Comparison of Techniques for VoIP Packet Loss Concealment

CSC 466 - A02

Jesse Browell

December 12, 2022

Summary	**3**
Introduction	**3**
Background	**4**
Methodology	**6**
Results	**7**
Discussion	**9**
Conclusion	**10**
References	**12**
Appendix	**12**

Summary

In this project, we evaluated the effectiveness of three different packet loss concealment (PLC) techniques on audio quality and comprehensibility: silence insertion, sound repetition, and the Opus codec. Our findings suggest that the performance of different packet loss techniques may vary depending on the level of packet loss present. At low levels of packet loss, the Opus codec performed the best, while at higher levels of packet loss, sound repetition was the most effective technique. These results provide some initial insights into the performance of different packet loss techniques, but further research is needed to fully understand their strengths and limitations.

Introduction

Packet loss is a common problem in voice over IP (VoIP) systems, where data packets carrying audio information are transmitted over a network. Packet loss can occur due to various factors such as network congestion, interference, or errors in transmission. When packet loss occurs, the quality of the audio can be significantly degraded, making it difficult for users to understand the conversation.

Packet loss concealment (PLC) techniques are used to mitigate this problem by replacing lost or corrupted packets with synthesized data in order to maintain the continuity and quality of the audio. In this project, we will demonstrate three different PLC techniques: silence insertion, sound repetition, and the Opus audio codec. We will evaluate their effectiveness by simulating different levels of packet loss on two audio files and comparing the results.

The goal of this project is to provide a practical demonstration of PLC techniques and to evaluate their effectiveness in real-world scenarios. Our results will provide insights into the strengths and weaknesses of each technique and may inform future research and development in this area.

Background

Packet loss is the failure of some number of packets to reach their destination. This can happen for a variety of reasons, including network congestion, faulty hardware, or even interference from other electronic devices. VoIP systems are highly sensitive to packet loss. This is because voice data is typically transmitted in real-time, and any delay or interruption can result in poor-quality audio. Packet loss can cause a variety of problems in a VoIP system, including choppy or garbled audio, long delays, or even complete failure of the transmission. This can be frustrating for users and can lead to a degraded experience.

PLC is a technique used in VoIP systems to mask the effects of packet loss. When packets are lost in transmission, the audio data they contain is not received, the receiving system is unable to play the audio needed for the call. PLC uses various algorithms to fill in the gaps caused by lost packets, so that the audio data is uninterrupted. This can help to improve the overall quality of the audio, making the conversation more natural and easier to understand. PLC is used in VoIP systems to improve the user experience and to help ensure that conversations are clear and free from interruptions. This project looks at three different techniques for PLC , silence replacement, sound repetition, and the features built into the Opus audio codec.

One technique for PLC in VoIP systems is silence replacement. This technique involves replacing the audio data that was lost in the transmission of packets with silence [1 p. 170]. This can help to mask the effects of packet loss and can make the conversation more natural and easier to understand. Silence replacement works by inserting silence into the audio stream in place of the lost data. This helps to fill in the gaps caused by packet loss and can prevent choppy or garbled audio. While silence replacement is not perfect, it can be an effective way to improve the overall quality of the audio in a VoIP system.

Another technique for PLC in VoIP (Voice over Internet Protocol) systems is sound repetition [1 p. 171-172]. This technique involves repeating the last known audio data in place of the lost data. This can help to mask the effects of packet loss and can make the conversation more natural and easier to understand. Sound repetition works by repeating the last period of audio data to fill in the gaps caused by lost packets. This helps to prevent choppy or garbled audio and can improve the overall quality of the conversation.

The Opus audio codec is a widely used codec for VoIP systems. It is known for its high-quality audio and its ability to adapt to a variety of network conditions, including packet loss. One of the key features of the Opus codec is its use of PLC to mask the effects of lost packets. As described in [2], the PLC technique depends on the mode of the last packet received. In CELT mode, the PLC finds a periodicity in the decoded signal and repeats the windowed waveform using the pitch offset. In SILK mode, the PLC uses linear predictive extrapolation from the previous frame. These techniques can help to improve the overall quality of the audio, making the conversation more natural and easier to understand.

Methodology

In this study, I aimed to evaluate the effectiveness of different packet loss techniques on audio quality. Packet loss is a common issue in audio and video transmission, and can result in degraded quality and reduced intelligibility. I selected two audio files and simulated packet loss at three different levels: 10%, 20%, and 40%. I used three techniques to simulate packet loss: silence insertion, sound repetition, and the Opus codec.

To implement silence insertion and sound repetition, I wrote custom Python code that randomly replaced packets with either silence or the previous packet, depending on the technique being tested. For the Opus codec packet loss simulation, I used the packet loss simulation demo built into the open-source Opus codec project. This allowed me to easily simulate packet loss using the Opus codec without having to implement the technique myself.

To evaluate the results, I used a speech-to-text program to generate transcriptions of the audio files with simulated packet loss. These transcriptions were then compared against the original audio files to assess the accuracy of the transcriptions and the overall quality of the audio. I also created my own subjective interpretation of each result. These results were then compared to determine which packet loss technique was the most effective at preserving audio quality and intelligibility at each different packet loss level.

Results

In this project, we evaluated the effectiveness of three different packet loss techniques on audio quality and comprehensibility: silence insertion, sound repetition, and the Opus codec. We simulated packet loss at three levels (10%, 20%, and 40%) and used both objective and subjective methods to evaluate the results.

The objective evaluation was based on speech-to-text transcription accuracy. For the first audio file, we found that all three techniques performed similarly at 10% packet loss, with a transcription accuracy of 100%. However, at 20% and 40% packet loss, the Opus codec performed worse than the other techniques. In particular, at 40% packet loss, the Opus codec had a transcription accuracy of only 54.55%, while sound repetition had an accuracy of 100%. For the second audio file, we found that silence and sound repetition had a transcription accuracy of 100% at all levels of packet loss, while the Opus codec had an accuracy of 92.31% at all levels. These results are summarized in tables one and two.

In addition to the objective evaluation, we also used subjective interpretation to evaluate the audio quality. We found that the Opus codec provided the best audio quality at 10% packet loss, while sound repetition was the most effective technique at 20% and 40% packet loss. The effects of packet loss were not evenly distributed with the Opus codec, with periods of normal sound and severely artifacted sound. At the lowest level of packet loss (10%), the effects of packet loss were much less noticeable with the Opus codec than with the silence and repetition techniques. However, the performance of the Opus codec degraded much more severely from low to high levels of packet loss in comparison.

Silence insertion was found to be the worst technique in all categories. Even at 10% packet loss it was found to create an unpleasant static sound in the audio. The effects of packet loss were consistent throughout the audio file with silence insertion, without any parts being impacted significantly more than others. However, at 20% and 40% packet loss, the impact of silence insertion was noticeable.

Sound repetition was found to create a static-like effect on the audio, although to a lesser extent than silence insertion. The effects of packet loss were consistent throughout the audio file with sound repetition, without any parts being impacted significantly more than others. Even at the lowest level of packet loss (10%), the impact of sound repetition was noticeable.

Discussion

Our findings suggest that the performance of the different packet loss techniques may vary depending on the level of packet loss present. At low levels of packet loss, it may provide the best audio quality, while at higher levels of packet loss, another technique may be more effective.

The results of the subjective and objective analyses in this project differed in some ways. In the subjective analysis, we found that the Opus codec provided the best audio quality at 10% packet loss, while sound repetition was the most effective technique at 20% and 40% packet loss. In contrast, the objective analysis showed that at 10% packet loss, all three techniques had similar transcription accuracy, while at 20% and 40% packet loss, the Opus codec performed worse than the other techniques in terms of transcription accuracy. These differences may be due to the fact that the subjective analysis was based on our own interpretations of the audio quality, while the objective analysis was based on a speech-to-text transcription. However, both types of analysis are valuable in understanding the effectiveness of different packet loss techniques, and further research may be necessary to fully understand the strengths and limitations of each approach.

However, there is still much work that could be done in this area. For example, there are many more packet loss techniques that could be tested. In particular, the use of neural network-based techniques warrants investigation, as they have demonstrated superior performance compared to other methods in some scenarios. This type of technique is being used in newer audio codecs such as Google's Lyra and Microsoft's Satin.

Additionally, a better simulation of packet loss could be performed. In the real world each packet does not have an independent chance to be dropped, and packets are often lost in sequence creating what is called bursty packet loss with periods of no packet loss and periods with high packet loss. This project does not take this into account. A more realistic simulation of packet loss would allow for a more accurate evaluation of the effectiveness of different packet loss techniques.

Overall, our study provides some initial insights into the performance of different packet loss techniques on audio quality and intelligibility. However, further research is needed to fully understand the best technique to use in a more real world environment.

Conclusion

In conclusion, this project aimed to evaluate the effectiveness of different PLC techniques on audio quality and intelligibility. We simulated packet loss at three levels (10%, 20%, and 40%) and used both objective and subjective methods to evaluate the results. The objective evaluation was based on speech-to-text transcription accuracy, while the subjective evaluation was based on our own interpretation of the audio quality.

Our findings suggest that the performance of different packet loss techniques may vary depending on the level of packet loss present. At low levels of packet loss, the Opus codec performed the best in terms of both transcription accuracy and audio quality, while at higher levels of packet loss, sound repetition was the most effective technique. This indicates that the best PLC technique may depend on the specific conditions of the transmission, such as the level of packet loss and the type of audio data being transmitted.

Additionally, our results showed some differences between the subjective and objective evaluations of the audio quality. While the Opus codec provided the best audio quality at low levels of packet loss according to the subjective analysis, the objective analysis showed that all three techniques had similar transcription accuracy at this packet loss level. This suggests that the subjective and objective methods may provide complementary insights into the performance of different packet loss techniques.

Overall, this project provides some initial insights into the performance of different packet loss techniques on audio quality and intelligibility. However, there is still much work that could be done in this area, including the testing of additional packet loss techniques and the development of more realistic packet loss simulations. Further research in this area may be valuable in improving the performance of VoIP systems and enhancing the user experience.

References

[1] T. Bäckström, “Packet loss and concealment,” Signals and Communication Technology, pp. 161–184, 2017

[2] J. M. Valin, K. Vos, and T. Terriberry, “Definition of the opus audio codec,” RFC Editor, 01-Sep-1970. [Online]. Available: https://www.rfc-editor.org/rfc/rfc6716. [Accessed: 05-Dec-2022].

Appendix

Appendix A: Tables

Technique	**10% Packet Loss**	**20% Packet Loss**	**40% Packet Loss**
Opus	100	94.55	54.55
Silence	100	100	92.59
Repeat	100	100	100

Table 1: Results of speech-to-text comparison on audio file one

Technique	10% Packet Loss	20% Packet Loss	40% Packet Loss
Opus	92.31	92.31	92.31
Silence	100	100	100
Repeat	100	100	100

Table 2: Results of speech-to-text comparison on audio file two

Appendix B: Member Statement

Jesse Browell is currently an undergraduate student at the University of Victoria working toward a Bachelor's degree of Software Engineering.