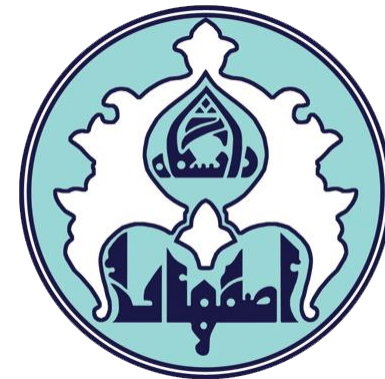




Digital Speech Processing

به نام خدا



گروه هوش مصنوعی، دانشکده مهندسی
کامپیوتر

گفتار پردازش رقمی

رقمی سازی و پیش پردازش سیگنال گفتار

حمیدرضا برادران کاشانی

پاییز ۱۴۰۱



سرفصل مطالب

- ❖ رقصی سازی
- ❖ نمونه برداری
- ❖ چندی سازی
- ❖ پیش پردازش
- ❖ فریم بندی
- ❖ پیش تاکید
- ❖ پنجره گذاری

Hamidreza Baradaran Kashani



رقمی سازی

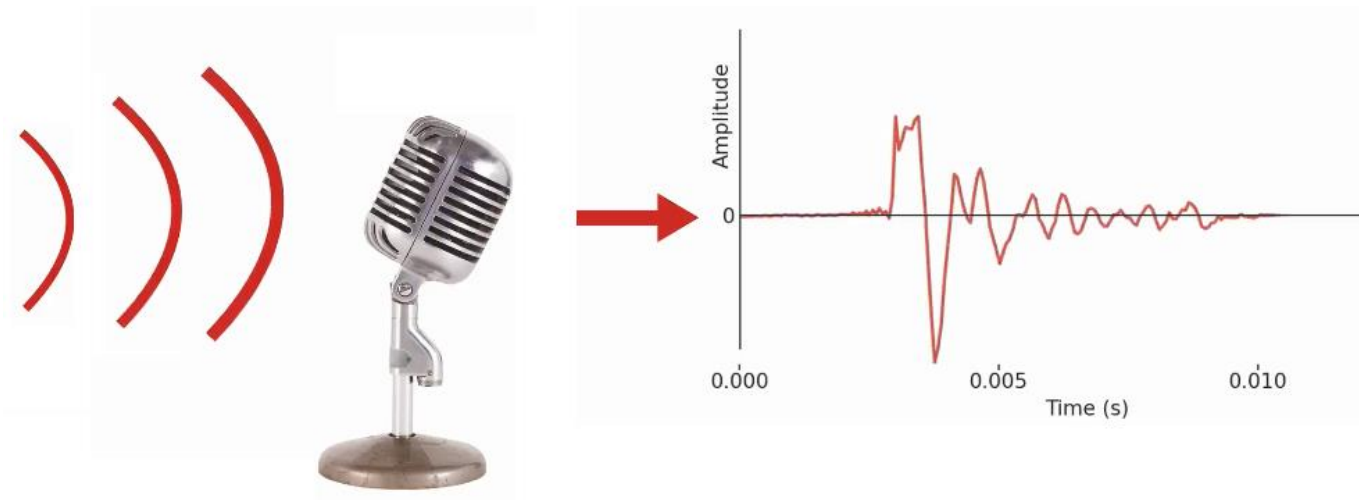
❖ اولین گام در آماده سازی سیگنال گفتار

❖ رقمی سازی (Digitization) سیگنال گفتار: تبدیل سیگنال آنالوگ به دیجیتال

❖ گفتار انسان در هوا بصورت موج انتشار می یابد.

❖ میکروفون این موج صوتی را دریافت کرده و تبدیل به یک سیگنال الکتریکی (ولتاژ) کرده که یک سیگنال پیوسته در دامنه و زمان است (سیگنال آنالوگ)

چرا کامپیوترها نمی توانند سیگنال آنالوگ را ذخیره کنند؟



Hamidreza Baradaran Kashani



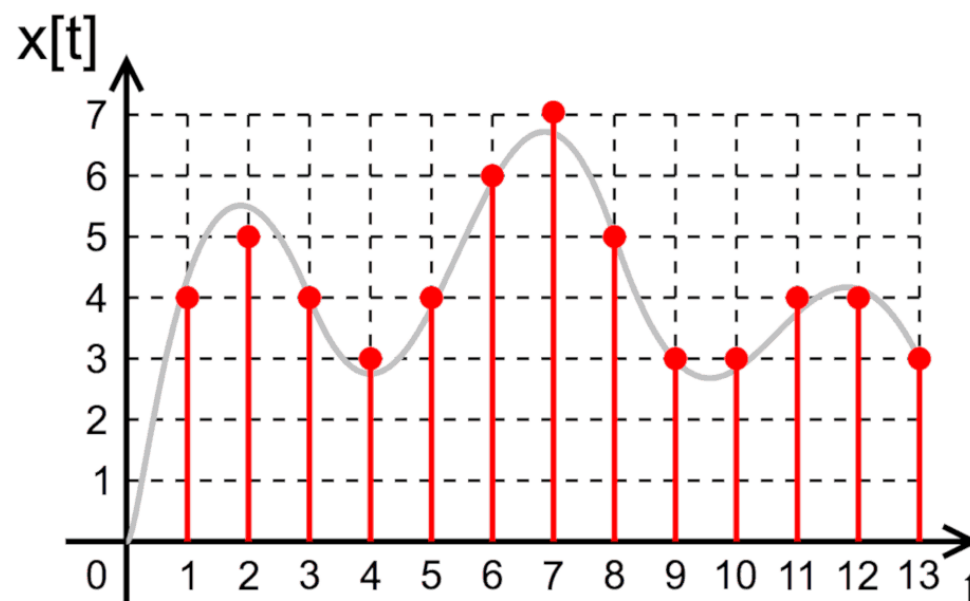
رقمی سازی

❖ کامپیوترها حجم ذخیره سازی محدودی دارند و فقط اعداد باینری را ذخیره می کنند

❖ رقمی سازی یعنی گسسته سازی سیگنال هم در دامنه و هم در زمان

❖ گسسته سازی سیگنال در زمان: نمونه برداری (Sampling)

❖ گسسته سازی سیگنال در دامنه: چندی سازی (Quantization)



Hamidreza Baradaran Kashani



نمونه برداری

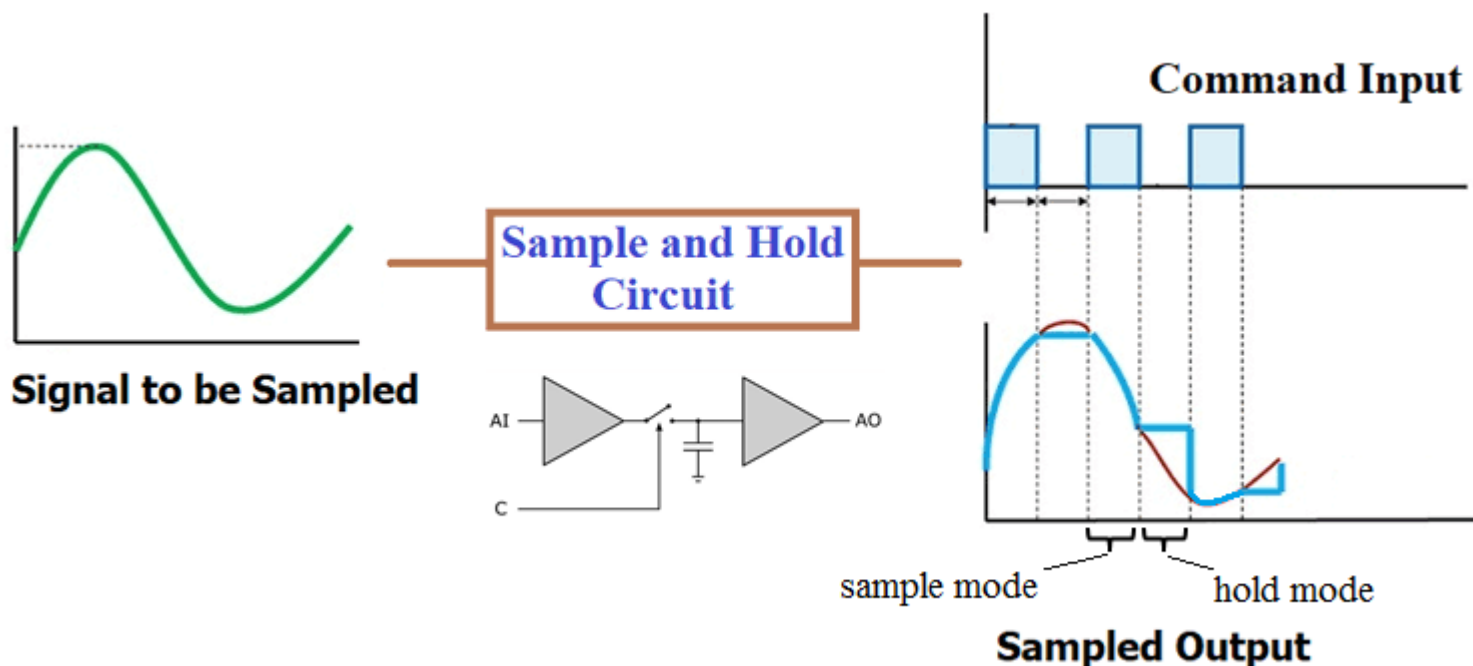
❖ نمونه برداری معمولاً توسط یک نمونه بردار با نام sample&hold انجام می شود.

❖ سرعت باز و بسته شدن کلید: فرکانس نمونه برداری

❖ کلید بسته: ولتاژ ورودی باعث

شارژ خازن و مقدار ولتاژ خروجی برابر ولتاژ ورودی

❖ کلید باز: نگه داشتن ولتاژ خروجی توسط خازن

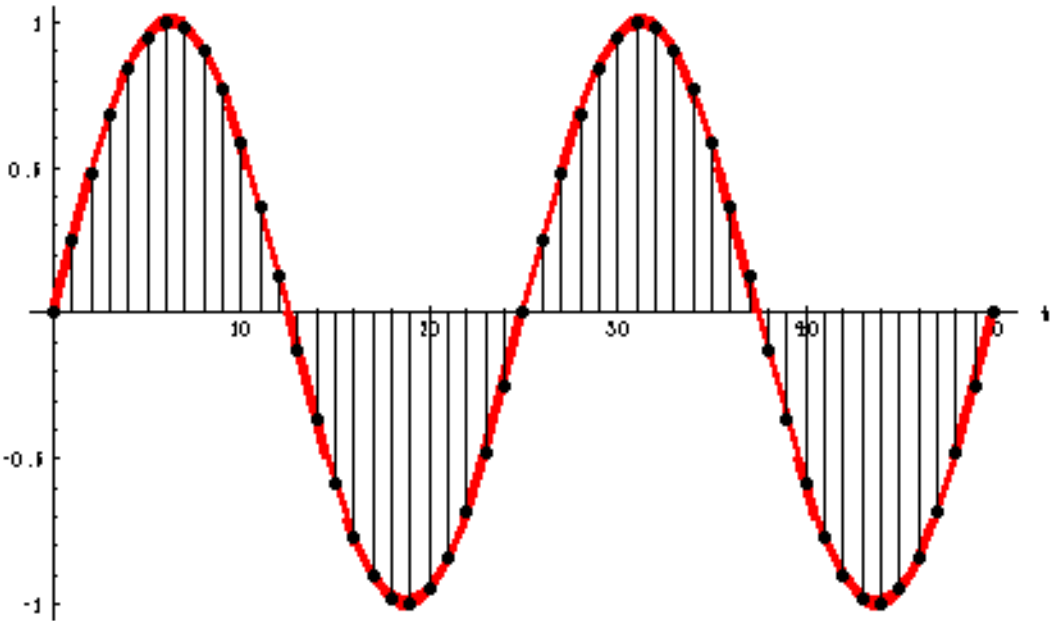




نمونه برداری

❖ مهمترین مساله:

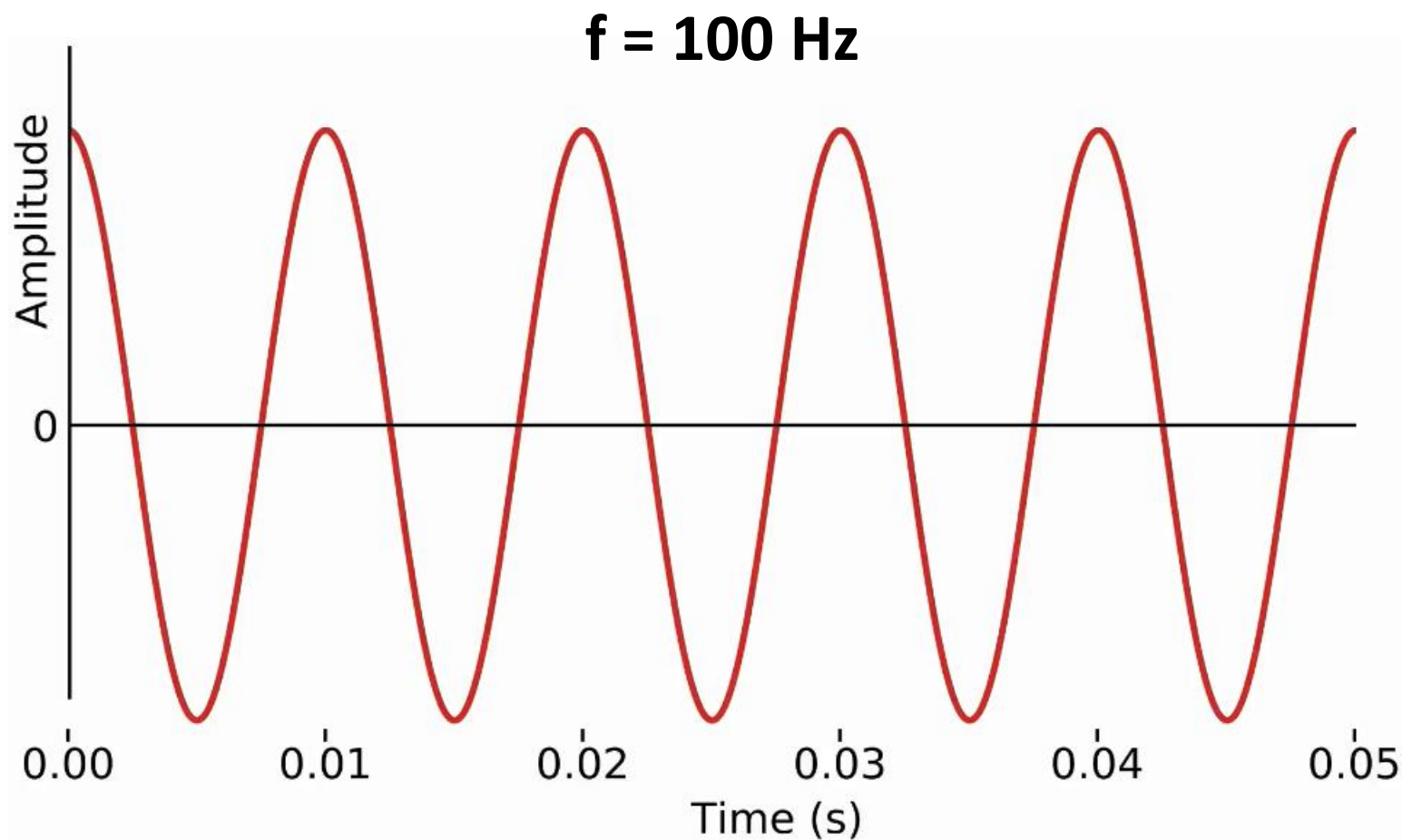
حداقل مقدار فرکانس نمونه برداری چقدر می تواند باشد؟



Hamidreza Baradaran Kashani



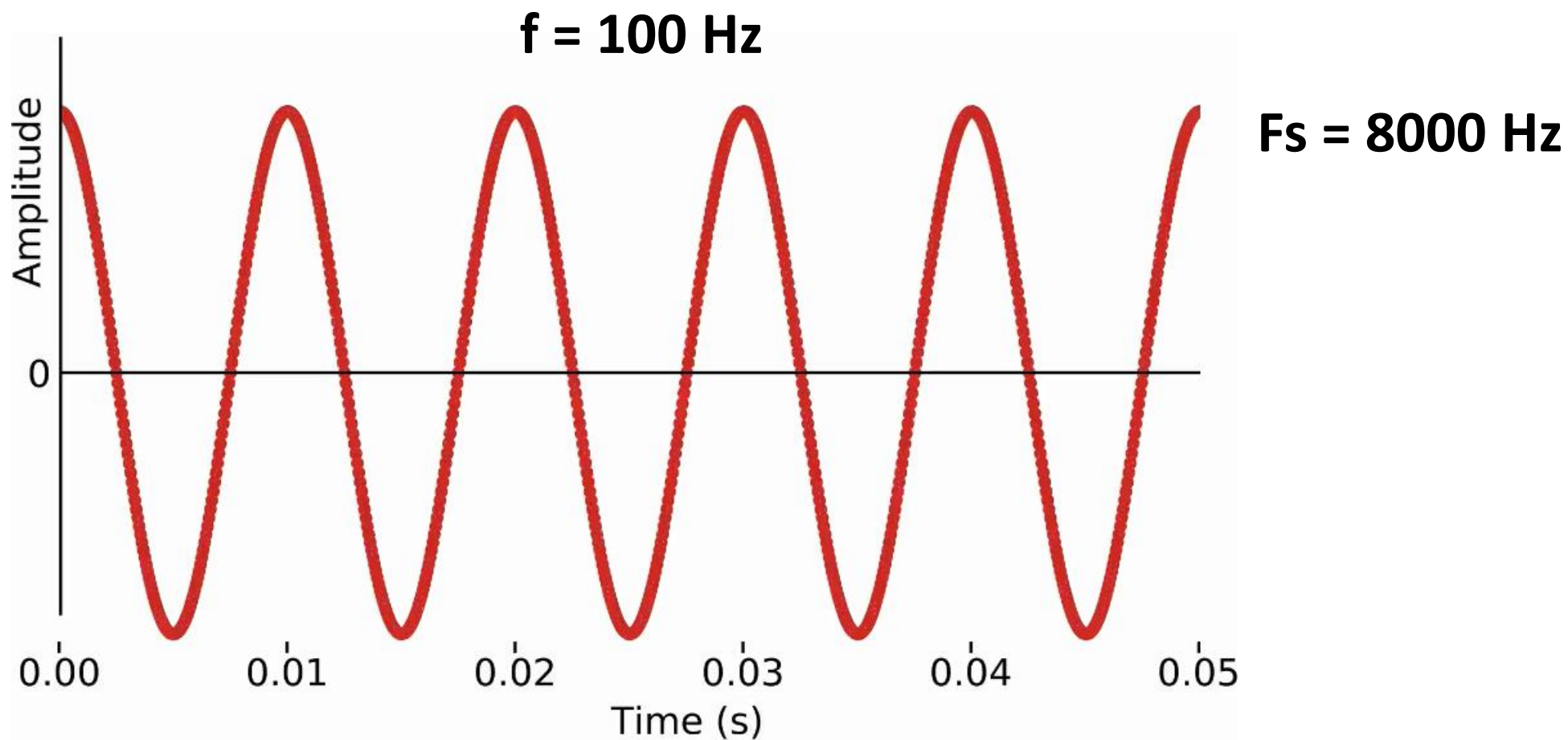
نمونه برداری



Hamidreza Baradaran Kashani



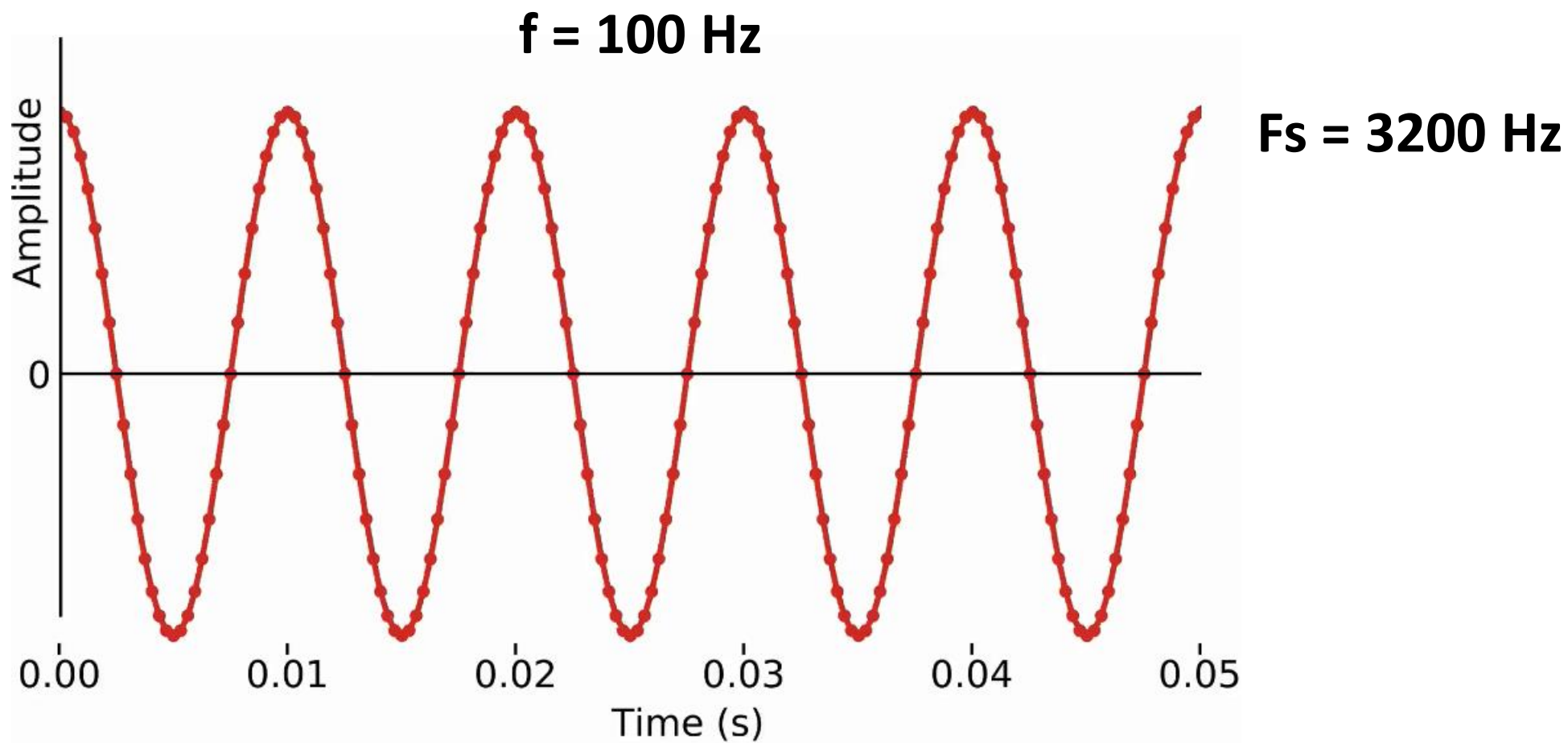
نمونه برداری



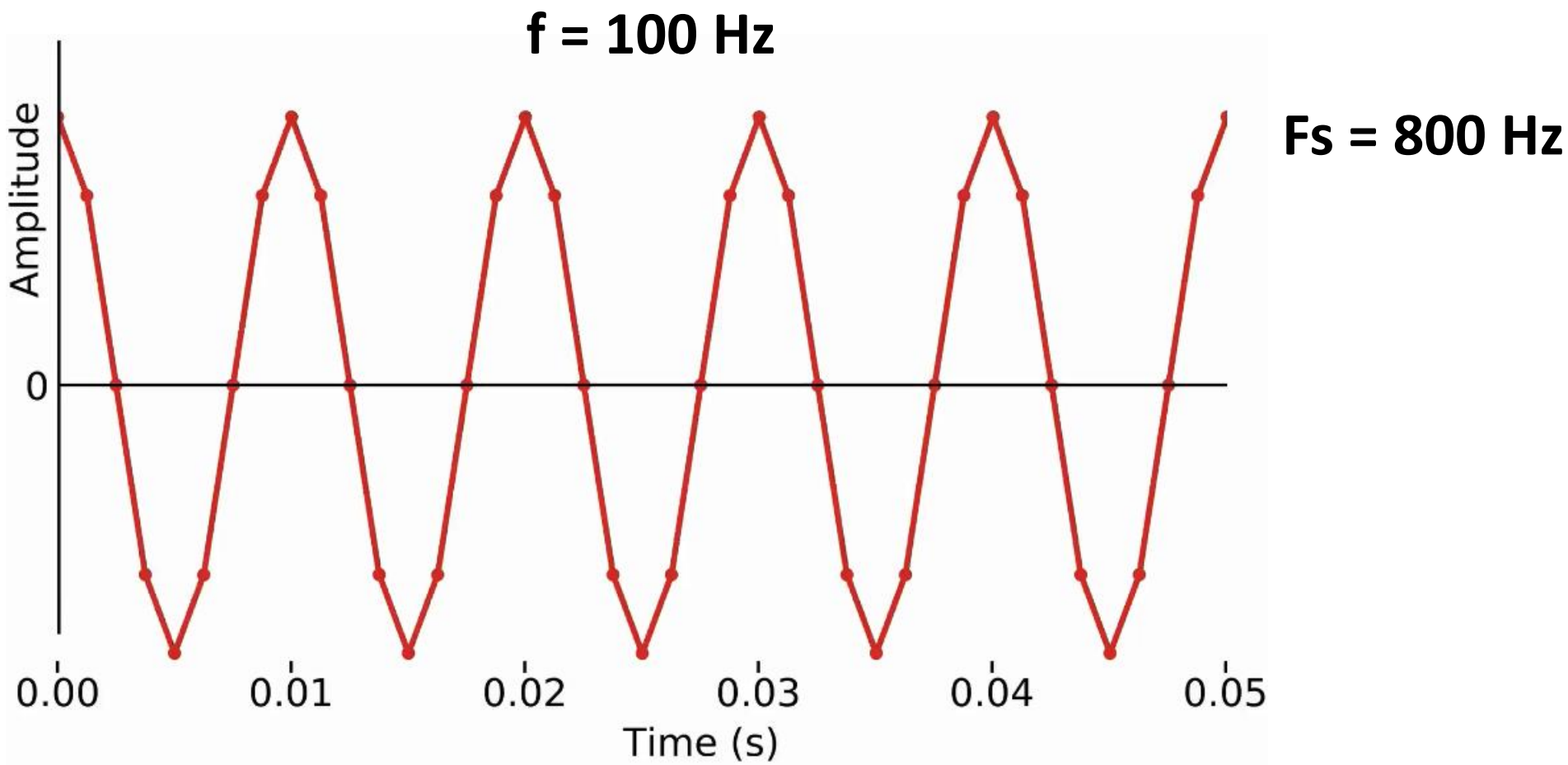
Hamidreza Baradaran Kashani



نمونه برداری



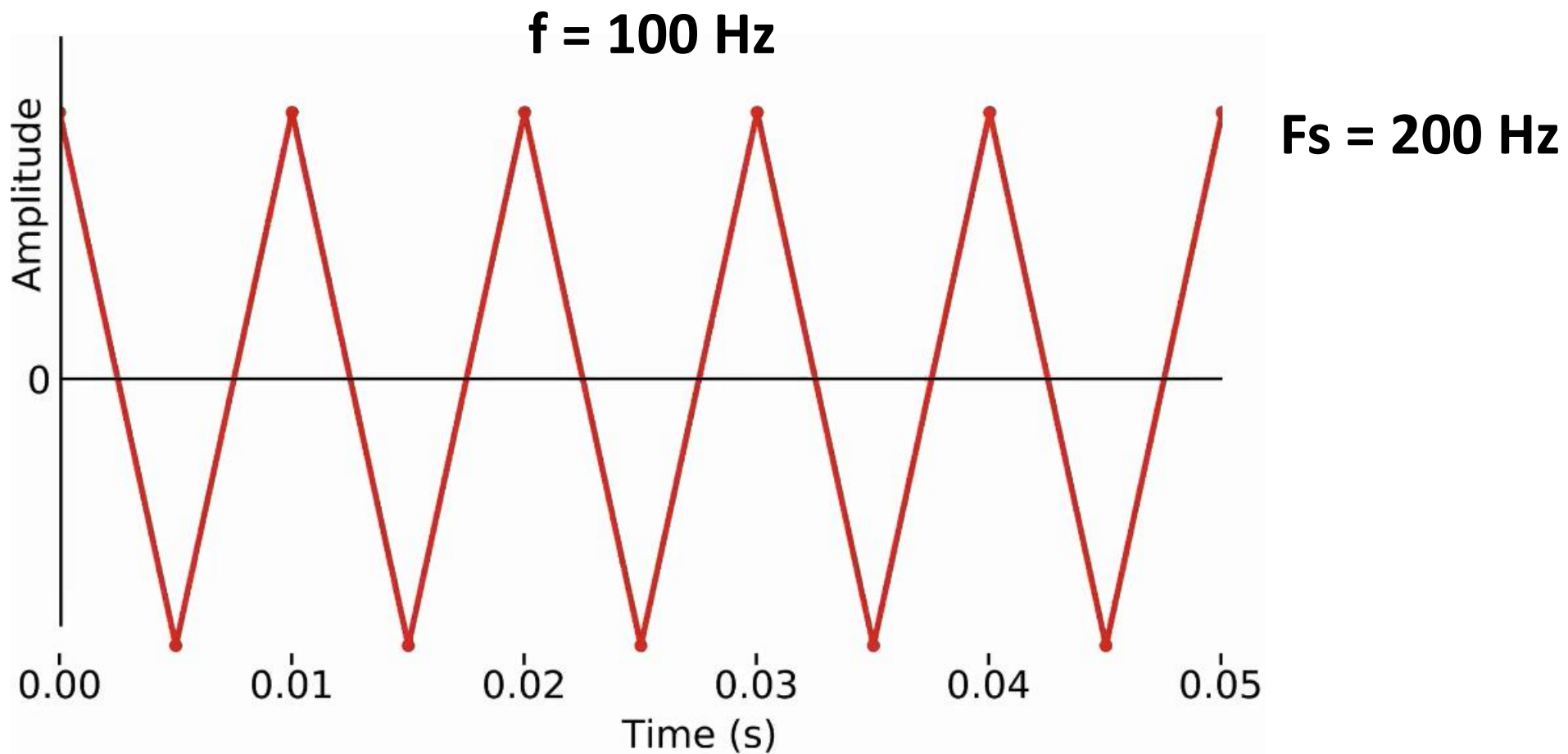
Hamidreza Baradaran Kashani



Hamidreza Baradaran Kashani



نمونه برداری



Hamidreza Baradaran Kashani



نمونه برداری

❖ در واقع مشاهده می شود که حداقل فرکانس نمونه برداری ۲ برابر بزرگترین فرکانس موجود در سیگنال است.

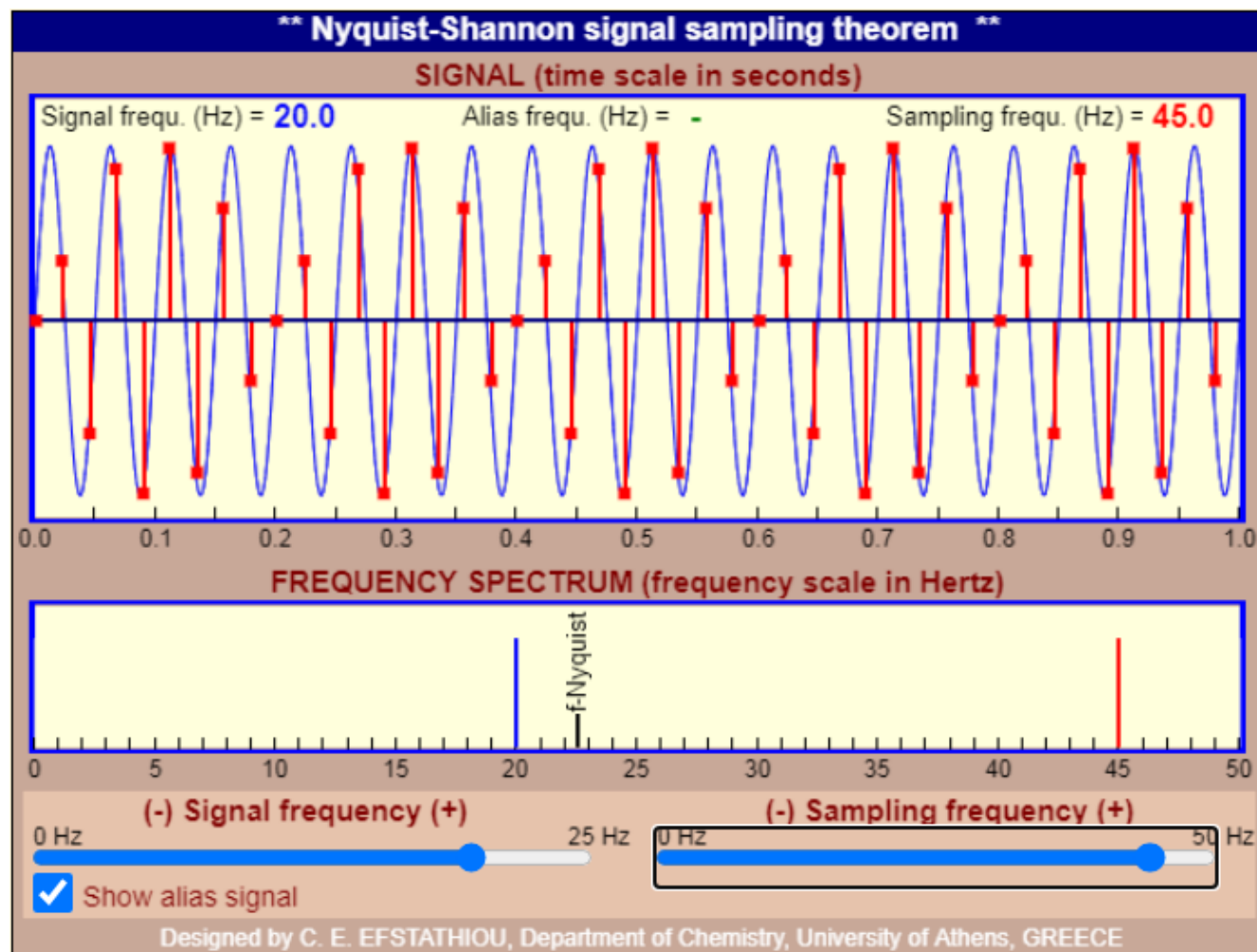
❖ به نصف فرکانس نمونه برداری، فرکانس نایکوئیست (Nyquist) می گوئیم.

❖ طبق قضیه نمونه برداری نایکوئیست-شنون،

❖ برای جلوگیری از رخ دادن پدیده **aliasing** هنگام بازسازی سیگنال از نمونه های نتیجه شده از نمونه برداری، فرکانس نایکوئیست سیستم نمونه بردار، بایستی بیشتر از دو برابر بزرگترین فرکانس موجود در سیگنال باشد.



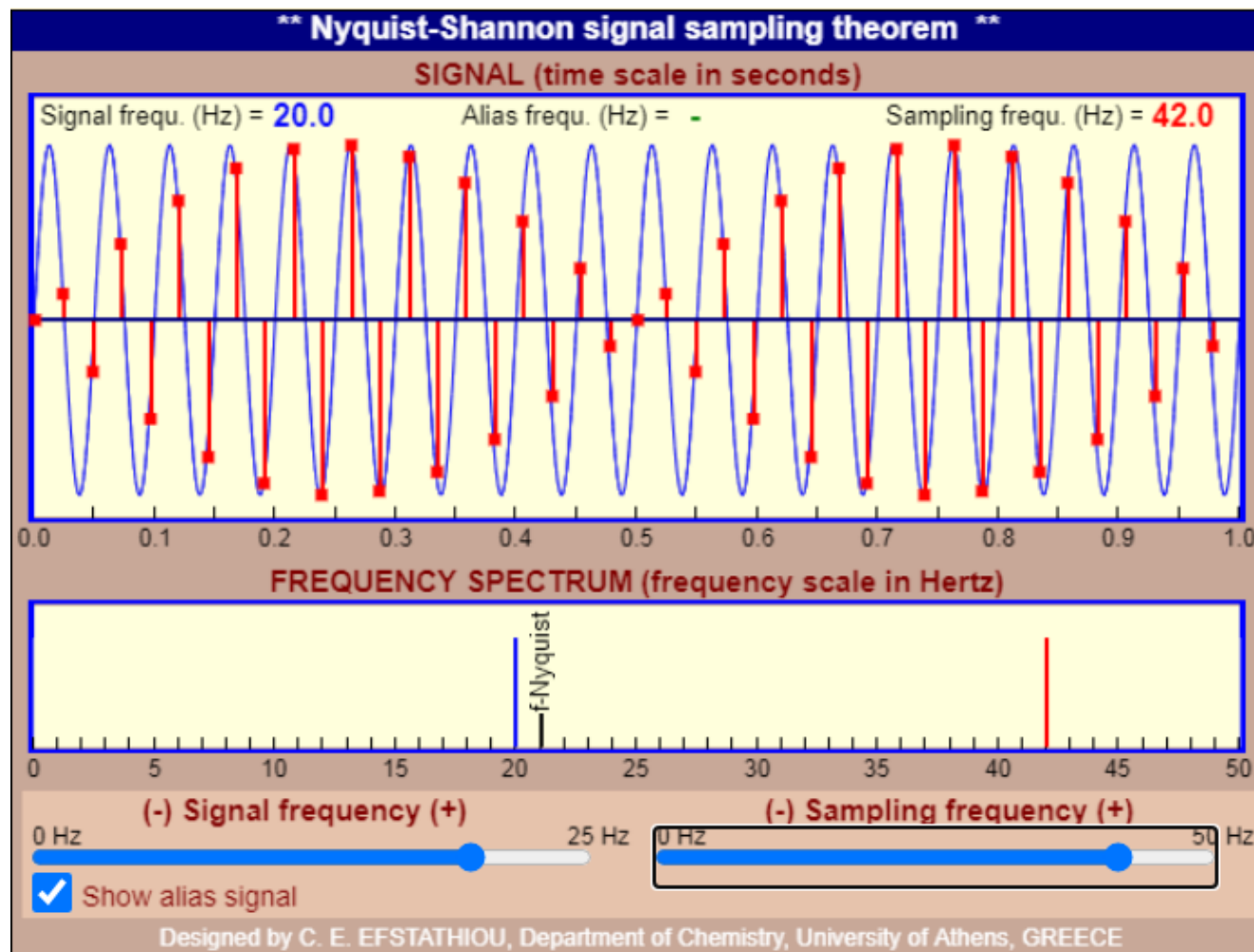
نمونه برداری



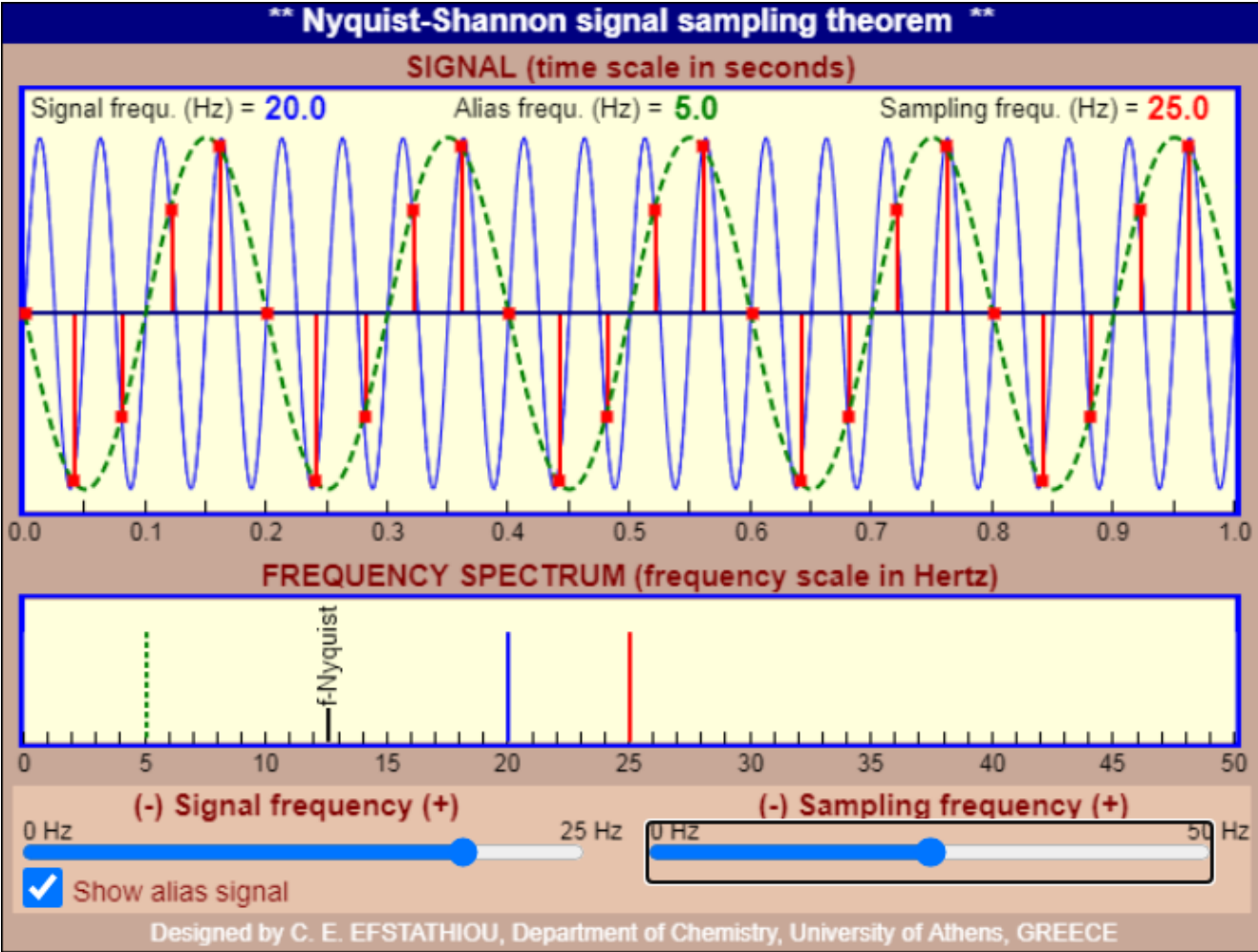
Hamidreza Baradaran Kashani



نمونه برداری



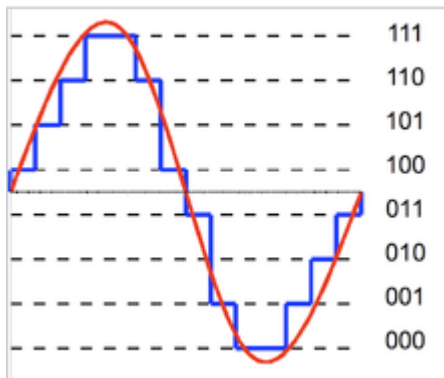
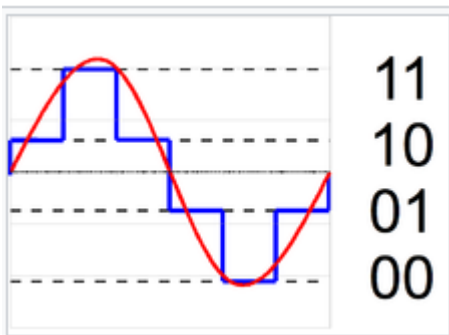
Hamidreza Baradaran Kashani





چندی سازی (Quantization)

- ❖ فرآیند چندی سازی در حوزه دامنه سیگنال را گسسته می کند.
- ❖ هر چه تعداد سطوح چندی سازی بیشتر باشد سیگنال به سیگنال آنالوگ نزدیکتر است.
- ❖ به تعداد بیت مورد استفاده برای چندی سازی ، رزولوشن یا **bit depth** می گوئیم.
- ❖ یک مقدار **bit depth** رایج در حوزه چندی سازی پردازش گفتار، مقدار ۱۶ بیت است. به عبارتی هر نمونه سیگنال آنالوگ با ۱۶ بیت ذخیره می شود.
- ❖ با ۱۶ بیت تعداد سطوح چندی سازی برابر $2^{16} = 65536$ است که از مقدار ۳۲۷۶۸ - تا ۳۲۷۶۷ است.



Hamidreza Baradaran Kashani



پیش پردازش

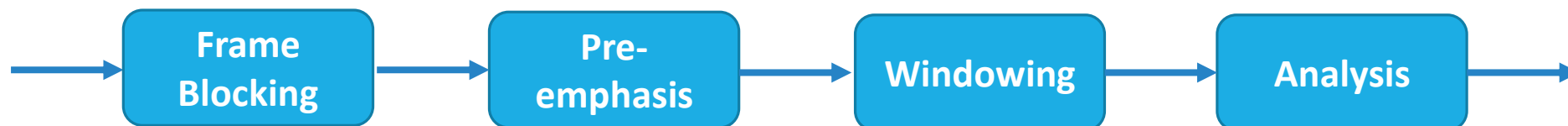
❖ قبل از آنالیز و استخراج ویژگی از گفتار ۳ مرحله پیش پردازشی بر روی سیگنال گفتار انجام می شود.

❖ اساس پیش پردازش گفتار مبتنی بر **آنالیز زمان کوتاه (Short-term Analysis)** است.

❖ فریم بندی (frame blocking)

❖ پیش تاکید (pre-emphasis)

❖ پنجره گذاری (windowing)



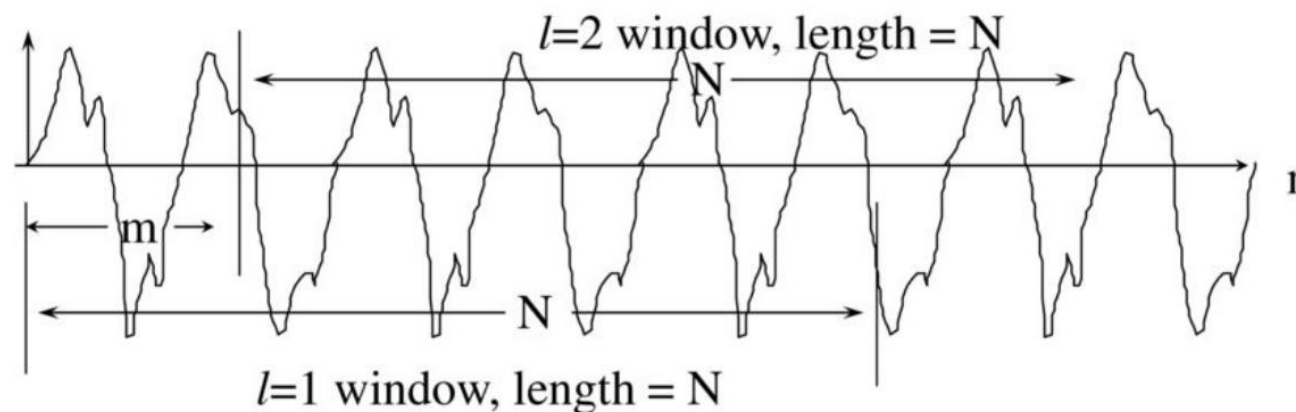
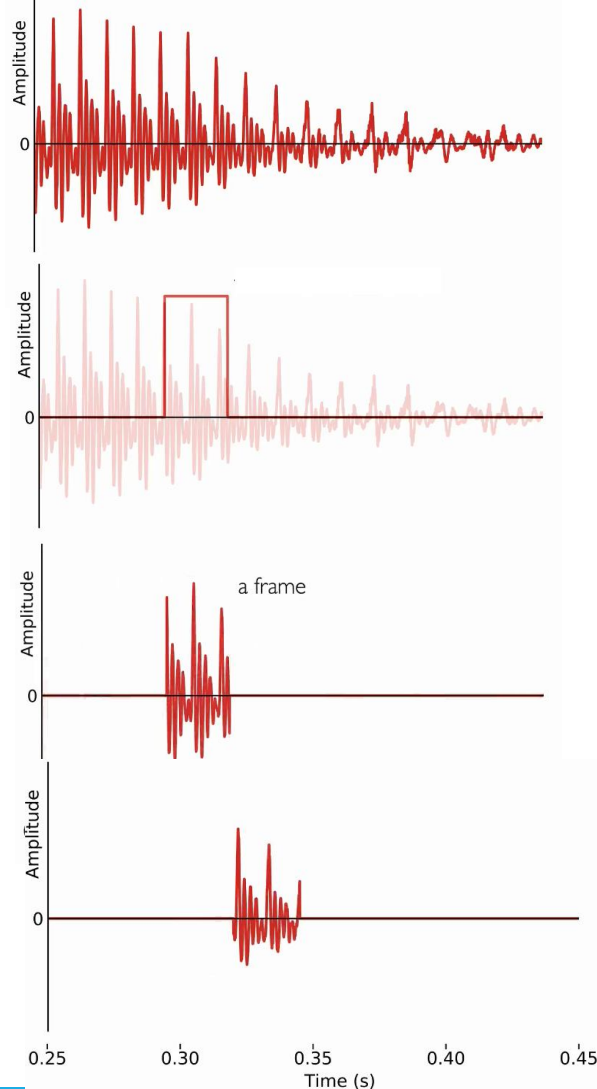


فریم بندی

❖ سیگنال گفتار غیر ایستان است. به همین دلیل گفتار به قطعات با طول بین ۲۰ تا ۳۰ میلی ثانیه بخش بندی می شود.

❖ هر بخش به عنوان یک فریم زمان کوتاه (short-term frame) گفتار مورد آنالیز قرار می گیرد.

❖ در فریم بندی، فریم های مجاور با یکدیگر همپوشانی دارند (معمولا از شیف فریم به اندازه ۱۰ میلی ثانیه استفاده می شود).



Hamidreza Baradaran Kashani

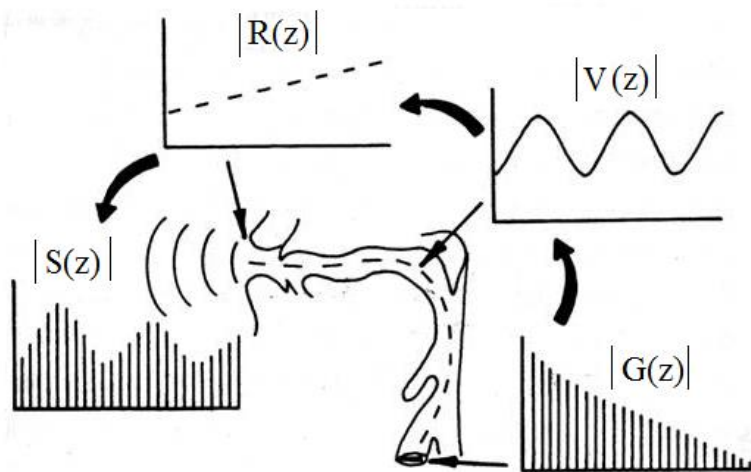
❖ طیف چاکنای ($G(z)$) دارای شیب -12db/oct است.

❖ تابع تشع لبها ($R(z)$) دارای شیب $+6\text{db/oct}$ است که در نتیجه اثر آن روی طیف خروجی چاکنای و مجرای گفتار، طیف نهایی گفتار ($S(z)$) دارای شیب -6db/oct است.

❖ برای رسیدن به تابع اصلی مجرای گفتار از یک فیلتر بالاگذر با عنوان فیلتر پیش تاکید با شیب $+6\text{db/oct}$ بر روی طیف گفتار خروجی از دهان استفاده می شود تا اثر شیب نزولی طیف بطور کامل برطرف شود.

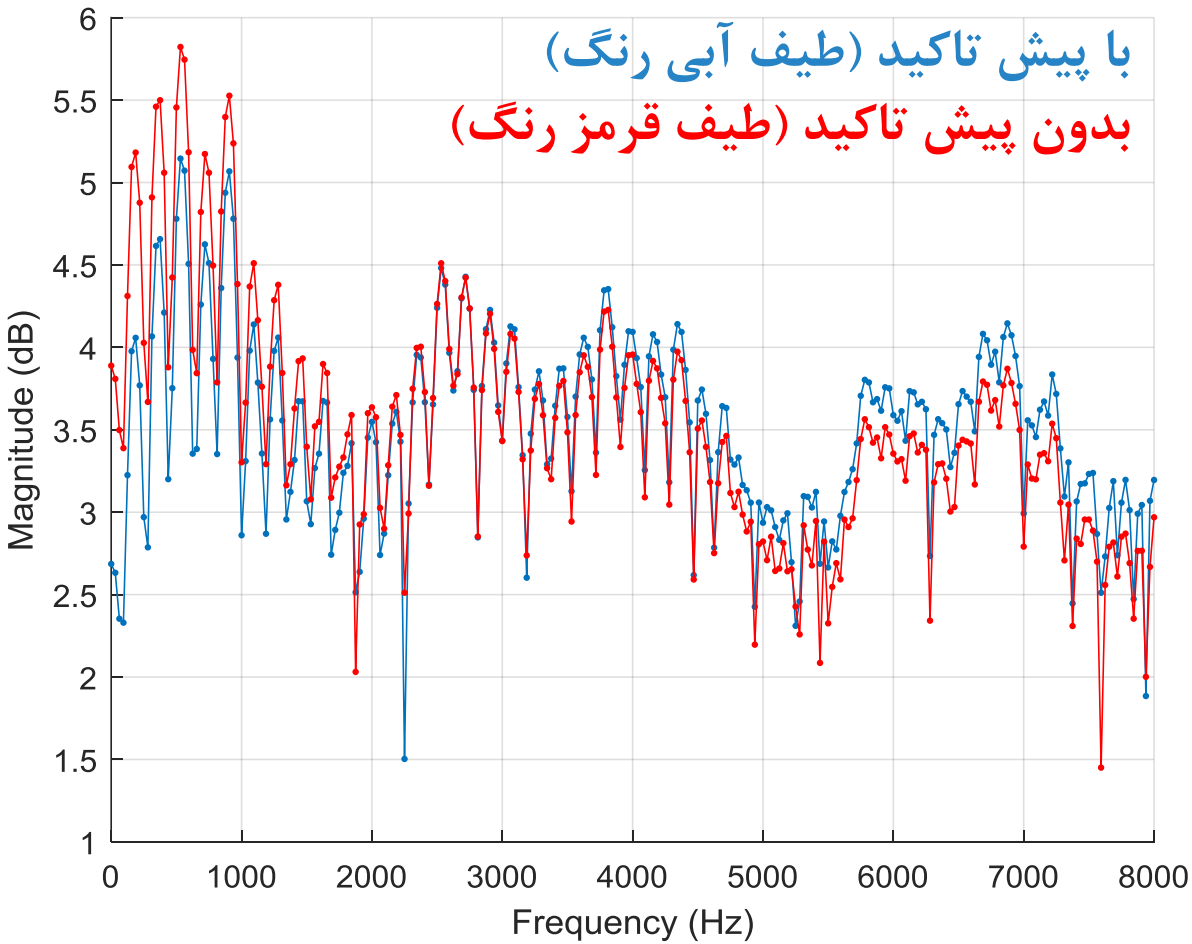
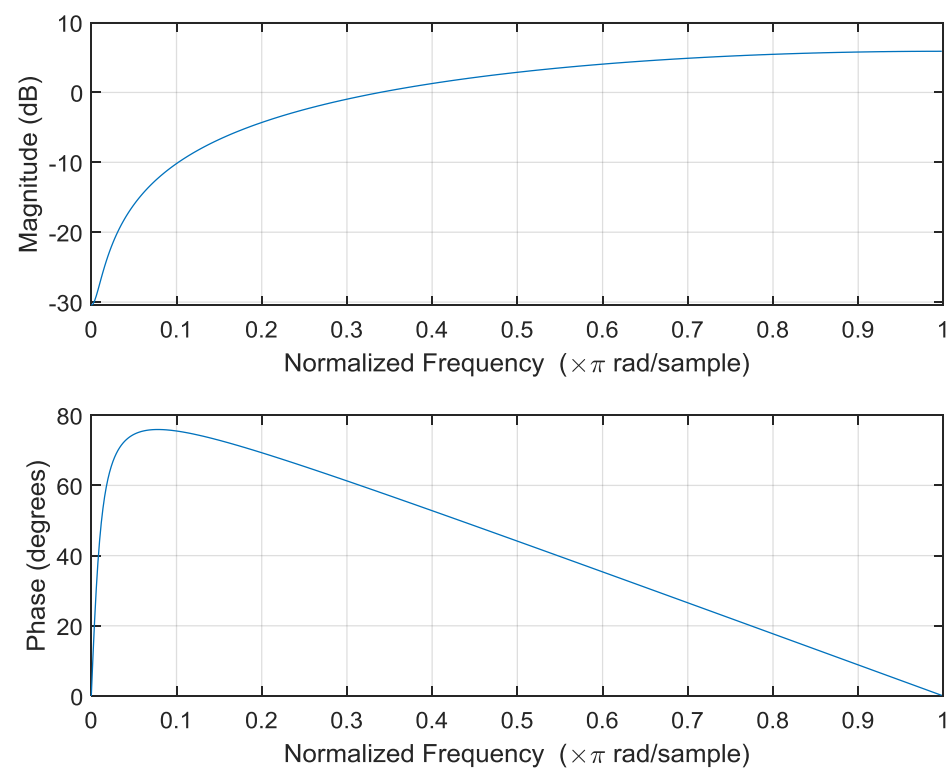
❖ فیلتر پیش تاکید بطور مستقیم بر روی نمونه های سیگنال (مثلا نمونه های یک فریم زمان کوتاه) اعمال می شود.

$$S'[n] = S[n] - a S[n-1], \quad a \in [0,1]$$





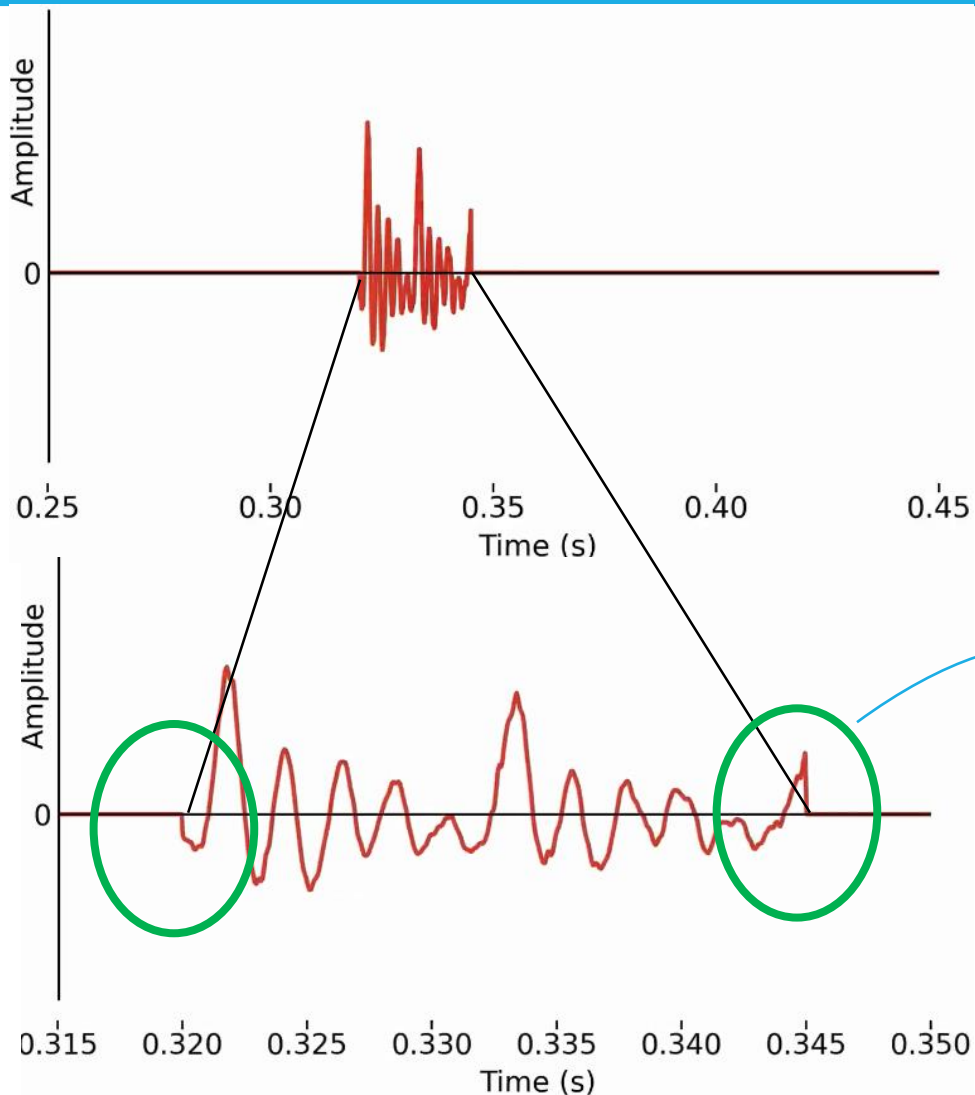
نمایش دامنه و فاز فیلتر پیش تاکید $a = 0.97$



Hamidreza Baradaran Kashani



پنجره گذاری



وجود تغییرات ناگهانی و ناپیوستگی
در ابتدا و انتهای فریم به خاطر
استفاده از پنجره مستطیلی

Hamidreza Baradaran Kashani



پنجره گذاری

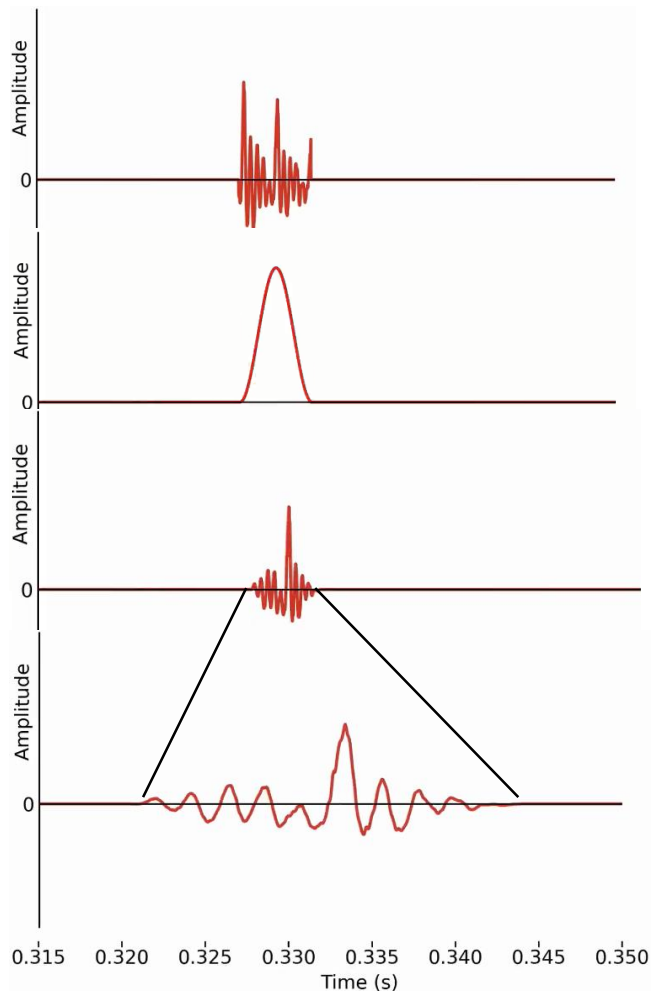
❖ استفاده از پنجره های مخروطی شکل برای کاهش اثر ناپیوستگی ابتدا و انتهای هر فریم

❖ ضرب یک تابع پنجره در نمونه های هر فریم

❖ نتیجه اعمال پنجره در حوزه فرکانس، متناظر با کانولوشن طیف های فریم و پنجره است.

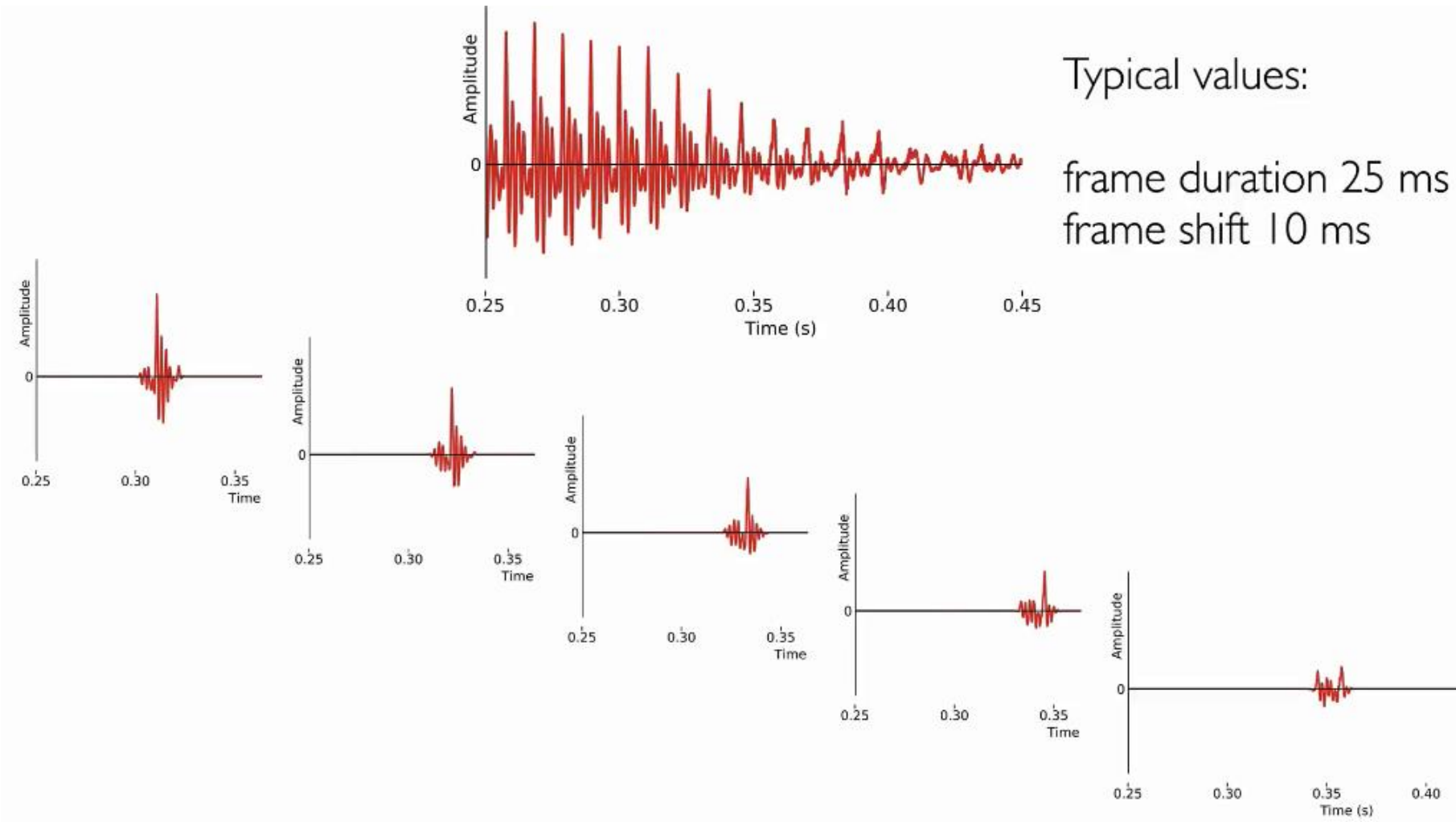
❖ استفاده از پنجره همینگ (Hamming)

$$w[n] = \begin{cases} 0.54 - 0.46 \cos\left(\frac{2\pi n}{N}\right) & 0 \leq n \leq N-1 \\ 0 & \text{otherwise} \end{cases}$$





پنجره گذاری

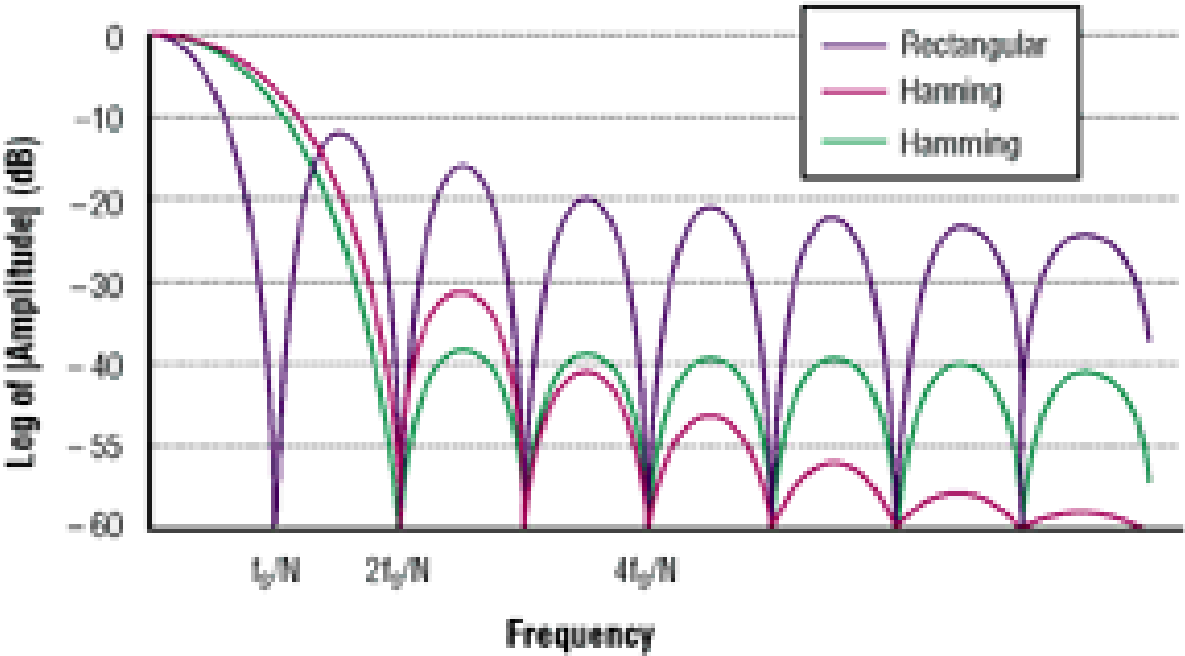


Hamidreza Baradaran Kashani



پنجره گذاری

- ❖ طیف پنجره همینگ در مقابل مستطیلی:
- ❖ پنجره همینگ عرض لوب اصلی بیشتر و دامنه لوب های فرعی کمتر

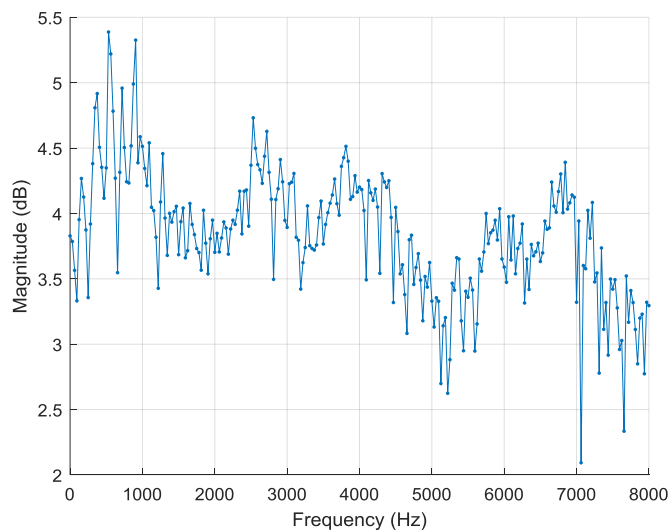
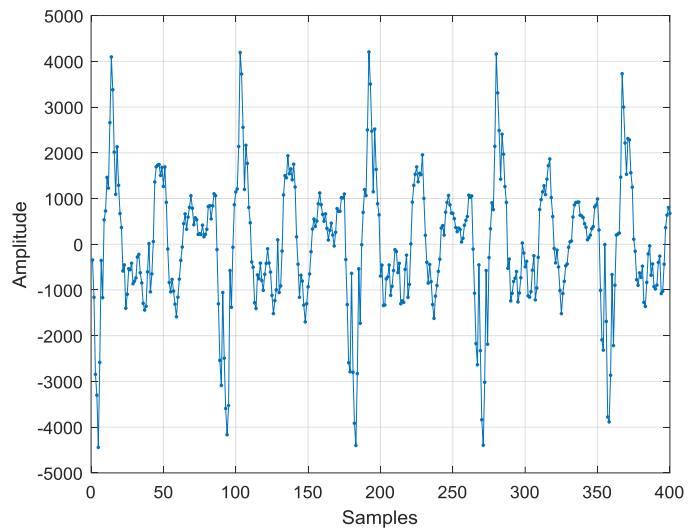


Hamidreza Baradaran Kashani

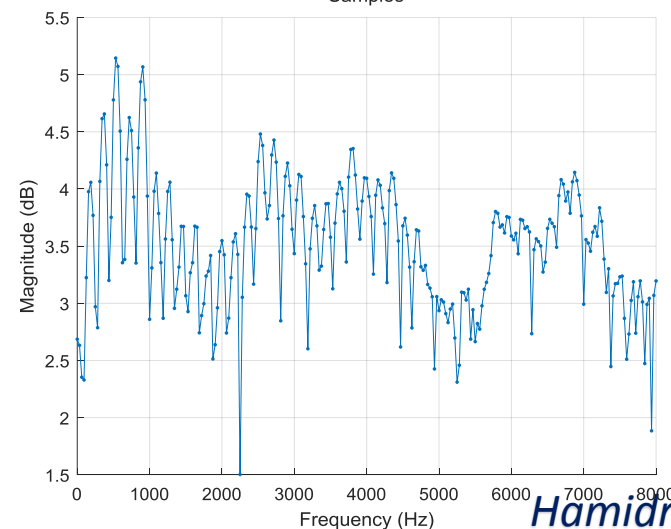
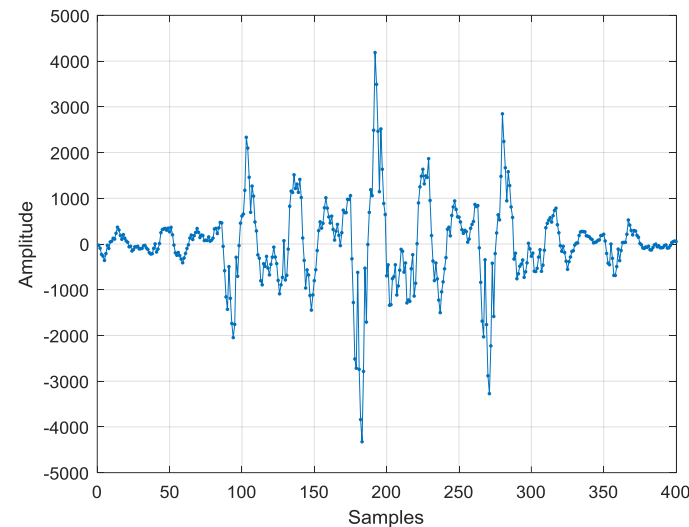


اثر پنجره مستطیلی (چپ) و همینگ (راست) بر روی یک فریم واکنش /A/

پنجره
مستطیلی



پنجره
همینگ

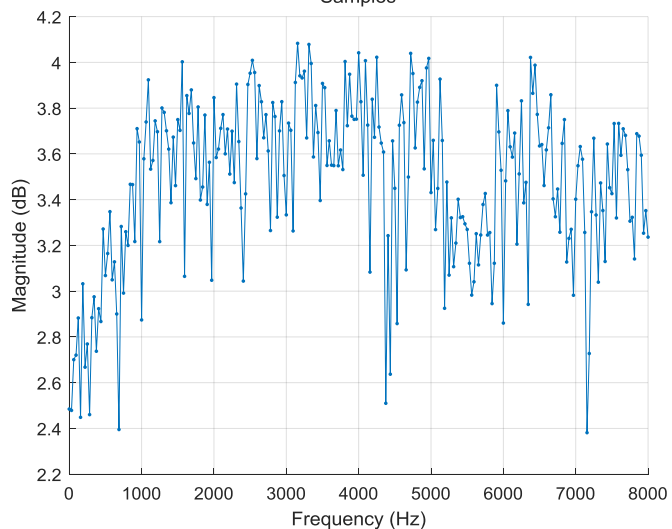
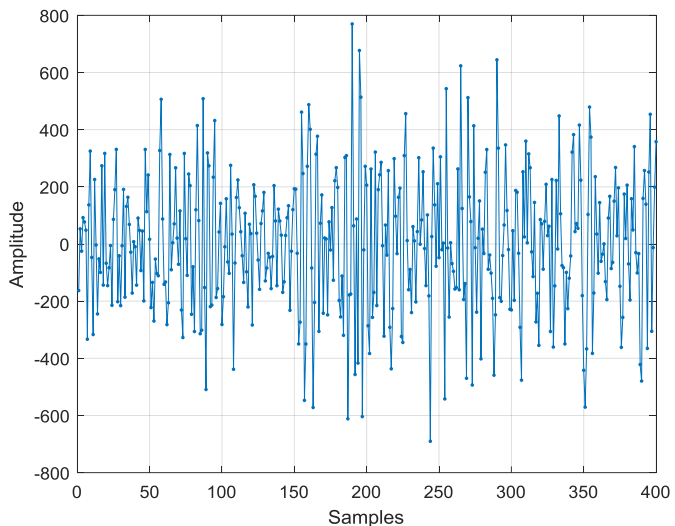


Hamidreza Baradaran Kashani

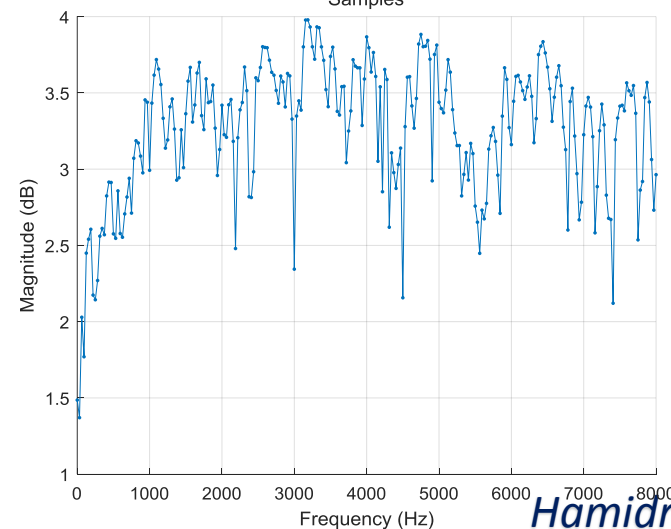
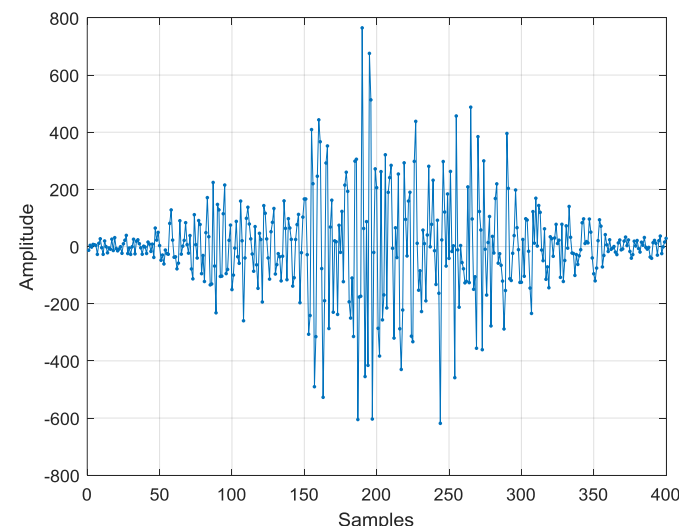


اثر پنجره مستطیلی (چپ) و همینگ (راست) بر روی یک فریم بیواک /p/

پنجره
مستطیلی



پنجره
همینگ



Hamidreza Baradaran Kashani



با تشکر از اساتید و همکاران گرامی:
آقای دکتر همایونیپور
آقای دکتر کبودیان
آقای دکتر ویسی