Noise types

- White noise
- Pink noise
- Speech-shaped noise

Long-term spectrum of speech

Long-term average vowel spectrum (Assmann and Katz 2000)
Audibility of speech

Speech recognition accuracy

- Word recognition scores measured in background noise can improve from near chance to near perfect over a range of signal-to-noise ratios (SNRs) as narrow as 10 dB (French and Steinberg, 1947).

Speech reception threshold

- Problem: comparisons between normal and hearing impaired individuals often lead to floor or ceiling performance in different conditions.

Speech recognition in noise

- Speech reception threshold, SRT (Plomp & Mimpen, 1969)
  - Speech-to-noise ratio required to achieve a specific level of intelligibility, typically 50%
  - Effects of speech materials
  - Effects of type of masker (e.g., speech-shaped noise vs. a single competing talker)
  - Effects of spatial separation of target & masker

Sensorineural hearing loss

- Listeners with cochlear hearing loss have difficulty recognizing speech when background noise is present.
  - Reduced audibility
  - Supra-threshold “distortions”
    - Impaired frequency selectivity
    - Loudness recruitment

Speech reception threshold

- One solution is to use an adaptive-tracking procedure to estimate the SNR required for a listener to achieve 50% correct.
- This procedure is referred to as the speech-reception threshold (SRT) (Plomp and Mimpen, 1979).
Speech recognition in noise

<table>
<thead>
<tr>
<th>Masker type</th>
<th>Listening situation</th>
<th>Deficit in SRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech-shaped noise</td>
<td>Speech+masker in front, unaided</td>
<td>2.5 - 7.0 dB</td>
</tr>
<tr>
<td></td>
<td>Speech+masker in front, aided</td>
<td>2.5 - 6.0 dB</td>
</tr>
<tr>
<td>Single talker</td>
<td>Speech+masker in front, unaided</td>
<td>6.0 - 12.0 dB</td>
</tr>
<tr>
<td>Single talker</td>
<td>Speech+masker in front, aided</td>
<td>4.0 - 10.0 dB</td>
</tr>
<tr>
<td></td>
<td>Speech+masker in front, spatially</td>
<td>12.0 – 19.0 dB</td>
</tr>
<tr>
<td></td>
<td>separated</td>
<td></td>
</tr>
</tbody>
</table>

Source: Moore, BCJ (2003) Speech Communication

Articulation Index

• How much does audibility contribute to difficulty understanding speech in noise?
• Articulation Index (AI) estimates the contribution of audibility (and other factors) to speech intelligibility

Articulation Index

1. Divides the speech and masker spectrum into a small number of frequency bands
2. Estimates the audibility of speech in each band, weighted by its relative importance for intelligibility
3. Derives overall intelligibility by summing the contributions of each band.

Articulation Index

• Most studies show that speech intelligibility is worse than predicted by the AI for hearing-impaired listeners, especially for moderate or severe hearing loss.

Articulation Index

• Conclusion: factors other than audibility must be responsible for the difficulties experienced by hearing-impaired listeners understanding speech in noise.
• What else?
  • Frequency selectivity
  • Temporal resolution

Frequency Selectivity

• Frequency selectivity is the ability to resolve the spectral components of complex sounds.
• Reduced frequency selectivity may lead to difficulty in understanding speech in noise.
Auditory filters

- Fletcher (1940) suggested that the peripheral auditory system could be modeled as a bank of linear bandpass filters with continuously overlapping center frequencies.

![Auditory filters](image)

Mapping frequency to cochlear position


![Mapping frequency to cochlear position](image)

Critical Bandwidth

- Fletcher (1940) band-widening experiment
  - The threshold for detecting a pure tone in the presence of a bandpass noise masker increases as the noise bandwidth increases, until the width of the band exceeds the critical bandwidth of the auditory filter.

![Critical Bandwidth](image)

Critical Bandwidth

- Sources of evidence for critical bandwidth:
  - Band-widening experiments (Fletcher, 1940)
  - Loudness summation (Zwicker et al., 1957)
  - Two-tone masking (Zwicker, 1954)
  - Discrimination of partials within complex tones (Plomp and Mimpen, 1968)

![Critical Bandwidth](image)
Critical Bandwidth

- Fletcher (1940) made the simplifying assumption that the auditory filter could be modeled as a rectangle, with flat top and vertical slopes.

Power spectrum model of masking

- Fletcher suggested that only a narrow band of frequencies in the region of the tone contribute to masking.
- He called this the critical bandwidth (CB).

Power spectrum model of masking

- But threshold changes gradually as the noise bandwidth increases, suggesting auditory filters with sloping rather than rectangular skirts (Patterson, 1976).

Power spectrum model of masking

- Detection of probe tone in the presence of a noise masker depends on the relative power of probe and noise passed by the auditory filter centered on the tone (Patterson, 1976).

Frequency selectivity

- Patterson (1976) estimated auditory filter shapes from the function relating tone threshold to notch width.
- The derived filters have a rounded top and steep skirts, with bandwidths 10-15% of filter center frequency.

Notched noise method

- Derived auditory filter shape
Auditory filter shapes as a function of frequency

Auditory filter shapes as a function of level

Excitation patterns

Excitation patterns

Excitation patterns

ERB-rate scale

- Moore and Glasberg (1989) auditory model
- ERB: Equivalent Rectangular Bandwidth
- ERB units provide approximately equal distances along the basilar membrane

\[ E = 16.7 \log_{10} (1 + \frac{f}{165.4}) \]
and

\[ f = 165.4(10^{0.06E} - 1) \]
Excitation pattern: ERB scale

Simulation of reduced frequency selectivity

Effects of reduced frequency selectivity on vowel /æ/

Distortion of spectral shape
- Broader auditory filters produce a “smeared” excitation pattern: reduced prominence of peaks, smaller peak-to-valley ratios.
- Introduction of noise fills up the valleys between the spectral peaks and reduces the distinctiveness of the spectral profile.

Simulation studies
- Simulation of reduced frequency selectivity (spectral smearing of the short-term speech spectrum) results in lowered intelligibility for listeners with normal hearing, particularly in noise (ter Keurs et al., 1993; Baer & Moore, 1994)

Distortion of temporal structure
- Broader auditory filters alter the temporal fine structure of the output.
  - Increased contribution of adjacent components
  - Increase in within-channel modulation
  - Diminished differences between adjacent channels
Loudness Recruitment

- When a sound is increased in level above absolute threshold, the rate of growth of loudness is greater than normal.
- At levels >90-100 dB SPL, loudness returns to normal (sound appears equally loud to hearing-impaired and normal listeners).

Fluctuating masker benefit

- Normal hearing listeners benefit from fluctuations in a noise masker. Compared to steady-state masker, fluctuating maskers can lead to ~6 dB release from masking.
- The advantage has been attributed to a process by which speech information is extracted during brief dips in the level of the fluctuating masker (dip listening).
- Hearing-impaired listeners (including cochlear implant users) do not benefit (or benefit very little) from masker fluctuations.

Temporal Modulation

Structure of Speech

Effects of reduced frequency selectivity on temporal structure

Loudness Recruitment

- Loudness recruitment is associated with reduced dynamic range (range between absolute threshold and highest comfortable level).
- Recruitment may reduce the ability to “listen in the dips” in a fluctuating masker, such as a competing voice.
- Recruitment distorts loudness relationships among components of speech sounds.
Temporal structure of speech

• Rosen (1992) proposed that the temporal structure of speech can be partitioned into three levels based on their rate of modulation:
  – **Envelope cues** - slow modulations (<50 Hz) associated with syllable structure
  – **Periodicity cues** (70-500 Hz) correspond to the rate of vocal fold vibration (voice pitch)
  – **Fine-structure cues** (> 250 Hz) correspond to rapid modulations associated with formant changes

Modulation spectrum of speech

• Houtgast and Steeneken (1985) showed that the intelligibility reduction caused by noise and reverberation can be modeled in terms of the corresponding reduction in **temporal envelope modulations**.

Modulation spectrum of speech

• Houtgast and Steeneken proposed a measure called the **Speech Transmission Index** (STI), based on the estimates of the amount of modulation preserved in different frequency bands. The STI is designed to predict the overall intelligibility of distorted speech.

Noise and reverberation tend to fill the dips in the temporal envelope of speech

Noise and reverberation tend to flatten formant transitions and fill gaps between them
Modulation spectrum of speech

- The capacity of a communication channel to transmit modulations in the energy envelope is referred to as the temporal modulation transfer function (MTF).
- The MTF for speech has a low-pass shape with a peak around 4-6 Hz, reflecting the syllable alternation rate in connected speech.

Signal processing to obtain the MTF

1. Filter the speech signal in octave bands between 0.25 and 8 kHz
2. Square and low-pass filter the output (30 Hz)
3. Analyze the resulting intensity envelope using 1/3 octave bandpass filters with center frequencies between 0.63 and 12.5 Hz.
Channel Vocoder

- Dudley (1939) developed the channel vocoder, a speech analysis-synthesis system that exploits the modulation structure of speech.
- Vcoders belong to a class of speech analysis/synthesis systems that perform a source-filter decomposition of the signal.

- In each channel, the amplitude envelope is extracted from the filtered waveforms.
  - A sequence of pulses is generated at the frequency of the fundamental for voiced sounds; or white noise if the signal is unvoiced.
  - The envelope is modulated by the pulsed and/or noise source, and summed across channels.

Channel Vocoder

- The channel vocoder filters the speech signal through a bank of bandpass filters with center frequencies distributed across the speech range.

Channel Vocoder

1. Filter speech with a set of bandpass filters.
2. Extract the waveform envelope in each channel.
3. Obtain the excitation signal (pulsed or noise) from the broadband signal.
4. Modulate the excitation signal with the filtered waveform envelope in each channel, then re-filter the result through the same bandpass filter.
5. Sum up the bands and scale to the appropriate amplitude.

Cochlear Implants
Cochlear Implants

- Cochlear implants provide reduced spectral activation – up to 24-channel electrode array replaces large array of mechanical-to-neural transducers (5 k inner hair cells).
- Imperfect electrode penetrations cause inappropriate activation of tonotopic map.
- Current spread to adjacent regions leads to spectral smearing (masking).
- Impedance mismatch can lead to imperfect amplitude mapping and inappropriate loudness growth.

Cochlear Implants and speech perception

- In quiet, speech recognition by cochlear implant users is fairly successful (70-80% words correct).
- Implication: Loss of spectral resolution (frequency selectivity) has less impact and fewer consequences for speech perception than predicted, at least in quiet, suggesting that temporal processing and gross spectral cues may be more important in speech perception than previously believed.

Cochlear Implants and speech perception

- However, cochlear implant users continue to experience difficulty in noise, and require much higher signal-to-noise ratios than normal hearing listeners to achieve similar levels of accuracy.