SUB-CHANNEL BELOW THE PERCEPTUAL THRESHOLD IN AUDIO

Heping Ding, *heping.ding@nrc-cnrc.gc.ca*
Acoustics and Signal Processing, Institute for Microstructural Sciences
National Research Council, Ottawa, Ontario, Canada

ABSTRACT

This paper explores the concept and possible ways of making use of an audio channel's capacity below the perceptual threshold as a hidden sub-channel. Such an audio channel can be a one with conventional telephony or with voice-over-IP. Since the sub-channel is hidden, not detectable by the human ear, a communications system equipped with this technology will be compatible with the existing system in terms of audio signal transmission. Two potential applications of the technology are concurrent services and the extension of audio bandwidth. The latter application provides the listener with an improved audio quality. Recordings demonstrating this will be played at the presentation.

1. INTRODUCTION

The standard public switched telephone network (PSTN), which has been part of our daily life for more than a century, is designed to transmit toll-quality voice only. This results from the fact that the PSTN, whether implemented digitally or in analog circuitry, is only able to transmit analog signals in a "narrow frequency band" (NB), about 300 - 3400 Hz, as illustrated in Fig. 1.

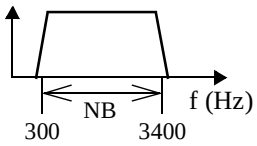

Fig. 1. Bandwidth of PSTN

Such a small bandwidth results in intelligibility and subjective quality that is inferior to that of other audio standards with a wider band. In addition, with the entire bandwidth occupied by voice, there is little room left for additional payload to be used for services and features that are readily available with many digital systems, such as digital private branch exchange (PBX) and voice over IP (VoIP), where a much larger equivalent bandwidth is available.

With NB being a nature of the PSTN infrastructure, we cannot directly count on the PSTN to provide a wider frequency band.

This document addresses the issue of virtually, as opposed to physically, extending the PSTN channel bandwidth for the purpose of implementing certain additional features. Section 2 reviews the existing work in the field. The goal of the research is discussed in Section 3. The proposed approach is discussed in Section 4 and possible applications in Section 5. A summary is given in Section 6.

2. EXISTING APPROACHES

Many efforts have been made to extend the capacity of the PSTN channel given the limited bandwidth. Existing approaches can be classified into certain categories below.

Time or frequency division multiplexing This category of techniques places voice and the additional payload in regions that are different in time or frequency, such as the cases with the calling line ID display and call waiting services, which are widely used in telephony. As a frequency division multiplexing example, [1] makes use of lower and upper frequency bands that are just beyond voice but still within the PSTN's capacity. Though relatively simple, these techniques inevitably cause voice interruption and/or distortion.

Voice coding (vocoding) These schemes are developed with the advancement of the studies on speech production and psychoacoustics, as well as of the digital signal processing theory and technology. Since all existing schemes, e.g., [2][3], transmit digital bits, a modem has to be used if they are to be implemented on PSTN or the like. Thus, the schemes are incompatible with the conventional end-user audio equipment.

Simultaneous voice and data (SVD) SVD technology is often used in dial-up modems that connect computers to the Internet through the PSTN. Typical examples are in [4][5]. SVD approaches change the audio signal and need SVD-capable modem hardware; therefore, they are not directly compatible with the conventional end-user audio equipment and are costly.

Audio watermarking These techniques embed information in an audio stream in ways so that it is inaudible to the human ear. An overview of the technology can be found in [6]. These techniques are aimed at high security, i.e., low probability of being detected or removed by an intentional attacker, and low payload rate. Our requirements for PSTN capacity extending are just the opposite; we want a high payload rate, while the security is considered less an issue.

3. OBJECTIVES

Our objectives for a scheme that extends the capacity of the PSTN are as follows.

Simplicity Firmware implementation should be simple and extra hardware should be none or minimized.

Compatibility with the existing end-user equipment A a conventional phone terminal, such as a plain ordinary telephone set (POTS), should still be able to access the basic voice service although it cannot access features associated with the additional payload. This is useful in audio broadcasting and conferencing operations and will facilitate the phase-in of the new technology.

High payload rate It should be higher than that offered by audio watermarking schemes while the stringent security requirement incurred by them can be eased. That is, the additional payload can possibly be tempered by an attacker who, however, is not able to obtain the information therein easily if an encryption scheme is used.

4. PROPOSED APPROACH

The presence of an audio component raises human ear's hearing threshold to another sound that is adjacent in time or frequency and to the noise in the audio component. Useful knowledge about the human auditory system and perceptual masking can be found in an overview article [7].

The concept behind our proposed approach is to make use of the masking principle and transmit audio components bearing certain additional payload below the perceptual threshold, which are not audible to the human ear but are detectable by a certain mechanism at the receiver, so that the payload can be retrieved.

While there can be various schemes of implementing the concept, only two examples are discussed in this paper.

4.1. Component Replacement

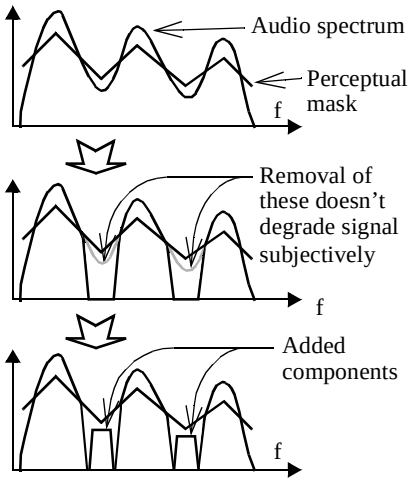

Fig. 2. Component replacement

This implementation scheme replaces certain audio components that are under the perceptual threshold with others that bear the additional payload, and is named as "component replacement." The scheme first breaks an audio stream in time-domain into frames, then processes them one by one.

As shown in Fig. 2, each frame is analyzed and a perceptual mask is estimated, a threshold below which signal components cannot be heard by the human ear. This can be done, for example, by an approach similar to that specified in [3]. Audio components below the perceptual mask are then replaced with components carrying the additional payload, which are still below the perceptual mask so that this operation will not result in audible distortion. Such a composite signal is sent through an audio channel, such as the PSTN, to the receiver. There may be channel degradations, such as parasitic or intentional filtering and additive noise, taking place along the way.

A POTS will treat the received signal as an ordinary audio signal and send it to its electro-acoustic transducer as usual, such as a handset receiver or a handsfree loudspeaker. Since the changes made by the replacement operations are under the perceptual threshold, they will not be audible to the listener.

At a receiver equipped with the component replacement scheme, the received composite signal is analyzed and the perceptual mask estimated. This mask is, to a certain accuracy tolerance, a replica of that obtained at the transmitter. As a result, the added components will also be below this perceptual mask. This makes them distinguishable from the original audio components, i.e., those that were not replaced. Thus, the added components can be identified and extracted from the received audio signal, and the additional payload therein can be restored.

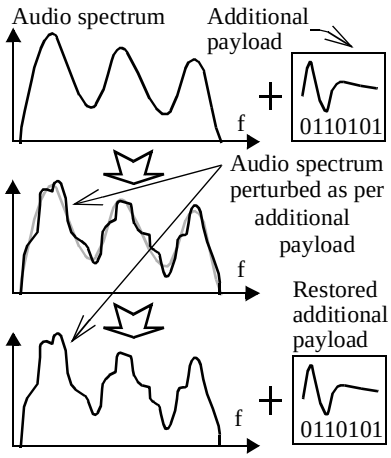

Fig. 3. Magnitude perturbation

4.2. Magnitude Perturbation

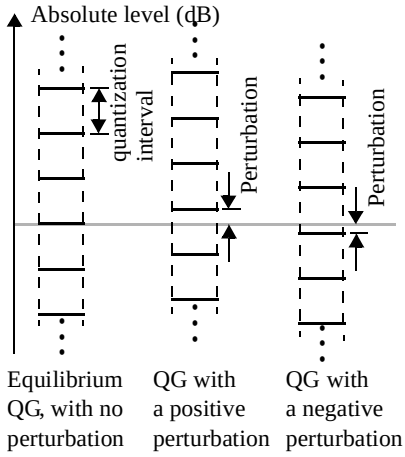

Fig. 4. Quantization grid (QG)

Instead of replacing audio components, this implementation scheme of the proposed approach adds certain noises that are below the perceptual mask to the original audio signal, which bear the additional payload. Since the noises are introduced as perturbations to the magnitudes of the audio components in the frequency domain, this scheme is named as "magnitude perturbation," shown in Fig. 3. The perturbations are in general uncorrelated with other noises such as the channel noise; therefore, the receiver is able to retrieve the perturbations in the presence of moderate channel noise.

The implementation is based on the concept of "quantization grid," shown in Fig. 4, which consists of a series of levels uniformly spaced in a logarithmic scale. The difference between two adjacent such levels is called the quantization interval (in dB). The ladder-like quantization grid can go up and down as a whole

depending on the perturbation introduced, but the relative differences between those levels remain the same, being the quantization interval.

The magnitude perturbation transmitter partitions the original audio stream into non-overlapped frames and processes them one after another. Each frame is first transformed into the frequency domain. The additional payload is encoded into the magnitude and the sign of the perturbation for each frequency bin. The magnitude of the perturbation must not exceed a certain limit, say, (quantization interval)÷3 dB, in order to avoid ambiguity to the receiver. Then, the quantization grid corresponding to each frequency bin is moved up or down as per the perturbation value determined. The magnitude of each signal component is perturbed by being quantized to the nearest level in its corresponding perturbed quantization grid. An inverse transform is performed on all the signal components so that a new time-domain frame with perturbations applied is formed.

Since, after the application of the perturbations, the magnitude of each signal component can only take a finite number of discrete values that are "quantization interval" dB apart, the magnitude perturbation scheme introduces noise to the audio signal. Obviously, the quantization interval must be large enough for the receiver to reliably detect the perturbations with the presence of channel noise, but small enough for the perturbation not to be audible.

The signal sequence consisting of non-overlapped consecutive such frames is transmitted to the receiver through an audio channel, such as that with a digital PBX, the PSTN, or VoIP, to the remote receiver. If PSTN is the media, there may be channel degradations, such as parasitic or intentional filtering and additive noise, taking place along the way.

A POTS will treat the received signal as an ordinary audio signal and send it to its electro-acoustic transducer. Since the changes made by the magnitude perturbation operations are under the perceptual threshold, they will not be audible to the listener.

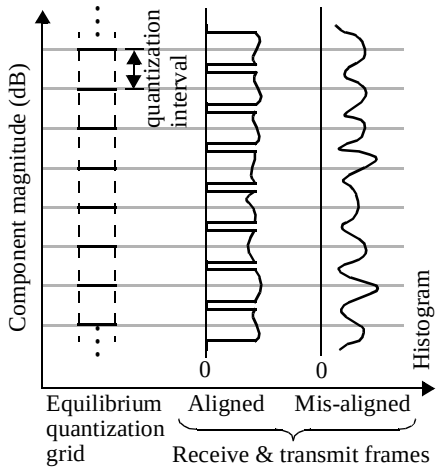

Fig. 5. Criterion for frame alignment

At a receiver equipped with the magnitude perturbation scheme, the received time sequence may need to first undergo certain equalization in order to reduce or eliminate the channel dispersion if the transmission channel contains analog elements, such as the PSTN. The received time sequence is then partitioned into frames. The frame boundaries are determined by using an adaptive phase locking mechanism, in an attempt to align the frames assumed by the receiver with those in the transmitter. The criterion to judge a correct alignment is that the magnitude distributions of components in all frequency bins are concentrated in discrete regions as opposed to being spread out. This is illustrated in Fig. 5.

5. APPLICATIONS

We discuss two classes of possible applications of the proposed approach.

5.1. Concurrent Services

With the proposed approach implemented in customers' terminals and in service providers' equipment, a hidden communications channel has been established between users in these two groups. They can then exchange information without interrupting or degrading the voice communications.

Some examples that the author can envision are as follows.

Instant calling line ID display The caller's identity is sent simultaneously with the very first ringing signal, so that the called party can immediately know who the caller is, instead of having to wait until after the end of the first ringing with the current technology.

Non-interruption call waiting A user while on the phone can get a message showing the identity of a third party that is calling, without having to hear a beep that interrupts the incoming voice.

Concurrent text message While on the phone talking to each other, two parties can simultaneously exchange text messages, such as e-mail and web addresses, spelling of strange names or words, phone numbers, ..., which come up as needed in the conversation. For this application, the phones need to be equipped with a keypad or keyboard as well as a display unit.

Simultaneous "display-based interactive services" and voice This feature allows the access to services like whether forecast, stock quotes, ticket booking, etc., and the results can be displayed on the phone's screen. With the proposed approach, these services can be integrated into voice so that the user can access both voice and such services at the same time.

In fact, the list for such concurrent services is endless, and it is up to service providers and system developers to explore the possibilities in this class of applications. The proposed approach just opens up a sub-channel for them to implement the features they can think of, and this sub-channel

- is compatible with the PSTN infrastructure,
- co-exists with audio,
- does not degrade audio quality, and
- is hidden to a POTS user

5.2. Audio Bandwidth Extension

It is well known that audio with a bandwidth beyond that of conventional NB (300 - 3400 Hz), i.e., PSTN, telephony can result in significant improvements in audio quality and intelligibility. Fig. 6. illustrates the concept of a wider bandwidth, XB as opposed to NB. In the figure, "lower band" (LB) stands for part of the XB

that is below NB, and "upper band" (UB) the XB part above NB. In addition, LB and UB will be denoted as LUB.

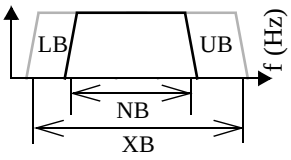

Fig. 6. Wider bandwidth: XB vs. NB

Since the PSTN's channel bandwidth cannot be extended beyond NB, we investigate the possibility of using our proposed approach to embed the LUB information into the NB signal at the transmitter and to restore it at the receiver. This way, the signal that is transmitted over the PSTN is NB physically, sounds the same as a conventional NB signal to a POTS user, and contains the information about LUB that can be decoded to restore the LUB at a receiver equipped with the approach.

In our application, the magnitude perturbation transmitter partitions the original audio stream, with a sampling rate of 16 KHz, into non-overlapped *N*-sample frames and processes them one after another. As shown in Fig. 7, a perceptual mask is calculated for frequency bins that are in the LUB range. After the audio signal has been band-split into NB and LUB, only those LUB components that are above the perceptual mask will be retained, and the magnitudes of NB frequency bins are perturbed as per the information about the locations and magnitudes of these LUB components. This information can be encoded into both the polarity and the magnitude of the perturbation for each frequency bin.

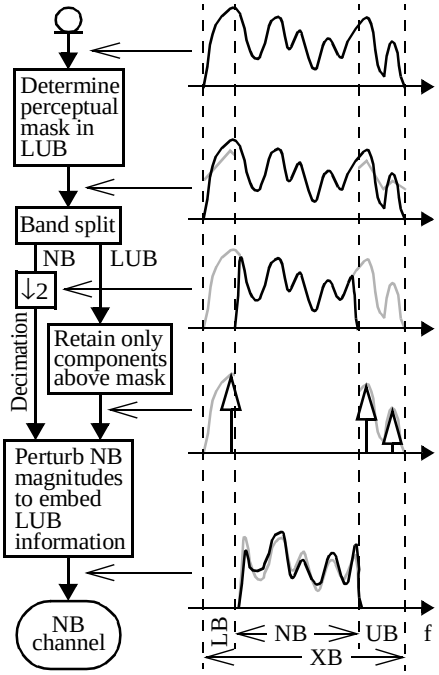

Fig. 7. Magnitude perturbation bandwidth extension transmitter

At a receiver equipped with the magnitude perturbation scheme, the received signal first undergoes some conditioning as discussed below Fig. 5. With that done, the perturbation that the transmitter applied to each NB frequency bin can be retrieved as the offset of the magnitude of the component from the nearest level in its corresponding equilibrium quantization grid, which is determined by inspecting the histogram of that bin, as shown in Fig. 5. Based on the retrieved perturbations, the information about the LUB components can be restored, as shown in Fig. 8.

Simulation demonstrating the audio bandwidth extension has been performed and it is shown that the additional payload can be successfully encoded into the perturbations and retrieved at the receiver.

6. SUMMARY

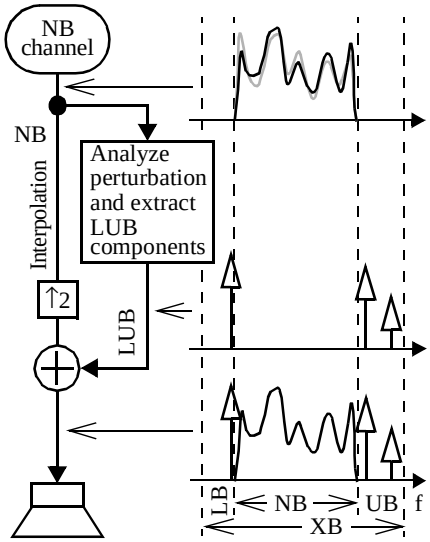

Fig. 8. Magnitude perturbation bandwidth extension receiver

We have discussed the need for extending the capacity of an audio channel associated with the traditional PSTN, digital PBX, or the VoIP, reviewed the existing techniques of doing that, and proposed an approach that creates a hidden sub-channel by either replacing audio components below a perceptual mask or introducing perturbations to certain parameters in the audio signal. Two classes of possible applications have been studied, one of which is an audio bandwidth extender with simulation showing that the concept of the proposed approach is viable.

REFERENCES

[1] Dwight W. Decker, *et_al*, "Speech and Data Multiplexor Optimized for Use over Impaired and Bandwidth Restricted Analog Channels," U. S. Patent 4,757,495, July 12, 1988.

[2] ITU-T Recommendation G.722.1, "Coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss," 1999.

[3] ISO/IEC JTC 1/SC 29/WG 11, ISO/IEC 13818-3 "Information technology - Generic coding of moving pictures and associated audio information - Part 3: Audio," (MPEG-2) April 15, 1998.

[4] Randy D. Nash, *et_al*, "Simultaneous Transmission of Speech and Data over an Analog Channel," U. S. Patent 4,512,013, April 16, 1985.

[5] Gordon Bremer, *et_al*, "Simultaneous Analog and Digital Communications with a Selection of Different Signal Point Constellations Based on Signal Energy," U. S. Patent 5,436,930, July 25, 1995.

[6] Ingemar J. Cox, *et_al*, *Digital Watermarking*, ISBN 1-55860-714-5, Academic Press, 2002.

[7] Ted Painter, Andreas Spanias, "Perceptual Coding of Digital Audio," *Proceedings of the IEEE*, Vol. 88, No. 4, pp. 451 - 513, April 2000.