



دانشگاه زنجان

دانشکده مهندسی

گروه برق

پایان نامه کارشناسی

گرایش:

الکترونیک

عنوان:

نوشتن برنامه ی تحت Matlab برای شناسایی افراد از روی صدای آنها

استاد راهنما:

دکتر فرشاد مریخ بیات

نگارش:

عاطفه شاقلی

شهریور ۹۱

پایان نامه کارشناسی

تقدیم به پدر و مادرم،

که وجودشان برایم همه "مهر" و وجودم برایشان همه "رنج" است.



فهرست مطالب و ضمایم

۱- مقدمه..... ۱

۲- تعاریف مهم..... ۲

۲-۱- سیگنال..... ۲

۲-۲- طبقه بندی سیگنالها..... ۳

۲-۳- پردازش سیگنال..... ۵

۲-۴- سیگنالهای صوتی..... ۷

۲-۵- پردازش سیگنالهای صوتی..... ۸

۲-۶- تشخیص صدا..... ۸

۲-۶-۱- تعریف تشخیص صدا..... ۸

۲-۶-۲- کاربرد تشخیص صدا..... ۸

۲-۶-۳- مزایا و معایب تشخیص صدا..... ۱۰

۳- پیشینه ی مساله..... ۱۱

۳-۱- روش اول..... ۱۱

۳-۲- روش دوم تشخیص صدا توسط مدل مخفی مارکوف..... ۱۳

۴- تجزیه و تحلیل برنامه ی نوشته شده در محیط متلب..... ۱۶

۴-۱- ورود صدا..... ۱۶

۴-۱-۱- مبنای مقدار دهی به فرکانس نمونه برداری..... ۱۷

۴-۱-۲- مشاهدات بدست آمده از تغییر فرکانس نمونه برداری..... ۱۷

۴-۲- جعبه ابزار پردازش سیگنال در متلب..... ۱۸

۴-۳- شرح ایده ی بکار رفته..... ۱۹

۱-۳-۴-مرحله ی اول ۱۹

۲-۳-۴-فیلترهای صدا..... ۲۴

۳-۳-۴-فیلتر متوسط گیر متحرک..... ۲۴

۴-۳-۴-مرحله ی دوم..... ۲۸

۴-۴-حساسیت به بلندی و آرامی صدای ورودی..... ۳۲

۵-۴-ویژگیهای جمله ای که خوانده میشود..... ۳۳

۵-نتیجه گیری..... ۳۵

۶-ضمائم و مراجع..... ۳۴

فصل ۱ - مقدمه

همواره یافتن ویژگی های خاص برای هر شخص، می تواند یک روش برای رمز گذاری یا شناسایی فرد باشد. صدای هر شخص از ویژگی هایی است که خاص خود اوست و به عنوان یکی از این روشها شناخته شده است.

در این پروژه سعی در نوشتن الگوریتمی در محیط مطلب داریم تا به کمک آن گوینده را شناسایی کنیم. در این پایان نامه ابتدا به بیان تعاریفی که در این پروژه حائز اهمیت است می پردازیم و در قسمت بعد شرحی مختصر از چند روشی که قبلاً بکار رفته است را ارائه داده و در آخر به تحلیل الگوریتم نوشته شده، مراحل طی شده و مشاهدات به دست آمده خواهیم پرداخت.

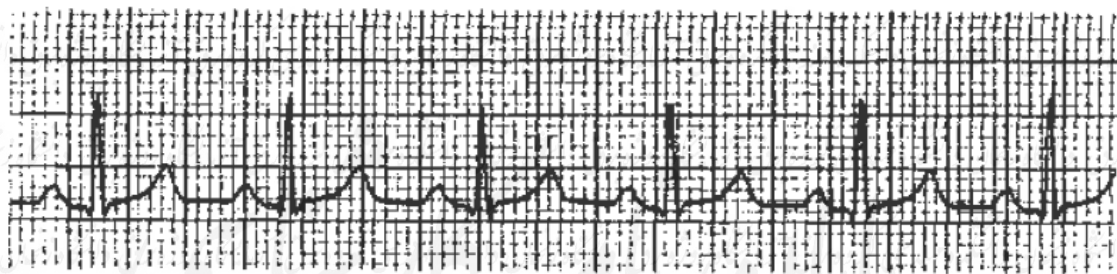
فصل ۲- تعاریف مهم

۲-۱- سیگنال

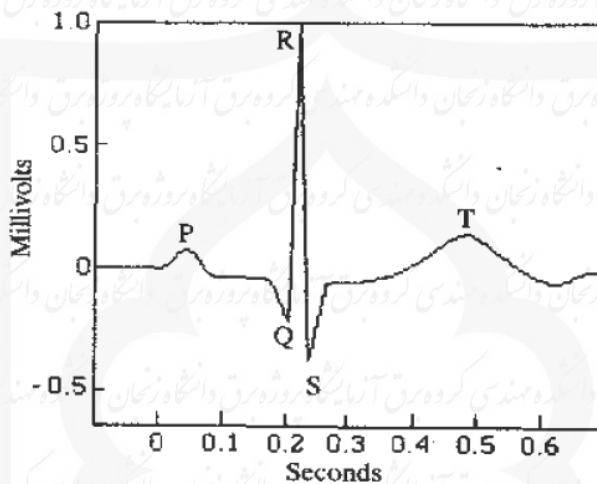
سیگنالها نقش مهمی در زندگی ما ایفا می کنند. نمونه هایی از سیگنالهایی که ما روزانه با آنها ارتباط داریم، سیگنال صوتی از قبیل سیگنال صحبت، موسیقی، تصویر و سیگنالهای ویدئویی هستند. یک سیگنال همواره به عنوان یک تابع که وابسته به تغییرات یک متغیر است، عمل می کند. برای مثال سیگنال موسیقی و یا صحبت به عنوان تابعی از زمان معرفی میشوند، یک تصویر سیاه سفید تابعی از نقاط سیاه و سفید است و یک سیگنال ویدئویی تابعی از تصاویر پشت سرهم است.

یک سیگنال همواره حامل اطلاعاتی مهم است که این اطلاعات توسط پردازش سیگنال، مفید و موثر واقع خواهند شد، چراکه یک پردازشگر، اطلاعات مفید و مطلوب سیگنال را استخراج می کند. روشی که برای پردازش سیگنال انتخاب میشود، وابسته به نوع سیگنال است. برای مثال شکل ۲-۱ یک سیگنال نوار قلب^۱ را نشان میدهد این سیگنال شامل چند نقطه‌ی مهم است که هر کدام می توانند اطلاعات بسیاری را در مورد بیمار به پزشک بدهند. همانطور که مشاهده می‌کنید در شکل ۲-۲ این نقاط مهم با P, Q, R, S, T نشان داده شده اند.

^۱ ECG



شکل ۱-۲ نوار قلب (برگرفته از [۱])



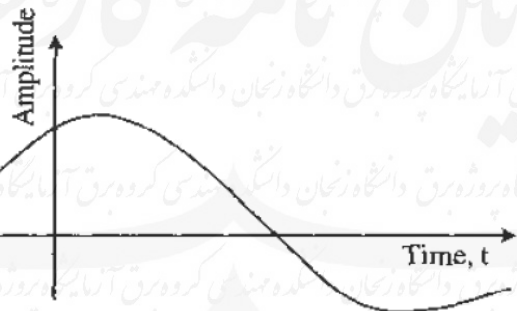
شکل ۲-۲ یک حلقه از نوار قلب (برگرفته از [۱])

۲-۲- طبقه بندی سیگنالها

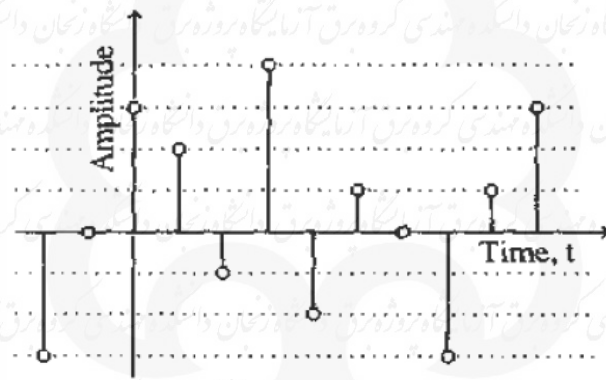
یک سیگنال میتواند توسط یک یا چندین منبع سیگنال، تولید شود. سیگنال تولیدی میتواند یک، دو و یا چند بعدی باشد. برای مثال سیگنال صحبت که تنها تابعی از زمان است، یک بعدی^۲ و سیگنال تصویر که شامل دو متغیر (سیاه-سفید بودن نقاط) است، دو بعدی^۳ است و سیگنال ویدئویی که شامل تصاویر دو بعدی است که در واحد زمان پخش می شوند، سه بعدی^۴ است.

^۲ ۱-D
^۳ ۲-D
^۴ ۳-D

یک سیگنال تک بعدی می‌تواند پیوسته یا گسسته باشد. هرگاه دامنه‌ی سیگنال به صورت پیوسته دارای مقدار باشد، پیوسته در زمان^۵ (شکل ۲-۳) و اگر دامنه سیگنال گسسته باشد، گسسته در زمان^۶ (شکل ۲-۴) نامیده میشود. برای مثال سیگنال صحبت به صورت پیوسته در زمان است و اگر از آن نمونه برداری کنیم گسسته در زمان میشود.



شکل ۲-۳ سیگنال پیوسته در زمان



شکل ۲-۴ سیگنال گسسته در زمان

یک سیگنال پیوسته در زمان که دارای دامنه پیوسته است در واقع یک سیگنال آنالوگ مینماید پس سیگنال صحبت یک سیگنال آنالوگ است و یک سیگنال گسسته در زمان که شامل مقادیر دیجیتالی

^۵ Continuouse-time
^۶ Discrete-time

شده در دامنه است، با عنوان سیگنال دیجیتالی معرفی میشود. برای مثال وقتی از سیگنال صحبت نمونه برداری می کنیم به سیگنال دیجیتالی می‌رسیم.

۲-۳- پردازش سیگنال

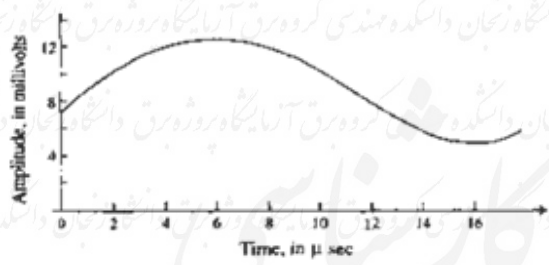
برای پردازش یک سیگنال دیجیتالی از روشهای پردازش سیگنال دیجیتال^۷ استفاده میشود. در حالی که برای پردازش یک سیگنال آنالوگ طبق (شکل ۲-۵) ابتدا بایستی توسط مبدل آنالوگ به دیجیتال^۸ و یا نمونه برداری^۹ سیگنال را به یک سیگنال دیجیتال تبدیل کرده و سپس آن را به روش های DSP پردازش کرد.



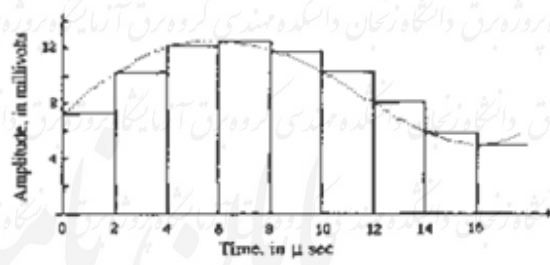
شکل ۲-۵ بلوک پردازش دیجیتالی یک سیگنال آنالوگ (بر گرفته از [۱])

در شکل‌های ۶-۲ الف، ب، ج، د، ه روند پردازش یک سیگنال تک بعدی مثل سیگنال صحبت را مشاهده میکنید. برای استفاده از ابزارهای پردازش دیجیتال، در سیگنال آنالوگ صدا ابتدا بایستی با نمونه برداری آن را به سیگنال گسسته و سپس به دیجیتال تبدیل کرد:

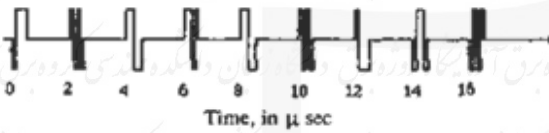
^۷ DSP
^۸ A/D
^۹ sampling



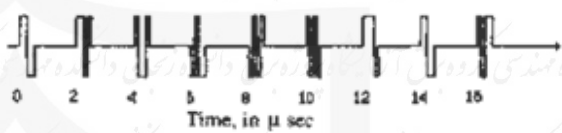
(الف)



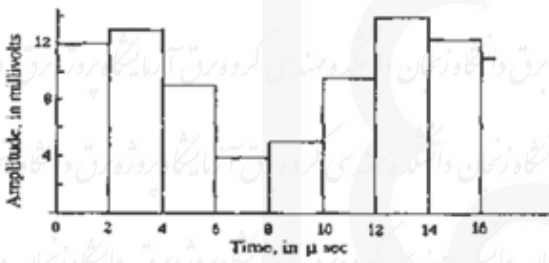
(ب)



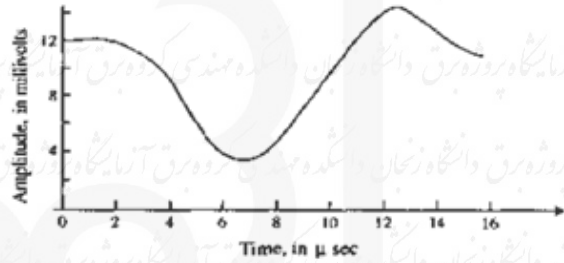
(ج)



(د)



(و)



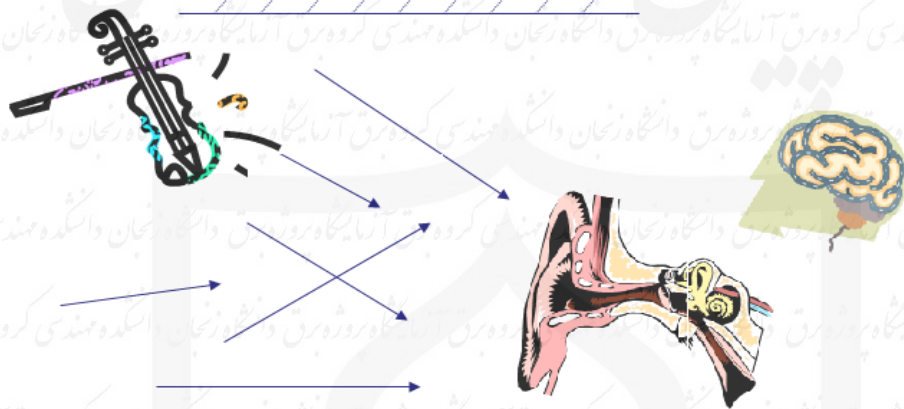
(ه)

شکل ۲-۶ (برگرفته از [۱]): الف. سیگنال آنالوگ ورودی ب. سیگنال نمونه برداری شده S/H ج. خروجی مبدل آنالوگ به دیجیتال (A/D) د. خروجی پردازشگر دیجیتال و خروجی مبدل دیجیتال به آنالوگ

ه. سیگنال آنالوگ خروجی (D/A)

۲-۴- سیگنالهای صوتی

صدا عبارتند از ارتعاش در یک وسیله. برای شنیدن صدا پرده گوش این ارتعاش را از مغز دریافت میکند که با عنوان صدا شناخته شده است (شکل ۲-۷).



شکل ۲-۷

یک صدای دیجیتال، مثل سی دی (CD) صوتی و بسیاری از فرمت (format) های صوتی در آن، شامل نمونه های گسسته از صدا هستند. نسبت نمونه برداری^{۱۰} عبارتند از نسبت مقدار ذخیره شده از صدا بر حسب زمان (به عنوان مثال در CD برابر با ۴۴.۱۰۰ K HZ است).

هرچه نسبت نمونه برداری به میزان کافی بزرگ انتخاب شود کیفیت صدای تولید شده و دقت کار بالاتر خواهد بود در حالی که کاهش این مقدار منجر به کاهش کیفیت صدای تولید شده و از دست رفتن نمونه های بارزش میگردد.

^{۱۰} Samplerate

۲-۵- پردازش سیگنالهای صوتی^{۱۱}

پردازش سیگنالهای صوتی که اغلب با عنوان پردازش صدا شناخته می‌شود، عبارتند از یک تغییر عمده در سیگنال صوت یا صدا برای رسیدن به هدف. در واقع پردازش سیگنال صوتی همان پردازش یک سیگنال آنالوگ تک بعدی است.

۲-۶- تشخیص صدا

۲-۶-۱- تعریف تشخیص صدا

تشخیص صدا و یا شناسایی گوینده^{۱۲} یکی از مسائل علوم رایانه و هوش مصنوعی است که هدف آن شناسایی یک فرد، تنها از روی صدایش است. [۲]

۲-۶-۲- کاربرد تشخیص صدا

بسیاری از اوقات شما با شنیدن صدای یک نفر بدون نیاز به دیدن وی می‌توانید او را بشناسید. هر فردی دارای صدایی است که تقریباً با هیچکس دیگر مشابه نیست. دلیل این امر حفره های صوتی هر شخص و نحوه حرکت دادن دهان در هنگام صحبت کردن است. بنابراین می‌توان سیستمی طراحی کرد که شخص را با استفاده از صدایش تشخیص دهد. چنین سیستمی ممکن است از شما درخواست کند که چند کلمه یا عبارت خاص را بیان کنید تا در آینده از آنها برای رمز عبور استفاده کند یا اینکه یک نمونه‌ی نسبتاً کامل از حرف زدن شما را دریافت کند تا بعدها بدون نیاز به استفاده از کلمه ای خاص شما را شناسایی نماید.

سیستم تشخیص صدا در واقع یک طیف نگاره^{۱۳} از صداست. طیف نگاره یک گراف است که فرکانسهای مختلف صدا را در محور عمودی و زمان را در محور افقی نشان می‌دهد. صداهای مختلف گرافهای متفاوتی می‌سازند. نمونه ای از یک سیگنال صحبت را در شکل ۲-۸ مشاهده می‌کنید که در مدت ۵ ثانیه ضبط شده است. طیف نگاره همچنین از رنگ‌ها با درجات خاکستری برای نشان دادن کیفیت آکوستیکی صدا استفاده می‌کند. نمونه ای از این قبیل طیف در شکل ۲-۹ آورده شده است. [۳]

^{۱۱} Audio signal processing

^{۱۲} Speaker Recognition, speaker identification

^{۱۳} spectrogram

فصل ۵- نتیجه گیری

با وجود تمام سیستمهایی که با دقت بالایی قادر به شناسایی گوینده هستند، بازهم سیستم های تشخیص صدا دارای امنیت بالا نیستند و به همین دلیل اغلب ترجیح داده میشود که در یک سیستم با اهمیت بالا از چند روش مختلف برای تشخیص صدا استفاده شود. برای مثال به علت حساسیتی که سیستم هایی از این دست به جمله ای که خوانده می شود دارند، پس بهتر است از روش تشخیص گفتار (تشخیص کلمات گفته شده) و سپس روش شناسایی گوینده استفاده کرد.

فصل ۶- ضمائیم و مراجع

۶-۱- مراجع

۶-۱-۱- مراجع فارسی

[۵] م. پیام جوادى / پروژه ی کارشناسی / دانشگاه تهران / ۱۳۷۰

[۶] محمدحسن معطر / مدل مخفی مارکوف و الگوریتم های آموزش / دانشگاه صنعتی امیرکبیر /

اردیبهشت ۱۳۸۵

[۳] روزنامه جام جم / هفته نامه کلیک

[۷] www.roshd.ir

۶-۱-۲- مراجع لاتین

[۱] Sanjit k.Mitra, Digital signal processing, second Edition, university of California, santa Barbara ,Publisher: Tata McGraw-Hill,Date:۲۰۰۶

[۴] KevinD. Donohue, Audio Signals and Systems, Electrical and Computer Engineering University of Kentucky

[۲] <http://en.wikipedia.org/wiki/Expectation-Maximization>

[۸] http://www.analog.com/static/imported-files/tech_docs/dsp_book_Ch۱۵.pdf

[۹]

http://www.eas.uccs.edu/wickert/ece۲۶۱۰/lecture_notes/ece۲۶۱۰_chap۶.pdf

۶-۲-۲- ضامم

ضمیمه ۶-۲-۱

```
۱: if (~isempty(daqfind))%empty and stop any sound that are  
current computer
```

```
۲: stop(daqfind)
```

```
۳: end
```

```
۴: ai = analoginput('winsound');
```

```
۵: addchannel(ai, ۱);
```

```
۶: fm=۲۰۰۰
```

```
۷: Fs=۴*fm;
```

```
۸: Ts=۱/Fs;
```

```
۹: ai.SampleRate = Fs;
```

```
۱۰: ai.SamplesPerTrigger = ۴۰۰۰; %۴ seconds of data
```

```
(۴۰۰۰/۱۰۰۰=۴seconds)
```

```
۱۱: secand=ai.SamplesPerTrigger/ai.SampleRate;
```

```
۱۲: ai.TriggerType = 'Immediate';
```

```
۱۳: start(ai)%start collect data from micriphone.
```

```
۱۴: [x,t] = getdata(ai,۴۰۰۰);%(object,number of extract
```

```
sample)
```

ضمیمه ۶-۲-۲

```
% output sound
```

```
۱: ao = analogoutput('winsound');
```

```
۲: addchannel(ao,۱);
```

```
۳: data =x;
```

```
۴: putdata(ao,data)
```

```
۵: start(ao)
```

ضمیمه ۶-۲-۳

```
۱: M=۱۰۰۰;
```

```
۲: for i=۱:size(x,۱)-M+۱
```

```
۳: xf(I,۱)=sum(x(i:i+M-۱))/M;
```

```
۴: end
```

ضمیمه ۶-۲-۴

```
۱: temp=۰;
```

```
۲: for i=۱:size(x,۱)
```

```
۳: temp=temp+sum(abs(x(i))^۲);
```

```
۴: end
```

```
۵: E=temp;
```

```
۶: x=x/E;%normaliz
```