How Is IndividualityExpressed in Voice?

Summer Term 12
Seminar on Personalization of HMI
Christian Wolff
Motivation

How Is Individuality Expressed in Voice?

„Greetings!“
„Good morning everyone!“
„Hello“
Motivation

„Good morning everyone!“

Speaker characteristics:
- pitch
- volume
- rate
- tone

11.05.2012
Motivation

 Speakers are different at all levels:
  - Realization of phonetic gestures
  - Choice of gestures used
  - Influence of context
  - Frequency of use

Derive speaker information from speech
Overview

1. Motivation
2. Human speech production
3. Influence of speaker characteristics
4. Applications
Overview

1. Motivation
2. Human speech production
3. Influence of speaker characteristics
4. Applications
How Is Individuality Expressed in Voice?

Volume in litres

- Inspiratory Reserve
- Tidal
- Expiratory Reserve
- Residual
Inhalation

Contraction of external intercostal muscles

Tension of diaphragm
Exhalation

Contraction of internal intercostal muscles (and muscles of the abdominal wall)

Relaxation of diaphragm
Respiration

- A typical inspiratory breath used for speech has an average volume of 1.5 litres.

- Minimum lung pressure for phonation production is 300 Pa, however 1000 Pa are typically used.

- Capability of the respiratory system is mainly influenced by gender, age or illness.
How Is Individuality Expressed in Voice?

Main articulator: Glottis position

Primary sound generation
Pharynx and oral/nasal cavities

Main articulators:

- Soft palate
- Tongue
- Lips
- Jaw

Criterion of phonetic classification
Speech Production Process Control

- Exact positioning and interaction between main articulators
- Influenced by speaking rates (fuzzy or overlapping articulation) or intoxication status
Sound Generation

Two separated processes:

1. Constriction in the larynx causes vibrations, which leads to rapid pressure variations (sound).
2. Passing through pharynx and oral/nasal cavities modifies the sound depending on shape and size of those cavities.
   - Final sound is influenced by the sound source as well as the filtering by the vocal tract tube.

- Source-filter model of speech production
Speech Generation

Manners of Articulation (exemplary)

Type of obstruction used in order to produce a consonant.

- **plosive:**
  
  *access to all cavities is closed*

- **nasal:**
  
  *oral cavity is completely closed, air exhales nasally*

- **fricative:**
  
  *continuous frication at the place of articulation (e.g. [f] or [s])*
The International Phonetic Alphabet

<table>
<thead>
<tr>
<th>CONSONANTS (PULMONIC)</th>
<th>Bilabial</th>
<th>Labiodental</th>
<th>Dental</th>
<th>Alveolar</th>
<th>Postalveolar</th>
<th>Retroflex</th>
<th>Palatal</th>
<th>Velar</th>
<th>Uvular</th>
<th>Pharyngeal</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plosive</td>
<td>p b</td>
<td>t d</td>
<td>t q</td>
<td>c j</td>
<td>k g</td>
<td>q g</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>Nasal</td>
<td>m ñ</td>
<td>n</td>
<td>n̄ n̄</td>
<td>l̄ l̄</td>
<td>L̄ N̄ l̄ N̄ l̄ N̄</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trill</td>
<td>B</td>
<td>r</td>
<td>η j̴</td>
<td>j̴ j̴</td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tap or Flap</td>
<td>V̔ r̴</td>
<td>r̴</td>
<td>R̴</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td>f v</td>
<td>θ ð s z</td>
<td>f z c j</td>
<td>x y χ r̴ h f</td>
<td>h f</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral fricative</td>
<td>l h̴</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximant</td>
<td>u j̴</td>
<td>l̴ j̴</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral approximant</td>
<td>l</td>
<td>l̴ w̴ l̴</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Overview

1. Motivation
2. Human speech production
3. Influence of speaker characteristics
4. Applications
Anatomical Influences

Anatomical Influences

Source-filter model of speech production

1. Length and mass of the vocal folds set the default pitch, range and voice quality of the source.

2. Length of the vocal tract influences its properties as filter.

- Considerable average differences between women and men

Available frequency range for women is about 184 - 880 Hz, for men about 87 - 415 Hz. Vibrational frequencies outside these ranges usually involve changes in the quality of vibration.
Anatomical Influences

- Variance in quality of speech produced by one breath
- Phonation technique influences air loss

Variance in quality of speech produced by one breath

Phonation technique influences air loss

<table>
<thead>
<tr>
<th>Volume in litres</th>
<th>Inspiratory Reserve</th>
<th>Tidal</th>
<th>Expiratory Reserve</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example in-depth: Age

- Many features involved in vocal aging:
  - growing/stiffening of the larynx
  - changes in flexibility of lungs and chest (lung capacity)
  - involution of muscles

- Hard to determine without extensive measurements and equipment (pneumotachograph)
Feature Levels

linguistic

prosodic

acoustic
Acoustic Features

- **Fundamental frequency**
  Also known as *pitch*. Describes the lowest frequency within the speech signal.

- **Jitter**
  Small frequency variation specified relative to the average fundamental frequency.
Acoustic Features

- **Intensity**
  Mean amplitude in a given region (loudness of the signal).

- **Shimmer**
  Small amplitude variation specified either absolute or relative to intensity.
Example in-depth: Age

- **Pitch is a very suitable feature:**
  - Young people have high pitch; voices are less diverse
  - Minimum pitch decreases with age
  - Average pitch increases or decreases depending on gender

- **Shimmer and jitter:**
  - High values for teenagers
  - Low values for adults
  - Highest values for seniors
Example in-depth: Age

- Almost all features have a higher variation for seniors:
  - Vocal tremor
  - Speech rate
  - Sound pressure level
Language Use

Language Use

Distinguishing characteristics:

- **Choice of language**
  - „exotic“ languages (Tewa, Quechua)
  - non-native languages and the influence of the speakers native language

- **Dialect**
  - geographical groups
  - age
  - other social influences
The process of generating a set of speech sounds for one specific utterance splits up in two components:

1. **Segmental component:** phonetic segments used to build the word sequence
   environment, native tongue

2. **Supra-segmental component:** stress pattern, timing and intonation of the sequence
   intention, emotion, intoxication
Overview

1. Motivation
2. Human speech production
3. Influence of speaker characteristics
4. Applications
Forensic Phonetics

Forensic Phonetics

- Verify the identity of speakers (recording, phone calls)

1. Naïve method: Identification by witness/victim

2. Expert speech analysis, including as many features as possible. Amount of positive feature matchings is critical for identification.
Mobile ShopAssist

How Is Individuality Expressed in Voice?

adult

teenager
How Is Individuality Expressed in Voice?
An Introduction to Speech Production and Description for Speaker Classification

Volker Dellwo, Mark Huckvale, and Michael Ashby

C. Müller (Ed.): Speaker Classification I, LNAI 4343
pp. 1–20, 2007
Springer-Verlag Berlin Heidelberg 2007