DIGITAL CODING SOUND SIGNALS AND COMPRESSION MASK

Mirsamikov Mirislom Mirzoid o'g'li,
second year master TUIT named after Muhammad al-Khorazmiy
Tashkent, Uzbekistan
Second-year master of mirslom@gmail.com

Abstract: In certain cases, one sound may be hidden by another sound. For example, talking next to railroad tracks can be completely impossible if a train passes by. This kind of effect is called disguise. It is said that a faint sound is masked if it becomes indistinguishable in the presence of a louder sound.

Keyword: digital coding, masking, compression, compression of the sound, the sound, the compression of audio data, frequency masking, Temporary masking decoder formats sounds.

Аннотация: В некоторых случаях один звук может быть скрыт другим звуком. Например, говорить рядом с железнодорожными путями может быть совершенно невозможно, если поезд проходит мимо. Этот вид эффекта называется маскировкой. Говорят, что слабый звук маскируется, если он становится неразличимым при наличии более громкого звука.

Ключевое слово: цифровое кодирование, маскирование, сжатие, сжатие звука, звука, сжатие аудиоданных, частотное маскирование, декодер временного маскирования форматов звуков.

Simultaneous masking - Any two sounds while listening at the same time affect the perception of the relative volume between them. A louder sound reduces the perception of a weaker one, up to the disappearance of its audibility. The closer the frequency of the masked sound to the frequency of the masking, the more it will hide. The masking effect is not the same when the masked sound is shifted lower or higher in frequency relative to the masking one. Low-frequency sound masks high-frequency. It is important to note that high-frequency sounds cannot mask low- frequency sounds .

Temporary masking - This phenomenon is similar to frequency masking, but here there is a masking in time. When the masking sound is stopped, the masked one continues to be inaudible for some time. Under normal conditions, the effect of temporary masking lasts much less. The masking time depends on the frequency and amplitude of the signal and can reach 100 ms .

In the case when the masking tone appears later than the masked one, the effect is called post-masking. When a masking tone appears earlier than a masked tone (such a case is also possible), the effect is called pre-masking.

Post-stimulus fatigue - Often, after exposure to high-intensity loud sounds, a person's auditory sensitivity decreases sharply. Recovery to normal thresholds can last up to 16 hours. This process is called a “temporary shift in the threshold of auditory sensitivity” or “ post-stimulus fatigue." The threshold shift begins to appear when the sound pressure level is above 75 dB and accordingly increases with increasing signal level. Moreover, the highest influence on the shift of the sensitivity threshold is exerted by the high-frequency components of the signal.

Lossy audio data compression is based on the imperfection of human hearing in the perception of sound information. The inability of a person in certain cases to distinguish between quiet sounds in the presence of louder ones, called the masking effect, was used in algorithms for reducing psychoacoustic redundancy. The effects of auditory masking depend on the spectral and temporal characteristics of the masked and masking signals and can be divided into two main groups:

Frequency (simultaneous) masking

Temporary (non-simultaneous) masking

The masking effect in the frequency domain is due to the fact that in the presence of large sound amplitudes the human ear is insensitive to small amplitudes of close frequencies. That is, when two signals are simultaneously in a limited frequency domain, the weaker signal becomes inaudible against the background of the stronger one.

Masking in the time domain characterizes the dynamic properties of hearing, showing a change in time of the relative threshold of audibility (the threshold of audibility of one signal in the presence of another), when the masked and masked signals do not sound simultaneously. In this case, it is necessary to distinguish between the phenomena of post-masking (changing the threshold of audibility after a high level signal) and pre-masking (changing the threshold of audibility before the arrival of the maximum level signal). A weaker signal becomes inaudible 5 to 20 ms before the masking signal is turned on and becomes audible 5 0 to 200 ms after it is turned on.

The best method of coding sound, taking into account the effect of masking, is strip coding. Its essence is as follows. A group of samples of the input audio signal, called a frame, is fed to a filter unit that divides the signal into frequency subbands. The output of each filter is the part of the input signal that falls into the passband of this filter. Further, in each band, using the psychoacoustic model, the spectral composition of the signal is analyzed and it is estimated which part of the signal should be transmitted without abbreviations, and which lies below the masking threshold and can be quantized to a smaller number of bits. To reduce the maximum dynamic range, the maximum sample in the frame is determined and the scaling factor is calculated, which brings this sample to the upper quantization level. This operation is similar to companding in analog broadcasting. All other samples are multiplied by the same factor.

A scaling factor is transmitted to the decoder along with the encoded data to correct the gain of the latter. After scaling, the masking threshold is estimated and the total number of bits is redistributed between all the bands.

Obviously, after eliminating the psychoacoustic redundancy of sound signals, their exact restoration during decoding is no longer possible. Methods of eliminating psychophysical redundancy can provide compression of digital audio data 10 to 12 times without significant loss in quality.

Many other tricks can serve as a way to reduce the amount of audio information data. Even a simple narrowing of the signal frequency band together with a decrease in the dynamic range can already be called compression of audio data. For example, the cellular audio compression standard uses both. In an effort to remove redundancy from sound, a codec with poor signal quality becomes selective for certain words, stubbornly swallowing them.

Generally accepted data compression methods, such as RLE , statistical and dictionary methods, can be used to compress sound files without loss, but the result depends heavily on specific audio data. Some sounds will compress well with RLE , but poorly with statistical algorithms. Other sounds are more suited to statistical compression, but with a dictionary approach, on the contrary, expansion can occur. Here is a brief description of the effectiveness of these three methods in compressing audio files.

RLE works well with sounds that contain a long series of repeating sound bites - samples . With 8-bit sampling, this can happen quite often. Recall that the difference in electrical voltage between two 8-bit samples is about 4 mV . A few seconds of homogeneous music, in which the sound wave will change by less than 4 mV , will generate a sequence of thousands of identical samples . With 16-bit sampling , obviously, long repeats are less common, and therefore, the RLE algorithm will be less efficient.

Statistical methods assign variable-length codes to sound samples according to their frequency. With 8-bit sampling , there are only 256 different samples , so a large sound file sampled sound can be distributed evenly. Such a file cannot be compressed well by the Huffman method. With 16-bit sampling, more than 65,000 audio fragments are allowed. In this case, it is possible that some samples will occur more often and others less often. With strong asymmetry of probabilities, good results can be achieved using arithmetic coding.

Vocabulary based methods suggest that certain phrases will occur frequently throughout the entire file. This occurs in a text file in which individual words or their sequences are repeated multiple times. The sound, however, is an analog signal and the values of the specific generated samples are heavily dependent on the operation of the ADC. For example, with 8-bit sampling , an 8 mV wave becomes a numerical sample equal to 2, but a wave close to it, say, at 7.6 mV or 8.5 mV may become a different number. For this reason, speech fragments containing matching phrases and sounding the same for us may slightly differ when they are digitized. Then they will fall into the dictionary in the form of different phrases, which will not give the expected compression. Thus, vocabulary methods are not very suitable for compressing sound.

You can achieve better results when compressing sound with the loss of part of the audio information by developing compression methods that take into account the characteristics of sound perception. They delete that part of the data that remains inaudible to the hearing organs. This is similar to compressing images with discarding information that is invisible to the eye. In both cases, we proceed from the fact that the initial information (image or sound) is analog, that is, part of the information is already lost during quantization and digitization. If we allow some more loss by doing it carefully, this will not affect the playback quality of the expanded sound, which will not differ much from the original. We will briefly describe two approaches called pause suppression and compaction.

The idea of suppressing pauses is to consider small samples as if they were not (that is, they are zero). Such a zeroing will produce a series of zeros, so the method of suppressing pauses is, in fact, a variant of RLE , adapted to compress sound. This method is based on the peculiarities of sound perception, which consists in the tolerance of a person's ear to discard hardly audible sounds. Audio files containing long sections of quiet sound will be better compressed by the pause suppression method than files filled with loud sounds. This method requires the participation of a user who will control the parameters that set the volume threshold for samples . In this case, two more parameters are required, they are not necessarily controlled by the user.

One parameter is used to determine the shortest sequences of quiet samples , usually 2 or 3. And the second sets the smallest number of consecutive loud samples , when silence or pause stops. For example, after 15 quiet samples , 2 loud and then 13 quiet ones can follow, which will be defined as one big pause of length 30, and a similar sequence of 15, 3 and 12 samples will become two pauses with a short sound between them.

Compaction is based on the property that the ear is better able to distinguish between changes in the amplitude of quiet sounds than loud ones. A typical ADC of computer sound cards uses linear conversion when translating voltage into a numerical form. If the amplitude has been converted to a number, then the amplitude will be converted to a number. The compression-based compression method first analyzes each sample of the sound file and applies a non-linear function to it to reduce the number of bits assigned to this sample . For example, with 16-bit samples , a compressed encoder can apply the following simple formula

Audio data compression (audio compression) is a type of data compression, encoding used to reduce the volume of audio files or to reduce the bandwidth for streaming audio. Sound file compression algorithms are implemented in computer programs called audio codecs. The invention of special compression algorithms for audio data is motivated by the fact that general compression algorithms are inefficient for working with sound and make it impossible to work in real time.

As in the general case, lossless sound compression is distinguished, which makes it possible to restore the original data without distortion, and lossy compression, in which such recovery is impossible. Lossy compression algorithms allow a greater degree of compression, such as audio CD can accommodate no more than one hour "uncompressed" music, compressive lossless CD accommodate nearly 2 hours of music, while lossy compression if the average bit rate - 7-10 hours.

Lossless Compression - The complexity of lossless audio compression is that sound recordings are extremely complex in structure. One of the compression methods is to search for samples and their repetitions, but this method is not effective for more chaotic data, such as digitized sound or photographs. It is interesting that if computer-generated graphics are much easier to compress without loss, then synthesized sound has no advantages in this regard. This is because even computer-generated sound usually has a very complex shape, which is a difficult task for inventing an algorithm.

D berating the difficulty lies in the fact that the sound usually changes very quickly and this is also the reason that the ordered sequence of bytes appear very rarely.

The most common lossless compression formats are:

Free Lossless Audio Codec (FLAC), Apple Lossless, MPEG-4 ALS, Monkey's Audio, and TTA.

Lossy compression - Lossy compression is extremely widespread. In addition to computer programs, lossy compression is used in streaming audio to DVD , digital television and radio, and streaming media on the Internet.

An innovation of this compression method was the use of psychoacoustics to detect sound components that are not perceived by the human ear. An example is either high frequencies, which are perceived only when their power is sufficient, or quiet sounds that occur simultaneously or immediately after loud sounds and therefore are masked by them - such sound components can be transmitted less accurately, or not at all.

To mask, the signal from the time sequence of samples of the amplitude is converted into a sequence of spectra of sounds in which each component of the spectrum is encoded separately. To implement such a conversion, methods of fast Fourier transform, MDCT, quadrature-mirror filters or others are used. Compression in a certain frequency domain may consist in that masked or zero components are not stored at all, or encoded with lower resolution. The result of this operation will be encoding with an average bit depth of 8-bit, but the result will be much better than when encoding the entire frequency range with 8-bit bit depth. However, it is obvious that low-resolution transcoded fragments of the spectrum can no longer be restored exactly, and thus are lost forever.

List of used literature

1. Kovalgin Yu.A., Vologdin E.I. Digital coding of audio signals. - St. Petersburg: CROWN-print, 2004
2. V.P. Deacons. Wavelets. From theory to practice. - M .: SOLON-R, 2002.-448 p.
3. Zainidinov Kh.N., Zainutdinova MB, Nazirova E.Sh., Yusupov I. Two-dimensional wavelet bases with compact carriers in the problem of signal samples - functions of two variables. PROCEEDINGS of the Scientific-Practical and Spiritual-Educational Conference Dedicated to the 1235 th Anniversary of Muhammad ai-Khwarizmi, International conference on importance of information-communication technologies in innovative development of sectors of economy, Tashkent, Uzbekistan, April 5-6, 2018 p.834-836
4. Digital processing of two-dimensional signals in the basis of haar wavelets - Computer Systems and Technologies. 19th International Conference, CompSys Tech`18, (Indexed by SCOPUS), Ruse, Bulgaria, September 13-14, 2018, Proceedings, p. 130-133 - H. Zaynidinov, M. Zaynutdinova , E. Nazirova