THE SPEECH

Study aid for learning of Communications Acoustics

VIHIM 000

Prof. Fülöp Augusztinovicz
BME Dept. of Networked Systems and Services
fulop@hit.bme.hu

2018. október 16., Budapest
Introduction

- The language:
  - Communication between the individuals of the human society
  - Main tool of the human thinking

- The speech:
  - Primary manifestation of the language
  - Most widespread form of human communication (but by far not the only one)
  - The very first, from the chronological point of view: well before writing
Suggested literature

- P. Ladefoged & I. Maddieson: *The sounds of the world’s languages*
  Blackwell Publishers, 1985


- Wendy J. Holmes: *Speech synthesis and recognition*. 2nd Ed.

  Springer, 2004
The natural chain of speech

Brain level

Organic level

Acoustical level

Organic level

Brain level
The structure of speech

- Sentence > word > syllable > **phonema** > speech sound

- Notion of phonemae:

The smallest contrastive linguistic unit which may bring about a change of meaning
The speaking organ / 2

- Larynx and the vocal cords

*2.4. ábra*
A gégező anatómiai felépítése
1 gyűrűsporc, 2 pájzzopor, 3-4 kannaporcok, 5 gégefedő porc, 6-7 hangszalagok, 8-9 állhangszalagok, 10-11 Morgani-féle üregek, 12 (nyitott) hangrés
The human singing

- Laryngoscope:

The speech production process / 1

- release of air from the lung
- the air at the larynx either
  - passes through: **unvoiced** sound (consonants like s, t or f)
  - Is set into vibration by the vocal cords: **voiced**
    (vowels or voiced consonants like m or z)
- operation of ports (or valves) -> production of various sounds:
  - if velar port is open: nasal consonants or vowels
    (m, n, of franc)
  - if linguo-palatal port is closed: k, g or ng
  - in the linguo-alveolar port: mostly consonants
    e.g. explosives (t, d) and fricatives (s, z, ʃ, ʒ)
  - in labial port: p, b, t, v
Example: articulation of voiceless plosives
The speech production process / 2

- The transmission system
  - system of resonators (aka vocal tract, articulation channel)
- Excitations
  - voicing
  - turbulence
  - shock wave
Excitation mechanisms

voice

shock wave

turbulence

\[ T = \begin{cases} 8 \ldots 12 \text{ ms} & \text{förajtnál} \\ 4 \ldots 8 \text{ ms} & \text{nőknál} \end{cases} \]

\[ T_1 = (0.2 \ldots 0.4) T \]

\[ T_2 = (0.3 \ldots 0.7) T \quad T_1 < T_2 \]
Classification of speech sounds

- **Vowels (V)**: ɪ, é, ü, ö, e á, a, o, u
- **Consonants (C)**
  - stops
    - nasals
    - plosives
      - voiced
      - voiceless
    - rhotic / trill
  - lateral fricatives
  - fricatives
    - voiced
    - voiceless
  - stop-fricatives (affricate)
    - voiced
    - voiceless

  - m, n, ny, ng
  - b, d, g, gy
  - p, t, k, ty
  - r
  - l, j
  - v, z, zs
  - f, s, sz, h
  - dz, dzs
  - c, cs
The international phonetic alphabet

```
| CONSONANTS (PULMONIC)                                                                 |
|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| Bilabial                        | Labiodental                      | Dental                          | Alveolar                        | Postalveolar                    | Retroflex                       | Palatal                        | Velar                           | Uvular                          | Pharyngeal                      | Glottophonal                    | ---------------------------------| 2005 IPA                        |                                  |
| Plosive                         | p                                | b                               | t                               | d                               | t                               | d                             | c                               | j                               | k                               | g                               | q                               | g                               | ?                               |
| Nasal                           | m                                | m                               | n                               | n                               | n                               | j                             | j                               | j                               | j                               | j                               | n                               |                                  |                                  |
| Trill                           | B                                |                                 | r                               |                                 |                                 |                               |                                 |                                 |                                 |                                 |                                 |                                 | R                               |                                  |
| Tap or Flap                     | v                                |                                 | r                               |                                 |                                 |                               |                                 |                                 |                                 |                                 |                                 |                                 |                                  |                                  |
| Fricative                       | f                                | v                               | θ                               | Ø                               | Ø                               | s                             | z                               | s                               | z                               | s                               | z                               | j                               | x                               | y                               | χ                               | θ                               | θ                               | θ                               | θ                               | θ                               | θ                               | θ                               | θ                               | θ                               | θ                               |
| Lateral                         |                                   |                                 |                                 |                                 |                                 |                               |                                 |                                 |                                 |                                 |                                 | L                               |                                 |                                 |
| Approximant                      | u                                |                                 |                                 |                                 |                                 |                               |                                 |                                 |                                 |                                 |                                 |                                 |                                 |                                 |
| Lateral approximant             | l                                |                                 |                                 |                                 |                                 |                               |                                 |                                 |                                 |                                 |                                 |                                 | L                               |                                 |

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.
```

http://www.internationalphoneticalphabet.org/ipa-sounds/ipa-chart-with-sounds/
Characteristics of vowels

FRF of the vocal tract

Line spectrum of the voice, female

Line spectrum of the voice, male

Final spectrum, modified by the transmission of the vocal tract
The resonant system

- The **nasal** and the **oral** cavities
Formants of vowels
The oral cavity as a tuned resonator

- Tuning element: the tongue
Characteristics of consonants

- **Fricatives**
  - turbulent excitation > resonance of the vocal tract

- **Plosives**
  - cannot be held for long time
  - no structure of formants
  - shock wave > time out of vocal cords > stabilised vowel
Articulation places of consonants

<table>
<thead>
<tr>
<th>Palatal</th>
<th>Velar</th>
<th>Uvular</th>
<th>Pharyngeal</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>c</td>
<td>j</td>
<td>k</td>
<td>q</td>
<td>G</td>
</tr>
<tr>
<td>j</td>
<td>1j</td>
<td>N</td>
<td></td>
<td>R</td>
</tr>
</tbody>
</table>
Database of Hungarian speech sounds

Music and singing: form of joy

https://www.youtube.com/watch?v=5hksKXZQ3us

https://www.youtube.com/watch?v=Tz7RQ-C8muk
THE SPEECH INTELLIGIBILITY
Influencing factors of intelligibility

1. articulation capability of the speaker
2. listening capability of the listener
3. characteristics of the sound field (room acoustical parameters)
4. level of background noise

Parameter 3 and 4 are expressed in form of ratios:

3. direct to reverberant ratio

and

4. signal to noise ratio
Measurement of intelligibility

- **Subjective tests**
  - test with meaningless syllables
    - invoking many test persons
    - result is the ratio of correctly understood syllables to the total number of syllables, in %
  - test with full sentences

- **Objective tests**
  - articulation index (AI): only background noise is considered
  - Speech Intelligibility Index (SII): reverberation is also taken into account
  - Speech Transmission Index (STI): subjective quantity is approached by objective measurements (international standard)
    - the speech consists of components of various frequencies, which are amplitude modulated by a low frequency envelope
    - in order to remain intelligible, this modulation should be retained
    - the transmission of this modulation without distortion is deteriorated both by multiple-path transmission (reverberation) and by background noise
Effect of noise: Articulation index (AI)

\[ AI = \sum g \cdot \Delta L, \]

where \( g \) is the weight of the octave band, \( \Delta L \) is the difference in loudness between the speech and the noise. The signal-to-noise ratio should be as high as possible, but at least 5-10 dB.

<table>
<thead>
<tr>
<th>f_{közép}</th>
<th>250</th>
<th>500</th>
<th>1000</th>
<th>2000</th>
<th>4000</th>
</tr>
</thead>
<tbody>
<tr>
<td>g</td>
<td>0.0018</td>
<td>0.0050</td>
<td>0.0075</td>
<td>0.0107</td>
<td>0.0083</td>
</tr>
</tbody>
</table>

A beszéd érthetősége és az AI artikulációs index közötti összefüggés a következő:

<table>
<thead>
<tr>
<th>AI</th>
<th>Beszédérthetőség</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>igen rossz</td>
</tr>
<tr>
<td>0.3</td>
<td>nem megfelelő</td>
</tr>
<tr>
<td>0.3 ÷ 0.5</td>
<td>megfelelő</td>
</tr>
<tr>
<td>0.5 ÷ 0.7</td>
<td>jó</td>
</tr>
<tr>
<td>&gt; 0.7</td>
<td>igen jó</td>
</tr>
</tbody>
</table>

Replacement rule of thumb: the signal to ratio should be as high as possible, but min. 5-10 dBA.
Effect of reverberation on the intelligibility

- Original sound sample
- In-situ recording
- Measured impulse response
- Artificially reverberated sound
Arificial reverberation / 1

Cave theatre in the Fertőrákos quarrier
Arbitrary reverberation / 2

Underground concert hall in the Baradla cave in NW-Hungary
Arficial reverberation / 3

Esztergom Basilica
Consequence: difficulties with PA systems

- Very difficult to develop/install good quality PA systems in reverberant spaces
- Example: the aula of this university (building K)
Concept of the calculation of STI

- Test signal: modulated band-limited noise
  - nowadays: by means of impulse response measurement

- Calculation of apparent S2N ratio:

- Frequency weighting:

\[ w_k = 0.13 \quad 0.14 \quad 0.11 \quad 0.12 \quad 0.19 \quad 0.17 \quad 0.14 \]
Evaluation of STI values

<table>
<thead>
<tr>
<th>STI value</th>
<th>Quality according to IEC 60268-16</th>
<th>Intelligibility of syllables in %</th>
<th>Intelligibility of words in %</th>
<th>Intelligibility of sentences in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 – 0.3</td>
<td>bad</td>
<td>0 – 34</td>
<td>0 – 67</td>
<td>0 – 89</td>
</tr>
<tr>
<td>0.3 – 0.45</td>
<td>poor</td>
<td>34 – 48</td>
<td>67 – 78</td>
<td>89 – 92</td>
</tr>
<tr>
<td>0.45 – 0.6</td>
<td>fair</td>
<td>48 – 67</td>
<td>78 – 87</td>
<td>92 – 95</td>
</tr>
<tr>
<td>0.6 – 0.75</td>
<td>good</td>
<td>67 – 90</td>
<td>87 – 94</td>
<td>95 – 96</td>
</tr>
<tr>
<td>0.75 – 1</td>
<td>excellent</td>
<td>90 – 96</td>
<td>94 – 96</td>
<td>96 – 100</td>
</tr>
</tbody>
</table>

Many figures of merit of intelligibility can be compared by using the **CIS** (Combined Intelligibility Scale) value:

\[
\text{CIS} = 1 + \log_{10}(\text{STI}).
\]
Characteristics of the STI

- It depends on
  - signal to noise ratio
  - reverberation
  - bandwidth
  - nonlinear distortions
  - etc.

<table>
<thead>
<tr>
<th>Sávszélesség</th>
<th>CIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>350-3000 Hz</td>
<td>0,62</td>
</tr>
<tr>
<td>350-5700 Hz</td>
<td>0,77</td>
</tr>
<tr>
<td>180-3000 Hz (nagyjából ez felel meg a végpont elektroakusztikai elemei sávszélességének)</td>
<td>0,75</td>
</tr>
<tr>
<td>180-5700 Hz</td>
<td>0,86</td>
</tr>
<tr>
<td>85-11000 Hz</td>
<td>1</td>
</tr>
</tbody>
</table>
Relationship of measurable SI descriptors