



تست (MRT) Modified Rhyme

- انتخاب CVC های با C های متفاوت
- مثال: Cat, bat, rat, mat, fat, sat

4- ارزیابی انسانی طبیعی بودن

امتیاز متوسط عقیده (Mean Opinion Score MOS)

- MOS بسیار پرکاربرد و بسیار معتبر است
- در این تست کاربر می تواند به دفعات گفتار را بشنود
- نمونه امتیازها را در تصویر 2 مشاهده می کنید.

Score	Speech Quality
1	Not Acceptable
2	Weak
3	Medium
4	Good
5	Excellent

تصویر 2 – نمونه امتیازهای MOS

تست Diagnostic Acceptability Measure (DAM)

این تست بسیار پیچیده است.

در این تست 19 پارامتر مختلف برای امتیازدهی وجود دارد. این پارامترها به 3 گروه اصلی تقسیم می شوند:

- کیفیت سیگنال
- کیفیت پس زمینه
- کیفیت کل

5- ارزیابی کامپیوتری طبیعی بودن

نمی توان از این تست ها برای ارزیابی قابل درک بودن استفاده کرد.

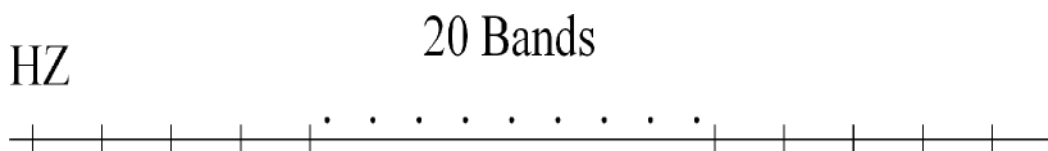
البته استفاده از سیستم های بازشناس گفتار را می توان نوعی ارزیابی برای قابل درک بودن دانست.

تست های کامپیوتری ارائه شده در این درس فقط برای ارزیابی طبیعی بودن گفتار می باشد.

Articulation Index (AI)

AI فرض می کند که باندهای مختلف فرکانسی مستقلند و کیفیت سیگنال را در باندهای مختلف محاسبه می کند.

در هر باند درصد سیگنال شنیده شده توسط شنونده محاسبه می شود (تصویر 3).



تصویر 3 – باندهای مختلف

شرایط قابل شنیدن بودن توسط شنونده:

- بالاتر بودن از سطح آستانه شنوایی انسان
- زیر آستانه درد اسنان
- بیشتر بودن از سطح ماسک کردن نویز

در AI معیار SNR در هر باند محاسبه می شود.

$$AI = \frac{1}{20} \sum_{j=1}^{20} \frac{Min(SNR, 30)}{30}$$

فرمول 1

Signal to Noise Ratio (SNR)

همان طور که از اسم این روش بر می آید (نسبت سیگنال به نویز) انرژی سیگنال به نویز را محاسبه می کند.

$$\begin{aligned} \varepsilon_{(n)} &= s_{(n)} - \hat{s}_{(n)} \\ E_{\varepsilon} &= \sum_{n=-\infty}^{\infty} \varepsilon_{(n)}^2 = \sum_{n=-\infty}^{\infty} [s_{(n)} - \hat{s}_{(n)}]^2 \quad E_s = \sum_{n=-\infty}^{\infty} s_{(n)}^2 \\ SNR_{(global)} &= 10 \log \frac{E_s}{E_{\varepsilon}} = 10 \log \frac{\sum_{n=-\infty}^{\infty} s_{(n)}^2}{\sum_{n=-\infty}^{\infty} [s_{(n)} - \hat{s}_{(n)}]^2} \end{aligned}$$

فرمول 2

Segmental SNR

به این صورت است که سیگنال را به فریم هایی تقسیم می کند و SNR را بر روی آن انجام می دهد و میانگین گیری می کند.

$$SNR_{(seg)} = \frac{1}{M} \sum_{j=0}^{M-1} 10 \log \left[\frac{\sum_{n=m_j-N+1}^{m_j} s_{(n)}^2}{\sum_{n=m_j-N+1}^{m_j} [s_{(n)} - \hat{s}_{(n)}]^2} \right]$$

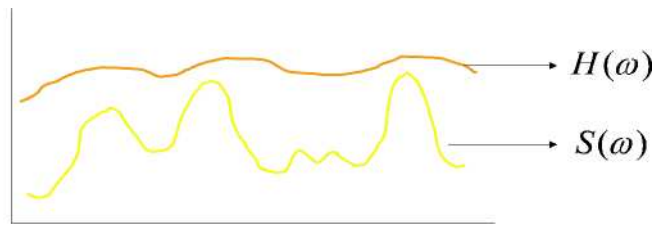
Frequency Weighted Segmental SNR

به هر باند فرکانسی وزنی داده می شود (مثلاً بر اساس مقیاس مل)

$$SNR_{(fw-seg)} = \frac{1}{M} \sum_{j=0}^{M-1} 10 \log \left[\frac{\sum_{k=1}^K W_{j,k} [E_{s,k}(m_j) / E_{\varepsilon,k}(m_j)]}{\sum_{k=1}^K W_{j,k}} \right]$$

فرمول 4

Itakura



$H(\omega)$ Is the envelope spectrum

$$S(\omega) = F\{R(\tau)\} \Rightarrow S(\omega) = |X(\omega)|^2$$

Use from All-Pole (AR) Model

$$H(\omega) = \frac{1}{1 - \sum_{i=1}^p a_i e^{-j\omega}}$$

تصویر 4 - معیار itakura

معیار itakura به صورت زیر است.

$$d(g_s(m), g_{\hat{s}}(m)) = \sqrt{\frac{1}{M} \sum_{l=1}^M [g_s(l, m) - g_{\hat{s}}(l, m)]^2}$$

فرمول 5

6 - خلاصه و نتیجه گیری

در این فصل بحث ارزیابی کیفیت را بیان کردیم.

7 - منابع درس

- 1- Rabiner, "Fundamentals of Speech Recognition"
- 2- Huang, Acero, "Spoken Language Processing"
- 3- Deller, "Discrete-time processing of speech signals"