Introduction to Speech Processing - Assignment 1

207377532, 206576142

April 30th, 2023

1 Theoretical Part

1.1 Multiplication and Convolution

$$F\left[x_1(t) \cdot x_2(t)\right] = \int_{-\infty}^{\infty} \left[x_1(t) \cdot x_2(t)\right] \cdot e^{-i\omega t}\, dt$$

From the definition of inverse Fourier transform, we get,

$$x(t) = F^{-1}\left[X^F(\omega)\right] = \frac{1}{2\pi}\int_{-\infty}^{\infty} X^F(\omega) \cdot e^{-i\omega t}\, d\omega$$

$$\Rightarrow F\left[x_1(t) \cdot x_2(t)\right] = \int_{-\infty}^{\infty} \left[\frac{1}{2\pi}\int_{-\infty}^{\infty} X_1^F(\rho) \cdot e^{i\rho t}\, d\rho\right] x_2(t) e^{-i\omega t}\, dt$$

By rearranging the order of integration, we get,

$$F\left[x_1(t) \cdot x_2(t)\right] = \frac{1}{2\pi}\int_{-\infty}^{\infty} X_1^F(\rho) \left[\int_{-\infty}^{\infty} x_2(t) e^{-i\omega t} e^{i\rho t} dt\right] d\rho$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} X_1^F(\rho) \left[\int_{-\infty}^{\infty} x_2(t) e^{-i(\omega-\rho)t} dt\right] d\rho$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} X_1^F(\rho) X_2^F(\omega - \rho)\, d\rho$$

By the definition of convolution:

$$= \frac{1}{2\pi}\left[X_1^F(\omega) * X_2^F(\omega)\right]$$

$\square$

1.2 Proving Nyquist Sampling Theorem

1.2.1

$$
\begin{aligned}
&\frac{1}{2\pi}\left(X^F(\omega) * S_T^F(\omega)\right)\\
&= F\left[x(t)\cdot s_T(t)\right]\\
&= F\left[\frac{1}{\sum_{n=-\infty}^{\infty}\delta(t-nT)}\cdot x_d(t)\cdot \sum_{n=-\infty}^{\infty}\delta(t-nT)\right]\\
&= F\left[x_d(t)\right]\\
&= X_d^F(\omega)
\end{aligned}
$$

$\square$

1.2.2

$$
\sum_n \int_{-\infty}^{\infty} X^F(\tilde{\omega})\cdot \overbrace{\delta\left(\tilde{\omega}-\left(\omega-\frac{2\pi n}{T}\right)\right)}^{\substack{equals\ 1 \iff \tilde{\omega}-\left(\omega-\frac{2\pi n}{T}\right)=0 \\ \iff \tilde{\omega}=\omega-\frac{2\pi n}{T}}}\ d\tilde{\omega}
$$

$$
= \sum_n X^F\left(\omega-\frac{2\pi n}{T}\right)
$$

$\square$

1.2.3

$$
\begin{aligned}
&\frac{1}{T}\sum_n X^F\left(\omega-\frac{2\pi n}{T}\right) = \frac{1}{T}\sum_n\int_{-\infty}^{\infty} X^F(\tilde{\omega})\cdot\delta\left(\tilde{\omega}-\left(\omega-\frac{2\pi n}{T}\right)\right)\ d\tilde{\omega}\\
&= \frac{1}{T}\int_{-\infty}^{\infty} X^F(\tilde{\omega})\cdot \overbrace{\sum_n \delta\left(\tilde{\omega}-\left(\omega-\frac{2\pi n}{T}\right)\right)}^{\frac{T}{2\pi}S_T^F(\tilde{\omega}-\omega)}\ d\tilde{\omega}\\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} X^F(\tilde{\omega})\cdot S_T^F\left(\tilde{\omega}-\omega\right)\ d\tilde{\omega}\\
&= \frac{1}{2\pi}X^F(\omega) * S_T^F(\omega)\\
&= X_d^F(\omega)
\end{aligned}
$$

$\square$

1.2.4

$\underline{\Leftarrow}$: Assuming that $f_s > f_{max}$, we get:

$$X_d^F(\omega) = \frac{1}{T}\sum_n \overbrace{X^F\left(\omega - \frac{2\pi n}{T}\right)}^{X^F\left(\omega-\frac{2\pi n}{T}\right)\neq 0 \iff |\omega-\frac{2\pi n}{T}|\leq\omega_{max} \overset{\star}{\iff} n=0} = \frac{1}{T}X^F(\omega)$$

$\underline{\star}$: $\underline{n=0}$:

$$|\omega - \frac{2\pi n}{T}| = |\omega| \leq w_{max}$$

$\underline{n \neq 0}$:

$$|\omega - \frac{2\pi n}{T}| = |\omega - 2\pi f_s n| = |\omega - \omega_s n|$$

$\underline{sign(\omega) = sign(\omega_s n)}$:

$$|\omega - \omega_s n| > |\omega - 2\omega_{max} n| \geq \omega_{max}$$

$\underline{sign(\omega) \neq sign(\omega_s n)}$:

$$|\omega - \omega_s n| > |2\omega_s n| > 2\omega_{max}$$

Therefore:

$$|\omega - \frac{2\pi n}{T}| > \omega_{max}$$

$\underline{\Rightarrow}$: Under the assumption that $X_d^F(\omega) = \frac{1}{T}X^F(\omega)$, we will assume towards contradiction that $f_s \leq 2f_{max}$. From the base assumption we get:

$$X_d^F(\omega_{max}) = \frac{1}{T}\sum_n X^F\left(\omega_{max} - \frac{2\pi n}{T}\right)$$

$\underline{n=0}$:

$$X^F\left(\omega_{max} - \frac{2\pi n}{T}\right) = X^F(\omega_{max})$$

$\underline{n=1}$:

$$|\omega_{max} - \frac{2\pi n}{T}| = |\omega_{max} - \omega_s n| \overset{2\omega_{max}\geq\omega_s>0}{\leq} |\omega_{max} - 2\omega_{max} n| =$$
$$= |\omega_{max} - 2\omega_{max}| = \omega_{max} \qquad \rightarrow X^F(\omega_{max} - \omega_s) \neq 0$$

But that is a contradiction:

$$X_d^F(\omega_{max}) = \frac{1}{T}(X^F(\omega_{max}) + X^F(\omega_{max} - \omega_s) + ...) \neq \frac{1}{T}X^F(\omega_{max})$$

$\square$

1.3 Fourier Transform

$$\forall\omega : X^F(\omega) = F_\omega(x(t))) = \sum_t x(t)e^{-it\omega}$$
$$= \sum_t (sin(2\pi\cdot 1000\cdot t) + sin(2\pi\cdot 5000\cdot t))\cdot e^{-it\omega}$$
$$= \sum_t sin(2\pi\cdot 1000\cdot t)\cdot e^{-it\omega} + \sum_t sin(2\pi\cdot 5000\cdot t)\cdot e^{-it\omega}$$
$$= \sum_t \frac{e^{2\pi\cdot t\cdot 1000\cdot i} - e^{-2\pi\cdot t\cdot 1000\cdot i}}{2i}\cdot e^{-it\omega} + \sum_t \frac{e^{2\pi\cdot t\cdot 5000\cdot i} - e^{-2\pi\cdot t\cdot 5000\cdot i}}{2i}\cdot e^{-it\omega}$$

$$= \frac{1}{2i}(\sum_t e^{2\pi\cdot t\cdot 1000\cdot i}\cdot e^{-it\omega} - \sum_t e^{-2\pi\cdot t\cdot 1000\cdot i}\cdot e^{-it\omega}$$
$$+ \sum_t e^{2\pi\cdot t\cdot 5000i}\cdot e^{-it\omega} - \sum_n e^{-2\pi\cdot t\cdot 5000i}\cdot e^{-it\omega})$$
$$= \frac{1}{2i}(\sum_t e^{-it(\omega-1000\cdot 2\pi)} - \sum_t e^{-it(\omega+1000\cdot 2\pi)}$$
$$+ \sum_t e^{-it(\omega-5000\cdot 2\pi)} - \sum_t e^{-it(\omega+5000\cdot 2\pi)})$$
$$= \frac{1}{2i}(\delta(\omega-1000\cdot 2\pi) - \delta(\omega+1000\cdot 2\pi) + \delta(\omega-5000\cdot 2\pi) - \delta(\omega+5000\cdot 2\pi))$$

Considering the fact that $\forall|\omega| > 5000\cdot 2\pi, X^F(\omega) = 0$ and the equation:

$$X_d^F(\omega) = \frac{1}{T}\sum_n X^F\left(\omega - \frac{2\pi n}{T}\right) = 8000\sum_n X^F(\omega - 8000\cdot 2\pi n)$$

For every integer $n \in (-\infty, \infty)$, the frequencies to appear in the Fourier Transform of the digital measured signals are:

$$X_d^F(1000\cdot 2\pi + 8000\cdot 2\pi\cdot n) = 8000\cdot X^F(1000\cdot 2\pi)$$
$$X_d^F(-1000\cdot 2\pi + 8000\cdot 2\pi\cdot n) = 8000\cdot X^F(-1000\cdot 2\pi)$$
$$X_d^F(5000\cdot 2\pi + 8000\cdot 2\pi\cdot n) = 8000\cdot X^F(5000\cdot 2\pi)$$
$$X_d^F(-5000\cdot 2\pi + 8000\cdot 2\pi\cdot n) = 8000\cdot X^F(-5000\cdot 2\pi)$$

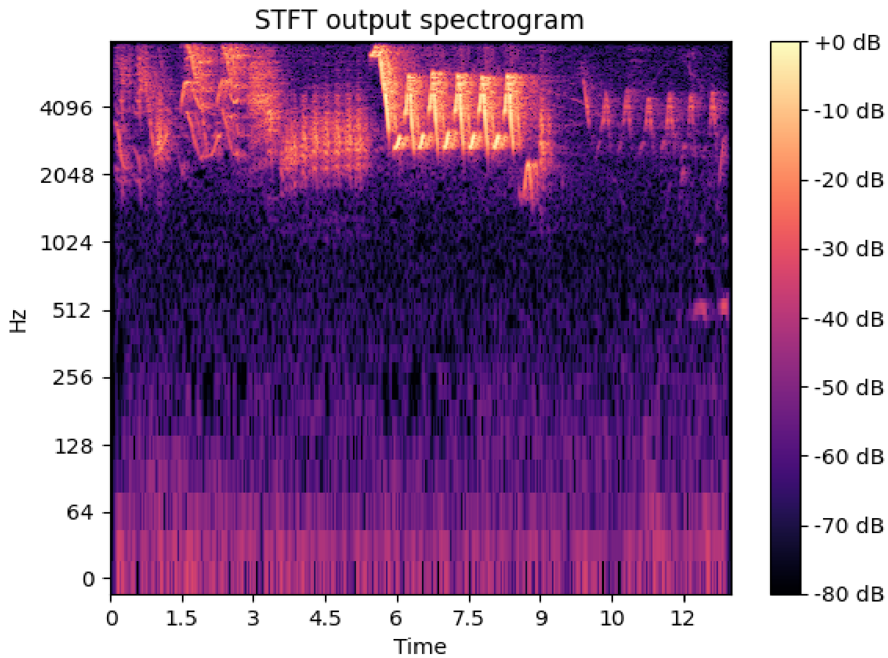
The signals of the frequencies to be sampled are: $X^F(1000\cdot 2\pi), X^F(-1000\cdot 2\pi), X^F(5000\cdot 2\pi), X^F(-5000\cdot 2\pi)$, but since the sample rate is $8000\cdot 2\pi$, the only signals we can reconstruct are $|\omega| < 4000\cdot 2\pi$, in that case, the signals: $X_d^F(1000\cdot 2\pi), X_d^F(-1000\cdot 2\pi)$. The rest of the frequencies are with $X_d^F(\omega) = 0$.

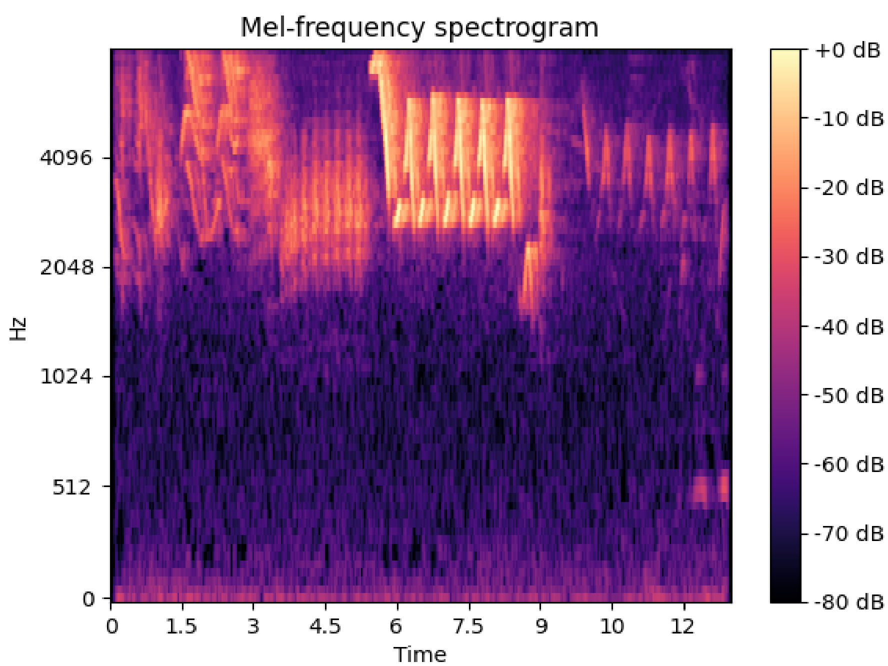
□

2 Practical Part

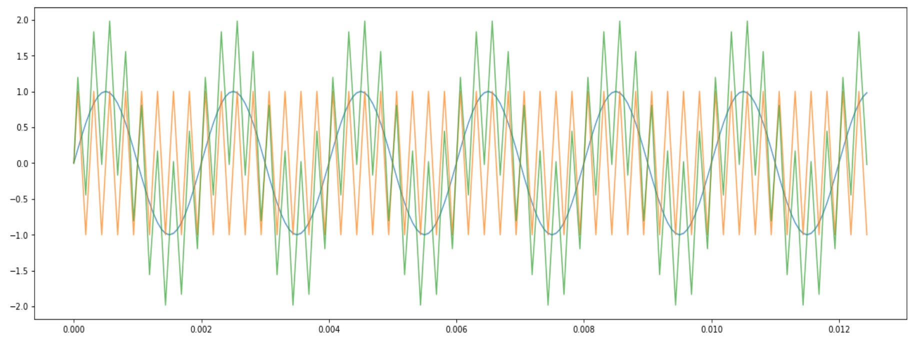
2.1 Basic Audio Handling

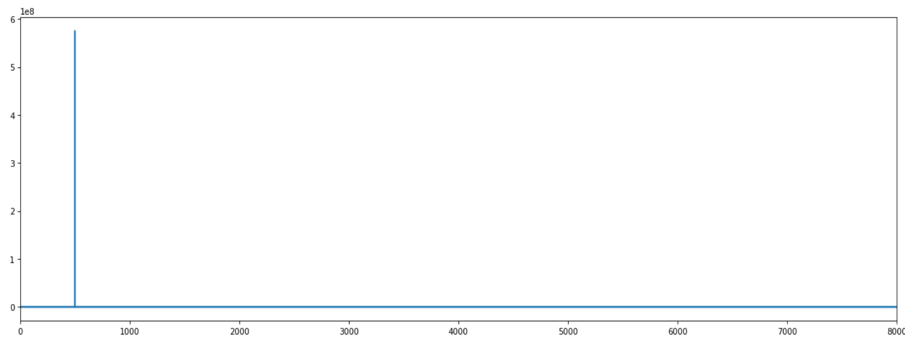
2.1.1 Part A

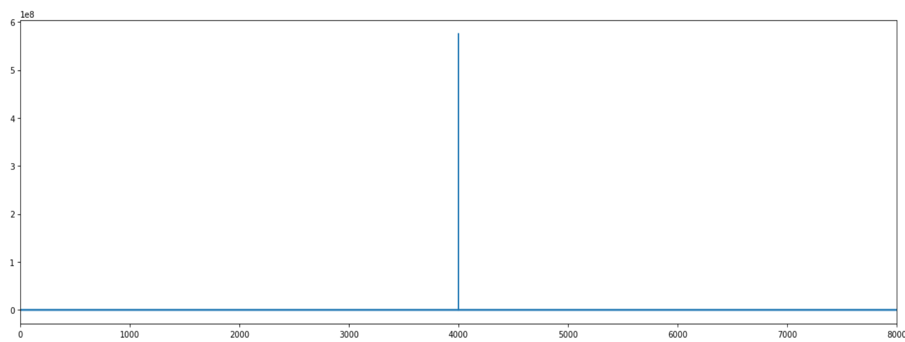


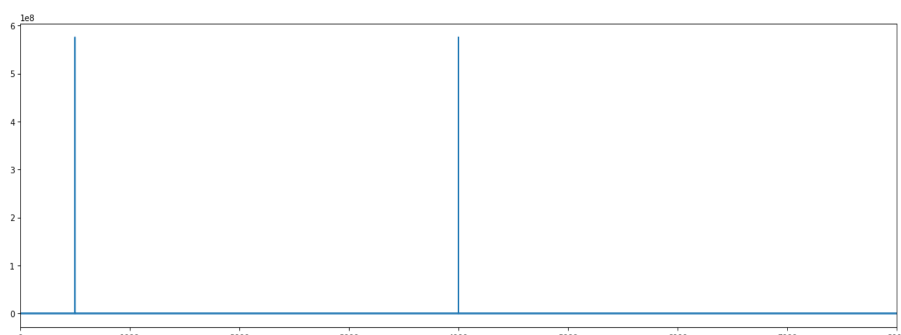


2.1.2 Part B

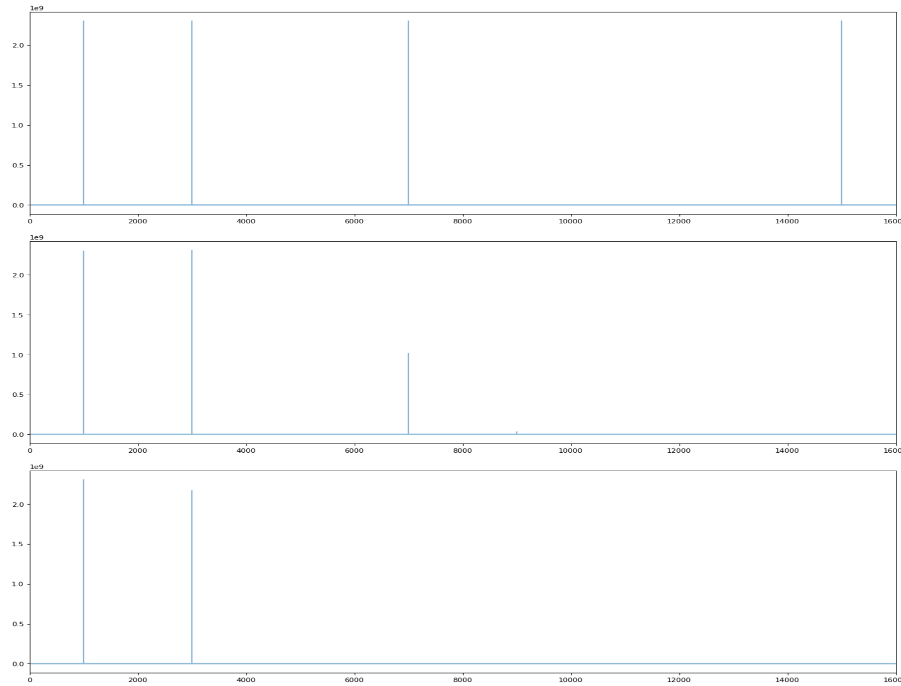

2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0.000
0.002
0.004
0.006
0.008
0.010
0.012


1e8
6
5
4
3
2
1
0
0
1000
2000
3000
4000
5000
6000
7000
8000


1e8
6
5
4
3
2
1
0
0
1000
2000
3000
4000
5000
6000
7000
8000


1e8
6
5
4
3
2
1
0
0
1000
2000
3000
4000
5000
6000
7000
8000

2.1.3 Part C

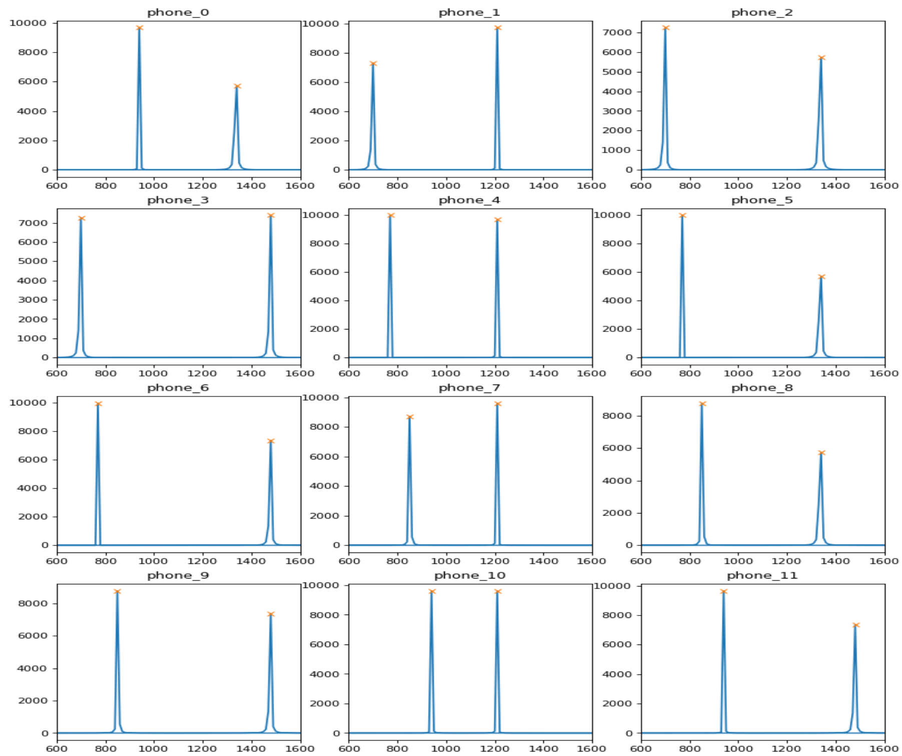


4. From Nyquist Sampling Theorem, after resampling with a sample rate of 32KHz, the signal frequencies from the resampled signals are in the range of 0-16KHz. Therefore, the x-axis of the plot would include the amplitudes values as a function of the FFT values in that range.

5. According to Nyquist Sampling Theorem, by downsampling from a rate of 32Khz to 16Khz, we can record only frequencies that are under 8Khz. That is why the frequncies of 1KHz, 3KHz and 7KHz are preserved, differently from the other frequency of 15KHz, which its singals is lost and so is its amplitude. Moreover, the frequency of 7KHz is relatively close to 8KHz compared to the other frequencies, that is why the signal amplitude is weaker than the others. From the same reason, we can also see a small noise over the frequency of 9KHz, resulting with a non-trivial amplitude.

Similarly, by downsampling from a rate of 16Khz to 8Khz, we can record only frequencies that are under 4Khz. That is why the frequncies of 1KHz and 3KHz are preserved, differently from the other frequencies of 7KHz and 15KHz, which their singals are lost and so are their amplitudes. Same as before, the frequency of 3KHz is relatively close to 4KHz compared to the frequency of 1KHz, resulting that the signal amplitude is weaker than the other ones.

2.1.4 Part D



2.2 Naive Audio Manipulation

2.2.1 Part A

3. The odd thing we observed while playing the new audio files, is that the original audio was preserved in both cases, but its pitch became slower and lower/faster and higher, in the streched/compressed versions respectfully.
We think it is caused from the interpolation, which stretched/compressed the audio waves, and as a result the frequency became lower/higher on the one hand, and on the other hand it is preserved the original data, it made the wave lengths longer/shorter accordingly.

2.2.2 Part C

3. The current time stretch is with much less noise compared to the previous naive one. We believe it is because of the hanning windows, which made the stft outputs more coherent around the center of the windows, therefore making diffenece of the stft outputs angles less noisy for the istft operation. Finally, the recovered signal was resulted with much less noise, thanks to that.