize the voice quality and compression rate when in variable rate mode. The background noise estimate gradually adjusts the adaptive rate thresholds to float above the level of the background noise. This enhances the efficiency of the vocoder and the quality of the recorded voice.

The vocoder also has an average rate limit capability. This is used when the desired maximum average encoded data rate must be limited for active speech. The vocoder has to be in variable rate mode before average rate limiting can be used. The average rate limiting capability forces some frames that would normally be encoded at the maximum rate to be encoded at a medium rate instead, thus lowering the average data rate. The percentage of frames that are forced to the lower rate is adjustable.

Vianix, LLC's proprietary vocoder uses a PCM Codec-Filter for digitizing and reconstructing the human voice. When recording, the vocoder receives digital data from the Codec and processes it further as described above. When playing back the vocoder processes the data packets into a digital form that the codec understands (PCM). The codec then converts the PCM data back into the human voice. Details of the Codec's operation are discussed below. Codecs are used primarily in the telephone network to facilitate voice switching and transmission. The name codec is an acronym from "CODer" for the analog to digital converter (ADC) used to digitize the voice, and "DECoder" for the digital to analog converter (DAC) used for reconstructing the voice. A codec is a single device that does both the ADC and DAC conversions.

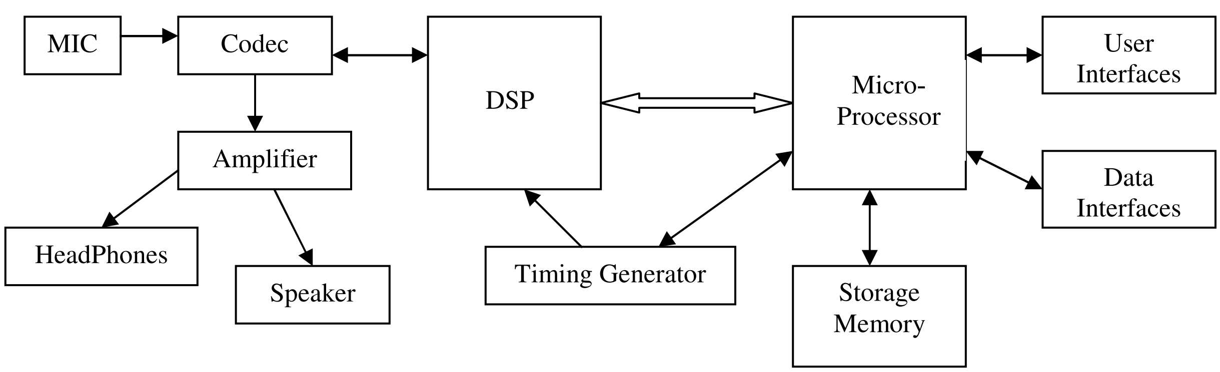
To digitize intelligible voice requires a signal to distortion ratio of about 30dB over a dynamic range of about 40 dB. This may be accomplished with a linear 13-bit ADC and DAC, but will far exceed the required signal to distortion ratio at larger amplitudes than 40 dB below the peak amplitude. This excess performance is at the expense of data per sample. The industry implemented two methods of data reduction that compress the 13-bit linear scheme to companded pseudo-logarithmic 8-bit schemes. These two companding schemes are: μ-law (pronounced Mu-law), primarily in North America and Japan; and A-law, primarily used in Europe. These companding schemes are the accepted standards worldwide. They follow a piecewise linear curve formatted as sign bit, three chord bits, and four step bits. For a given chord, all sixteen of the steps have the same voltage weighting. As the voltage of the analog input increases, the four step bits increment and carry to the three chord bits, which increment. When the chord bits increment, the step bits double their voltage weighting. This results in an effective resolution of six bits (sign + chord + four step bits) across a 42 dB dynamic range (seven chords above 0, by 6 dB per chord). This is a much more efficient method for digitally representing voice data, which is why it is used in the telephone networks.

The analog to digital conversion process samples the analog input signal at a specific rate. In a sampling environment, Nyquist theory says that to properly sample a continuous signal, it must be sampled at a frequency higher than twice the signals highest frequency component. Voice contains spectral energy above 3.4 kHz, but its absence is not detrimental to intelligibility. This has been proven through years of implementation in the telephone networks. To reduce the digital rate, which is proportional to the sampling rate, a sample rate of 8 kHz was adopted by the industry, consistent with a bandwidth of 3.4 kHz. This sampling requires a low-pass filter to limit the high frequency energy above 3.4 kHz from distorting the in-band signal. Because telephone lines are also subject to 50/60 Hz power line coupling, a high-pass filter before the ADC was also adopted as industry standard. After filtering, the analog signal is digitized and μ-law encoded. Then the μ-law data is transferred to the DSP.

The digital to analog conversion process reconstructs a staircase version of the desired in-band signal from the μ-law encoded data, which has spectral images of the in-band signal modulated about the sample frequency and its harmonics. These spectral images are called aliasing components, which need to be attenuated to obtain the desired signal. The industry standard is to use a low-pass filter to attenuate these aliasing components. This low-pass filter is generally called a re-construction or smoothing filter. This method results in very accurate reproduction of the human voice.

The current implementation of ***Vianix, LLC's*** proprietary voice compression technology is a dual processor system. There is a Digital Signal Processor (DSP) that runs the core speech processing algorithms and there is a microprocessor that manages buffering, storage, timing, data interfaces, and user interfaces (Refer to Figure 1). When recording, sound is turned into an analog electric signal by a microphone. This signal is then amplified and sent to the codec. The codec filters the signal then digitizes and μ-law encodes it, as described above. This μ-law encoded data is sent serially to the DSP. The DSP collects 20ms of the μ-law encoded data creating a frame. The DSP then analyzes the frame, determines a processing rate based on its setup parameters and the data, and then processes the frame using ***Vianix, LLC's*** CELP algorithm. The processed data is then transferred to the microprocessor while the DSP begins operating on the next frame. The microprocessor reformats the buffered processed data for more efficient use of the storage space and files it based on the operating mode the user has selected. The voice is now recorded. When playing back, the microprocessor reads a processed frame of data from storage and formats it for the DSP. When the DSP signals, the processor transfers the data to the DSP. The DSP then decodes the type of frame it is and processes it. This processed data is then serially transferred to the codec. The codec reconstructs the voice, as described above. The reconstructed analog signal (voice) is then amplified. If playback is through a speaker the amplifiers operate in a BTL (Bridge-Tied Load) configuration which increases the power transferred to the speaker. If playback is through a headset then the amplifiers operate in standard single ended mode. The amplified analog signal is then turned back into sound waves (voice) by the speaker or the headphones. Refer to the comparison section to see how ***Vianix, LLC's*** technology outperforms other technologies.

Figure 1: Current Implementation



Possible Technology Improvements:

Vianix, LLC is continuing its Research and Development (R&D) efforts to improve its technology. One of the main areas of research is in reducing the data rate even further while maintaining as much voice quality as possible. ***Vianix, LLC*** has made significant strides forward in this area already. By advancing the speech coding algorithms even further, average data rates of around 2.4 kbps with a MOS score of approximately 3.8 can be obtained, based on current results from the R&D lab. (Refer to the technology comparison section for an explanation of MOS scores) These algorithm advancements are in the early stages, but appear to achieve the data rate and quality goals. If the algorithms are proven to be effective and stable it is a simple matter of updating the software to take advantage of the new algorithms.

The R&D lab is also investigating incorporation of a Dual Tone Multiplexed Frequency (DTMF) decoder. DTMF tones are the tones you've heard while pressing the keys of a touch tone telephone. Each key produces a slightly different signal, which is a mixture of two frequencies. For example, pressing the '1' key produces a signal of 1209 Hz and 697 Hz mixed together. Decoding these signals is not difficult and should not pose a problem to integrate with the vocoding algorithms, enabling the ***Vianix*** technology to form the basis of a voice mail system.

To improve the overall versatility and recording quality, the R&D lab is researching an intelligent Automatic Gain Control (AGC) for the input signals. AGC is a method used to maintain the volume level of the signal on which it is operating. A problem arises when there is actual silence on the signal. A standard AGC circuit will increase its gain to the maximum level, which will result in the recording of static or noise. ***Vianix, LLC's*** intelligent AGC system will recognize the difference between quiet speech and silence, and will adjust the gain accordingly. It will also take into account the "Loud Talker" scenario and attenuate speech that is too loud. This feature will allow the modules to accurately record signals (Voice) that are at significantly different volumes. These volume differences can occur when speakers are at different distances from the microphone. They also occur because people speak at different volume levels. The AGC function will adjust for these variations, optimizing the recording volume for the compression algorithms. This will allow the module to be used in many different situations, with speakers at various distances and who speak at differing volume levels, while improving the intelligibility of the recorded voice.

The R&D lab is also researching an advanced filtering technique for the input and output filters of the codec. It is a digital filtering technology, which requires significantly less power to produce high-quality voice signals compared to earlier codec technology. These filters are being designed to meet ITU G.714 requirements. When these filters are implemented, they will lower power consumption, which in turn increases battery life and they will maintain if not improve the signal filtering characteristics.