

# Speech Information Systems

## Lab 4

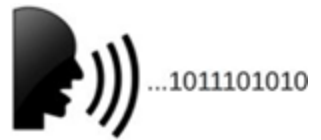
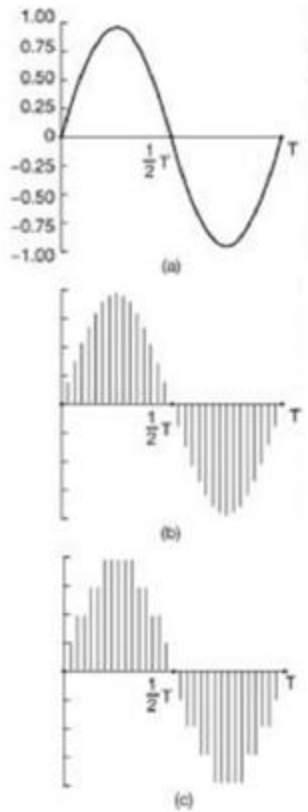
### Sampling, Quantization, Speech Coding

**Author:** Bartalis Mátyás, Ali Raheem Mandeel, and

Ibrahim Ibrahimov

**Email:** [ibrahim@tmit.bme.hu](mailto:ibrahim@tmit.bme.hu)

**Academic Year:** 2024/2025 – Spring

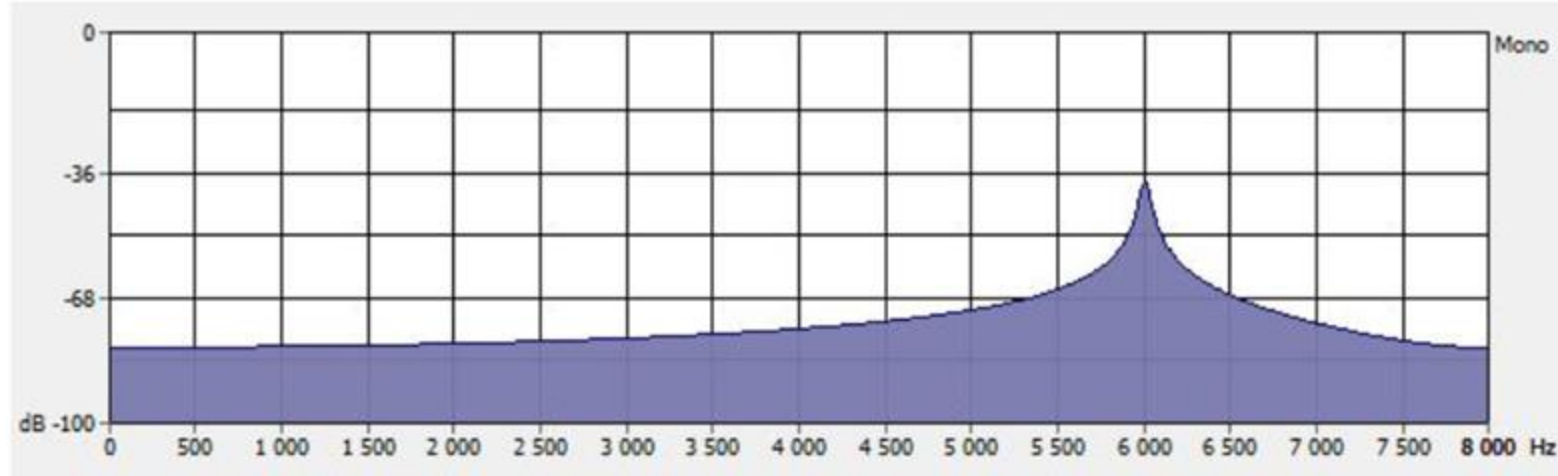


# **SAMPLING**

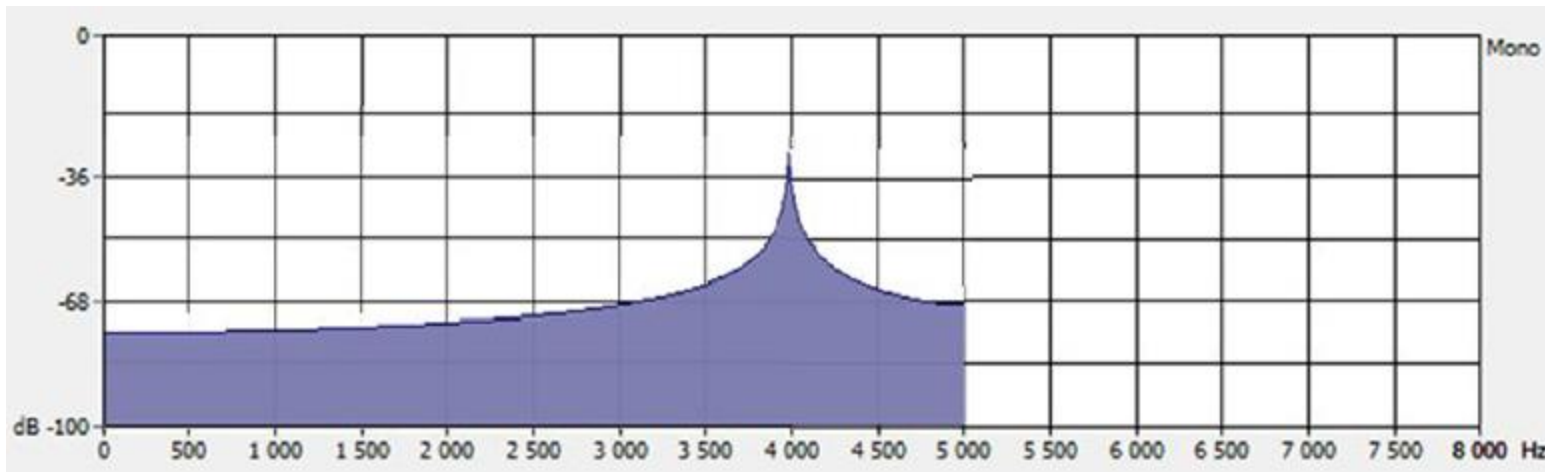
A **6 kHz sine wave** is sampled at **10 kHz** without filtering. Where does the original sine wave appear in the sampled signal?

- 4 kHz
- 16 kHz
- 5 kHz
- 3 kHz
- 2 kHz
- 1 kHz

### a) Original Signal



### b) Sampled Signal Without an Anti-Aliasing Filter

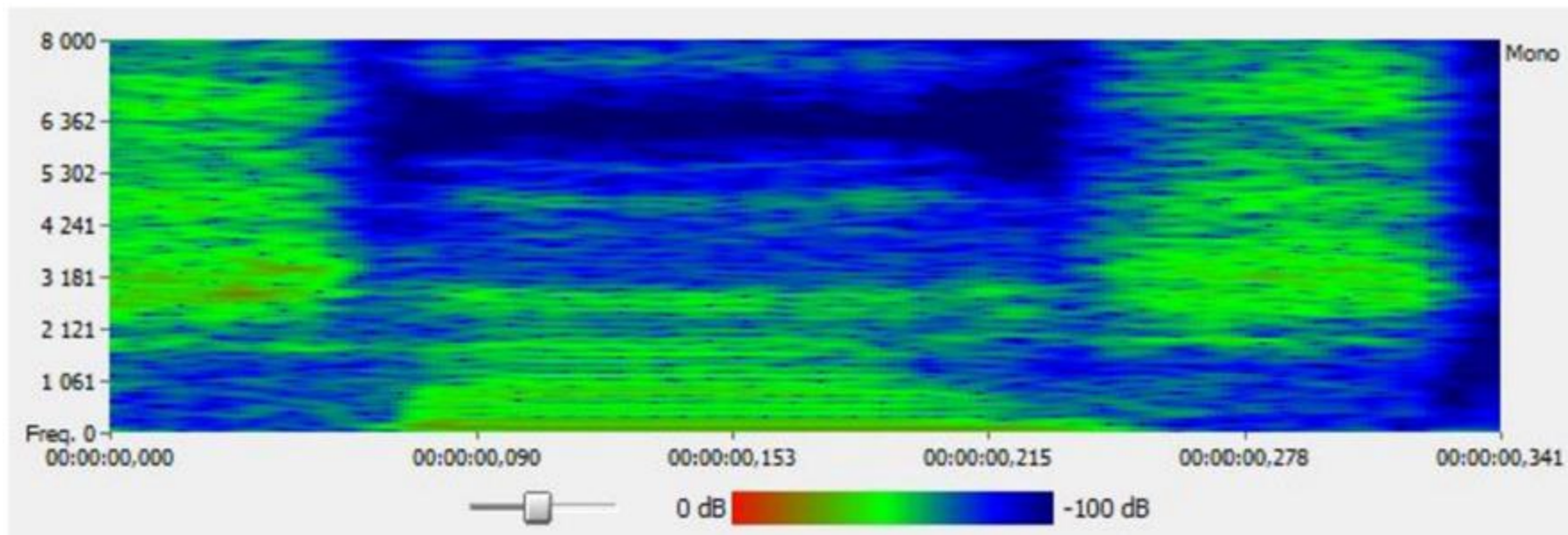


A **sampling-reconstruction system** operates with a **5.5 kHz sampling frequency** without an **anti-aliasing filter** at the input, but with a **high-quality low-pass filter with a cutoff frequency of 2.5 kHz** at the output.

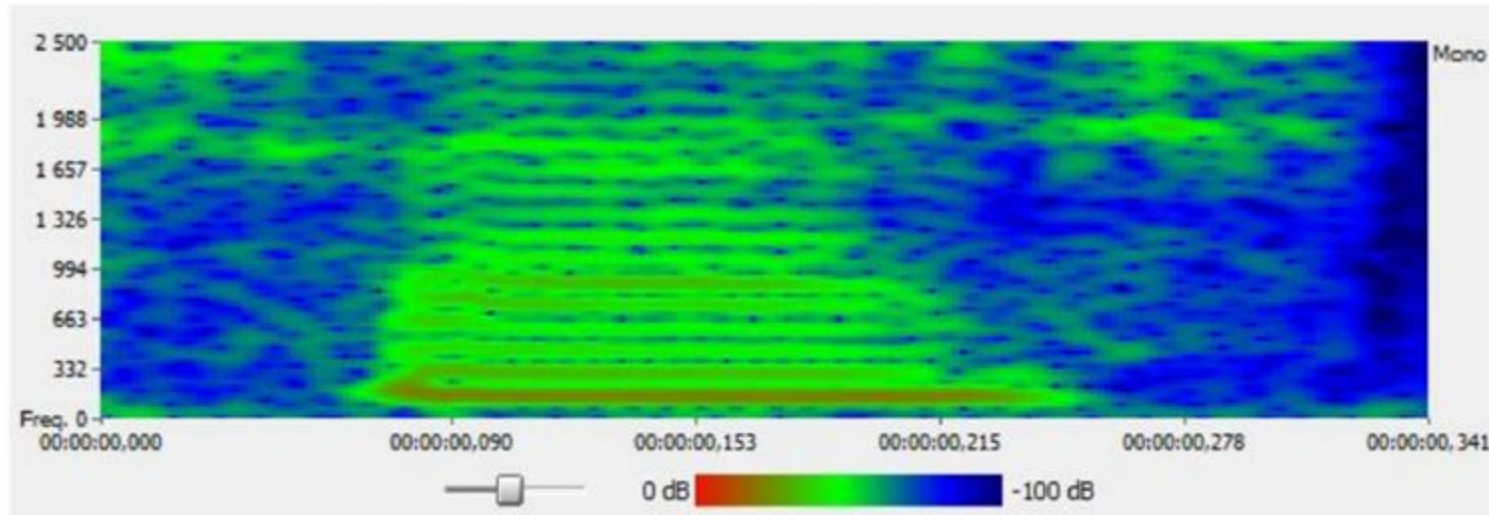
### **Tasks:**

- a) **Analyze the effect** of inputting the Hungarian word "*Sás*" (pronounced by an adult female speaker). Describe the characteristics of the output signal. Justify your answer with **numerical values**.
- b) **Determine the optimal system parameters** (sampling frequency, input, and output filters) for **distortion-free transmission**.

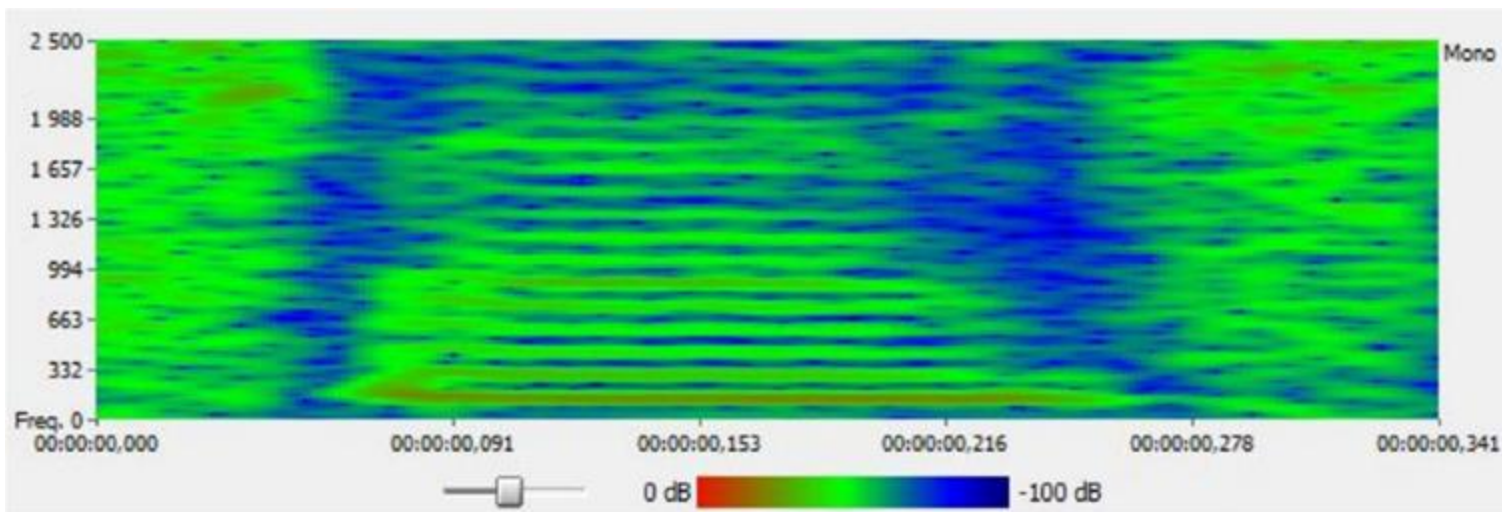
- “*Sás*”



### a) Original Signal



### b) Sampled Signal Without an Anti-Aliasing Filter



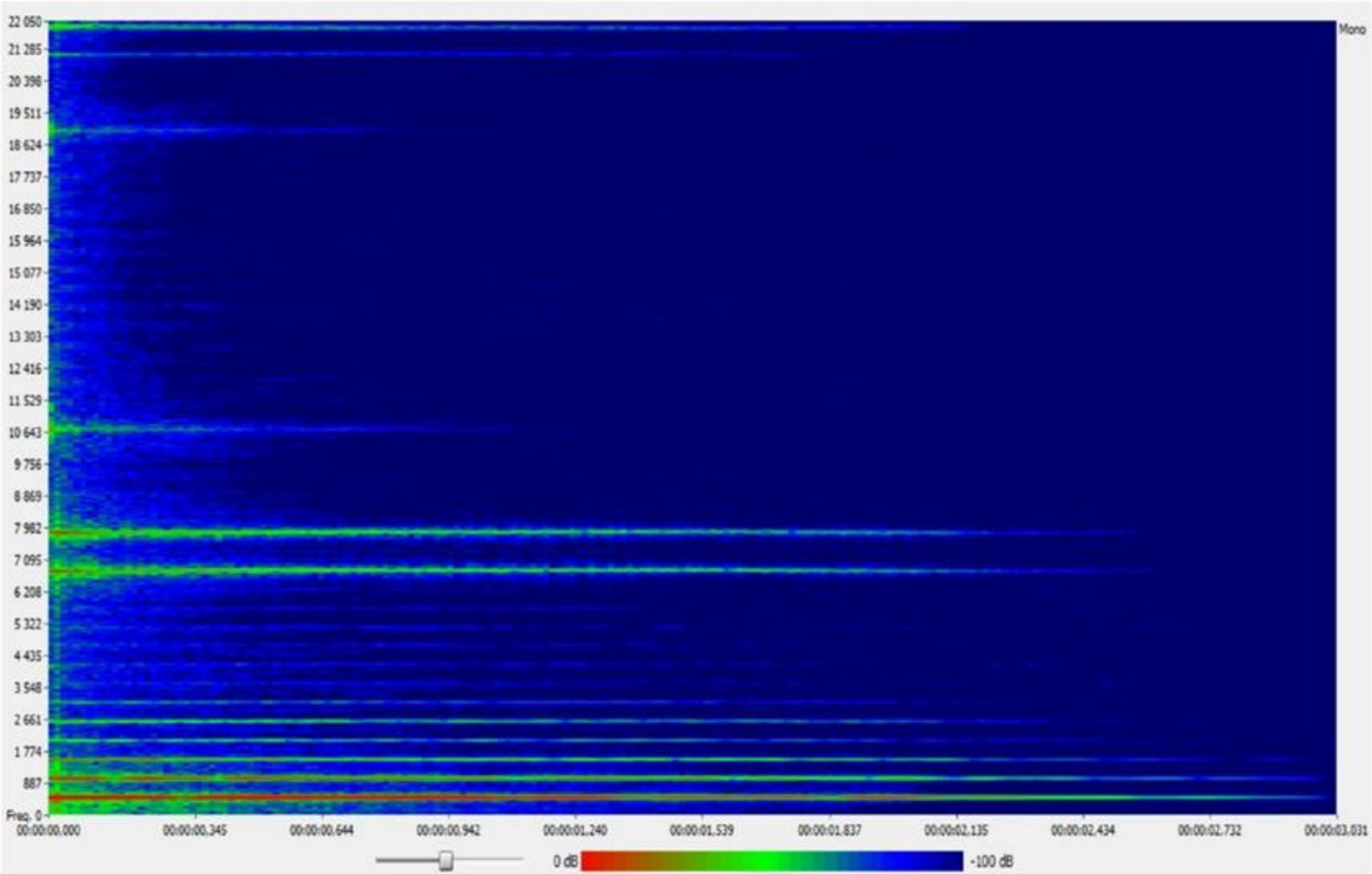
A **speech sampling system** operates with an **unknown sampling frequency** and **PAM-type reconstruction filter with  $H(f)$  characteristics**. The system alternately receives **speech signals** and **piano solos**. The **speech sounds intelligible**, but the **piano sound is distorted**.

### **Tasks:**

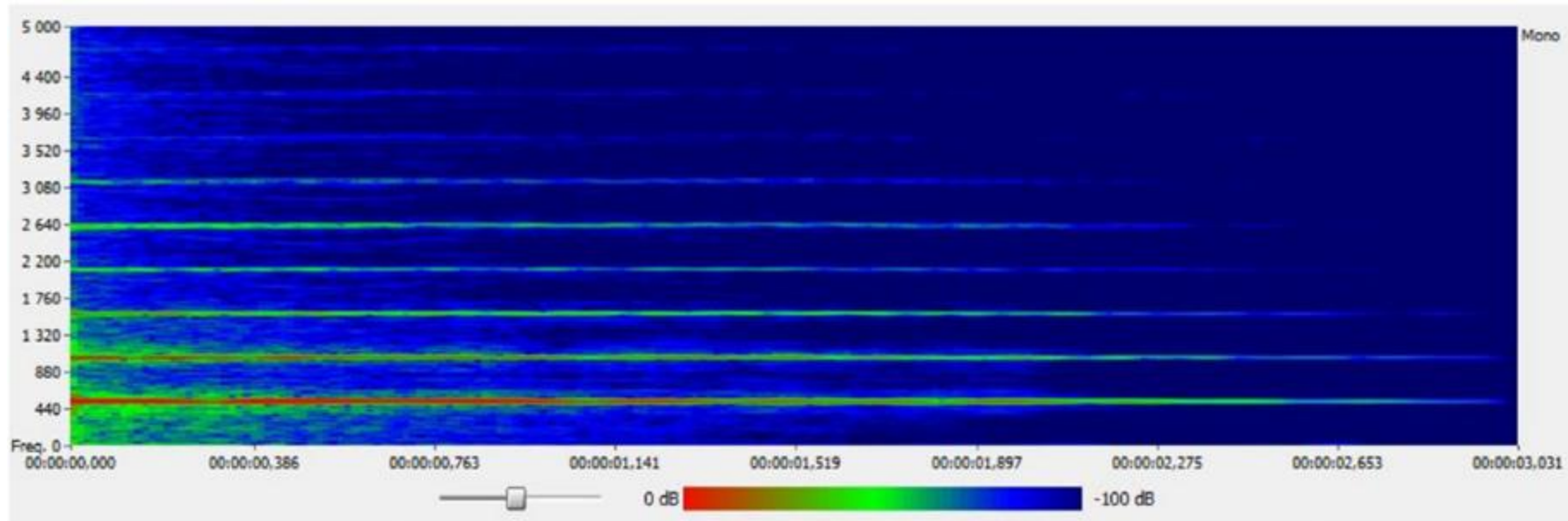
- a) **Identify system parameters** that could cause this behavior. Multiple correct answers are possible.
- b) **Propose a solution** that reconstructs the original signal with **reasonable complexity**.



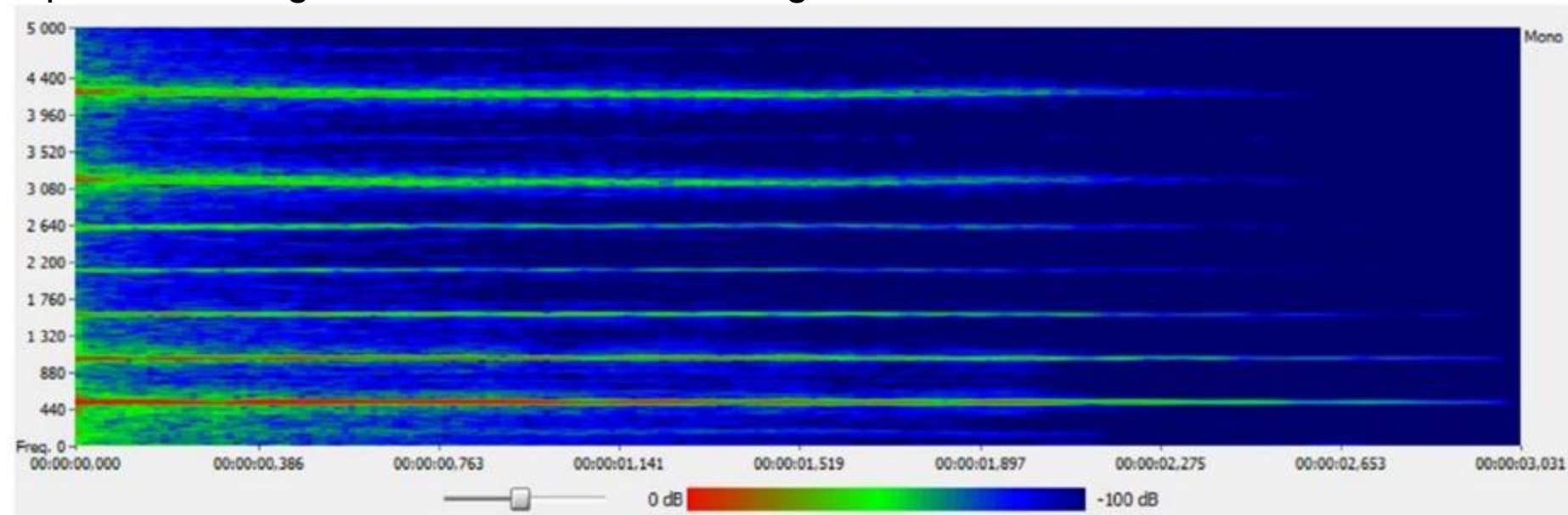
Piano



a) Original Piano Signal



b) Sampled Piano Signal Without an Anti-Aliasing Filter



- A text (**A4-sized, 50 lines, 80 characters per line**) is to be read aloud and **digitally stored**. Estimate the required **storage space** for the digitized audio using **44.1 kHz, 16-bit PCM format**, assuming a **female speaker**.
- **Outline the calculation steps** without a calculator. **Provide the final result in MB**.

Complete the following sentence with the correct statement:

**"When sampling speech at 10 kHz without an anti-aliasing filter and reconstructing it with a 4.5 kHz low-pass filter..."**

- The audio quality will certainly be better than telephone quality.
- Aliasing is prevented.
- Speech components near 9 kHz are filtered out.
- The output signal will be completely unintelligible and unusable.
- The signal can always be perfectly reconstructed.
- The signal will be mostly intelligible but noisy.

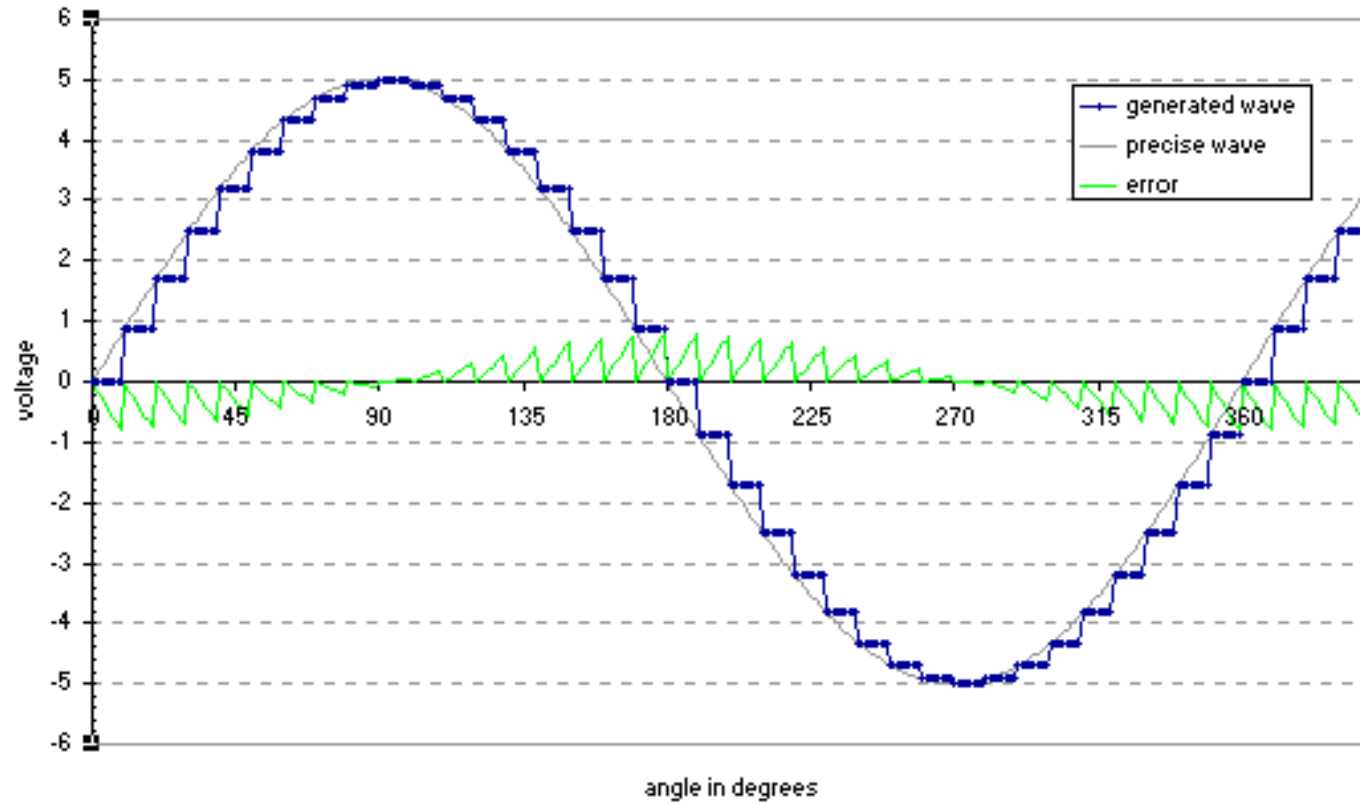
# QUANTIZATION

**A linear quantizer has a maximum signal-to-quantization noise ratio of 72 dB.**

**How many bits does the quantizer use?**

- **12 bits**
- **72 bits**
- **20 bits**
- **16 bits**
- **8 bits**
- **32 bits**

error in generating a sinewave from a series of values



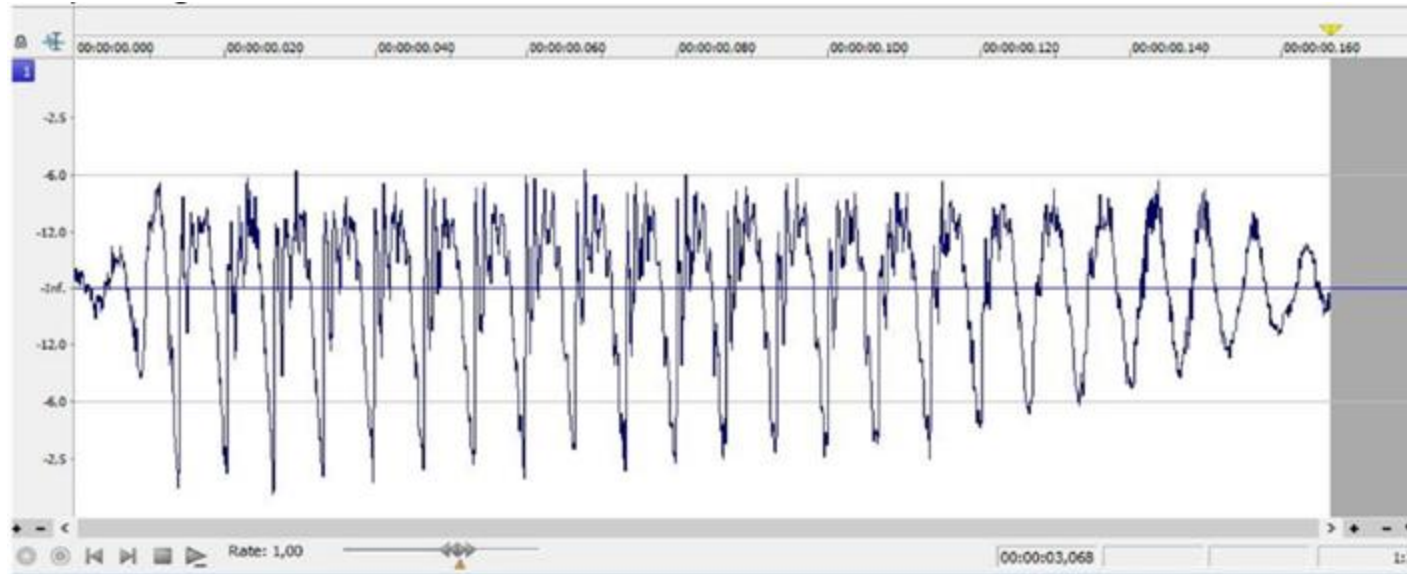
Complete the following statement with the correct option:

**"Quantization step size..."**

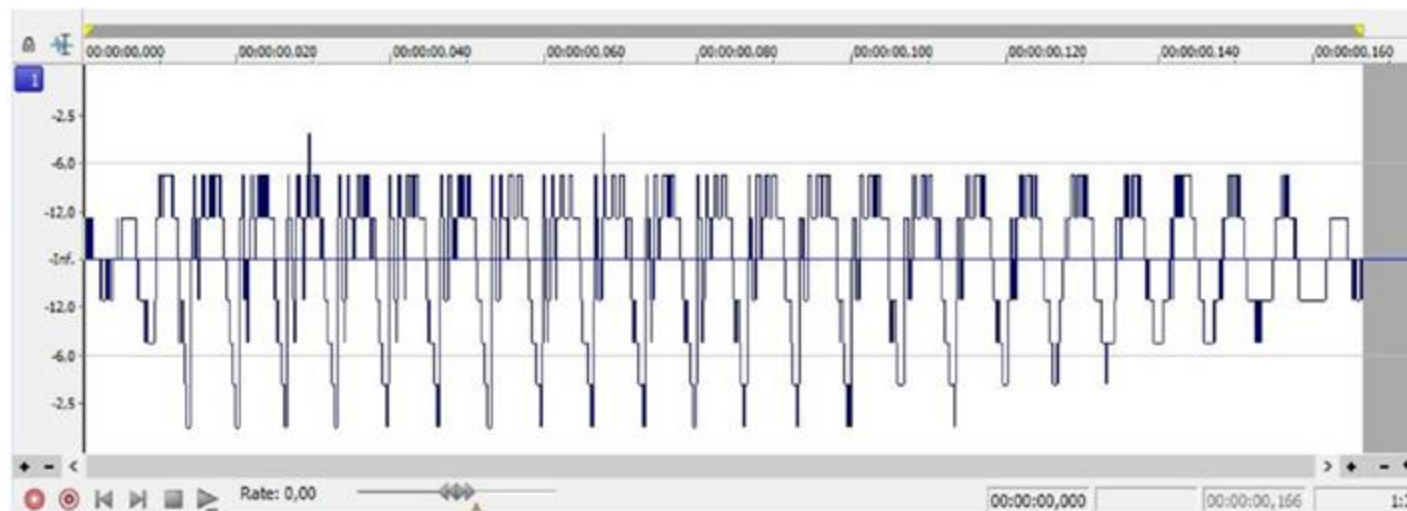
- Is independent of the sampling frequency.
- Depends on the sampling frequency.
- Can be perfectly reversed without loss.
- Does not affect audio quality.
- Is only used with uniform spacing for speech signals.
- Depends on the fundamental frequency of the speech.



a) Original Signal



b) Signal Quantized at a Low Bit Depth



# **SPEECH CODING**

Rank the following **speech coding methods** by **perceived speech quality** and justify your ranking.

- 8 kHz, 8-bit A-law quantization
- 2 kHz, 32-bit linear quantization
- 13 kbps GSM full-rate coder

Is there a linear relationship between the bitrate of speech coders and speech quality?

A speech database of 15 hours of audio was recorded in the studio at a sampling rate of 44.1kHz in mono. Julie encoded the speech database with a 64 kbps MP3 encoder and then gave the material to John in MP3 format. John decoded the resulting audio and then re-encoded it with an encoder using the A-law PCM algorithm. The re-compressed audio took up about 412 MB on John's hard drive.

Task:

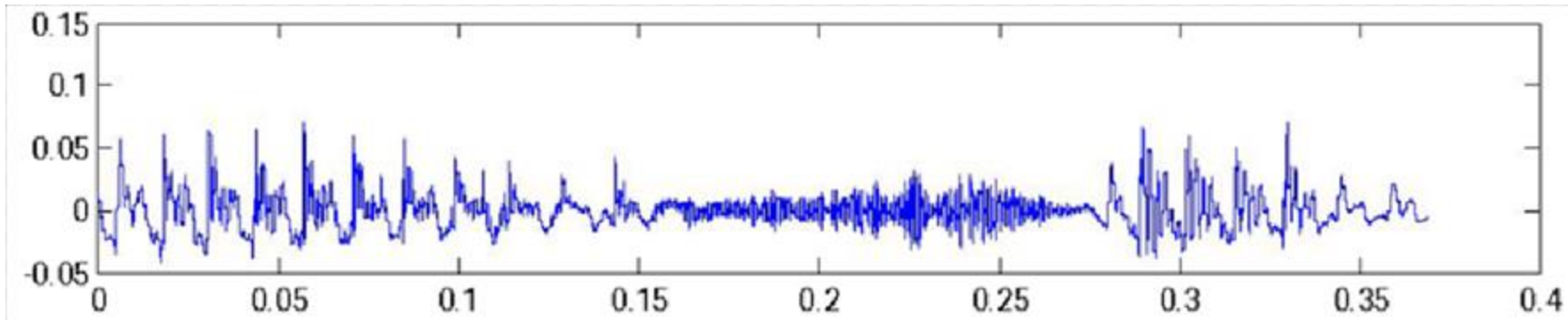
- Will the data that Julie gave to John fit in a 1 GB cloud storage?
- Is the audio compressed by John of sufficient quality for a telephone speech information system?
- Does John need to perform any signal processing steps before encoding?

- A **1-hour male speech sample** (8 kHz, 8-bit PCM) is **compressed using LPC-10**.
- The LPC-10 coder uses:
  - F0 values stored at 12 bits
  - Residual energy and LP coefficients stored at 10 bits each
  - Frame duration of 25 ms

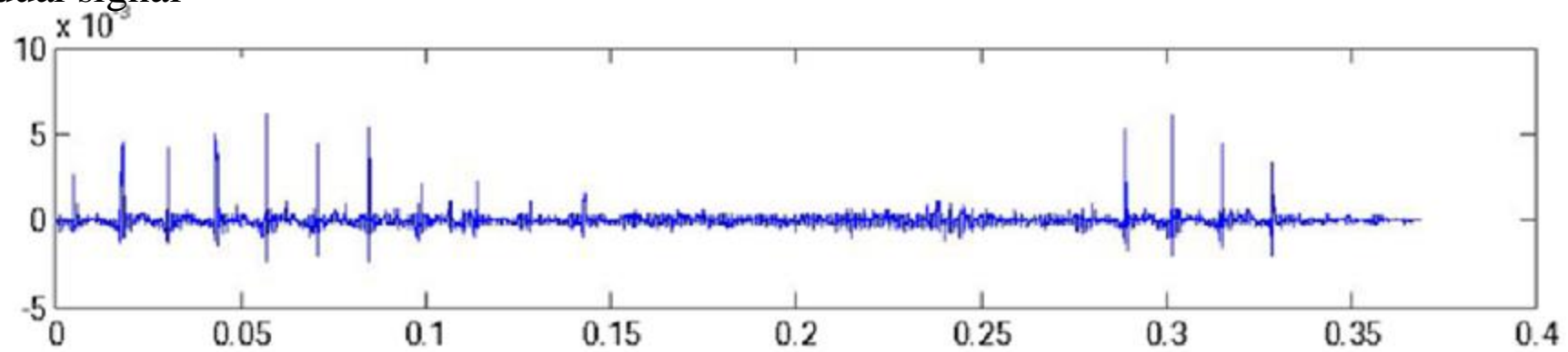
### **Tasks:**

- a) Calculate the compression ratio of the LPC-10 coder.
- b) Describe how to synthesize monotone speech using LPC-10.
- c) Explain the difference between whispered and normal speech and how to transform the LPC-10 speech to whispered speech.

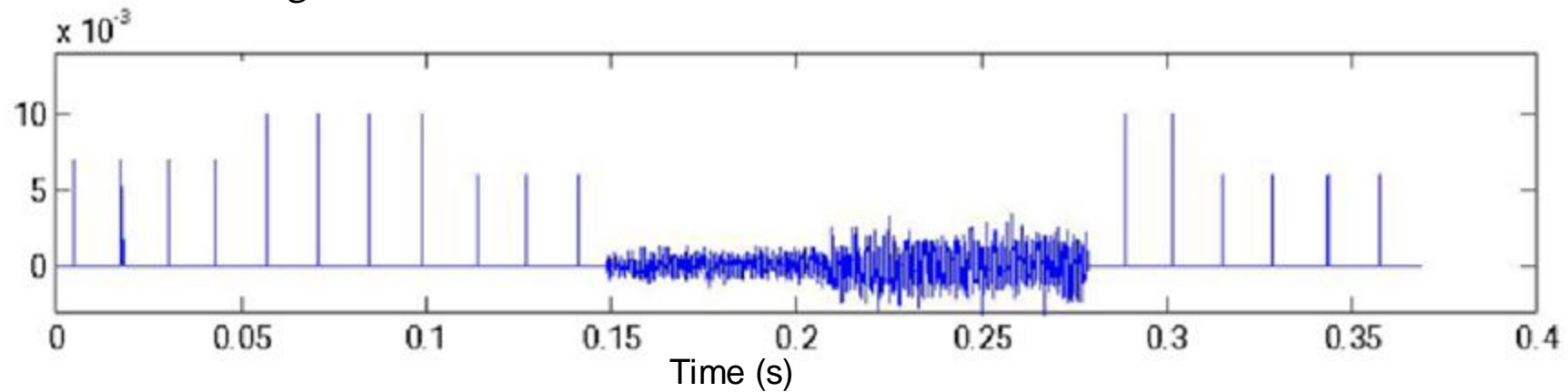
(a) Original Speech Signal



(b) residual signal



(c) LPC-10 excitation signal



a) Original Speech Signal



b) LPC-10 Compressed Speech



c) LPC-10 Monotone Speech



d) LPC-10 Whispered Speech



## What is characteristic of the **residual signal** in **LPC analysis**?

- It contains signal components not captured within the analysis window.
- Its energy is greater than the original signal.
- It can be discarded without affecting reconstruction.
- It requires fewer bits for quantization than the original signal.
- It contains formant frequency information.
- LPC synthesis can be performed without the original residual signal.



# BONUS TASK

- A **clean speech and music recording** is digitized at **24 kHz, 16-bit** with an **8 kHz low-pass filter** at the input.
- Playback uses:
  - 16-bit linear DAC (0-5V output range)
  - 24 kHz playback sampling rate
  - 8 kHz low-pass output filter

### **Tasks:**

- a) How much does the signal-to-noise ratio degrade when using an 8-bit linear DAC instead of a linear 16-bit DAC?
- b) What would change if the input filter cutoff frequency were doubled?

## Sampling, quantization, speech coding - 2

