# THE LATENT SPACE OF SPEECH IN DEAF PEOPLE

## SPEECH INTELLIGIBILITY, AND SEGMENTAL AND PROSODIC CHARACTERISTICS IN THE SPEECH OF HEARING IMPAIRED SPEAKERS:

PhD, Martina Ozbič, assist.,
University of Ljubljana, Faculty of Education, Slovenia

Martina.ozbic@guest.arnes.si

Key words: **Deafness, ICF, speech production, intelligibility of speech.**

# background

- In early infancy, hearing loss significantly affects the development of speech in children with hearing impairment; normal development of speech is often disrupted and speech is less intelligible.
- The International Classification of Functioning, Disability and Health (ICF) lists three main categories of speech and voice functions: voice functions, articulation functions and fluency and rhythm of speech functions.

## Considering that disorders of both speech and voice occur in the speech of hearing-impaired speakers, the present research will:

- 1) analyse the speech of 91 hearing impaired speakers (aged from 5 to 23, M = 13 years, 56% males, 44% female) exhibiting an average hearing loss of air conductivity in the right ear of 99.06 dB (SD = 19.18), and in the left ear of 98.18 dB (SD = 19.78), according to the functions from ICF listed above, and to analyse the overall speech intelligibility and three important aspects of it;

- 2) compare latent space in the HI subjects' speech and the classification of ICF speech and voice functions, to see whether they overlap partially or totally; and
- 3) investigate the correlations between derived factors and speech intelligibility of HI subjects from a closed list of words; i.e., the prediction value of the factors.

# Our hypotheses were as follows:

- *H1: the factor analysis will derive three factors of fluency and rhythm, articulation, and voice quality, in accordance with the ICF, and the latent structure will be shown to fit with the classes of ICF speech and voice functions.*
- *H2: All of the factors exhibit statistically significant correlation with speech intelligibility, with the highest correlation reflecting prosodic functions.*
- *H3: The derived factors predict the degree of speech intelligibility.*

# The assessed variables referred to

- The audiological status (from 125 to 8000 Hz of air conduction, in both the right and left ears) and
- to speech, voice, and articulation production, according to the ICF:

- b310 Voice functions:
  - b3101: quality of voice (voice, resonance and timbre variables):
    - control of voice: quality of voice: breathy speech/hoarse voice, evaluated as absent (0) or present (1);
    - quality of timbre or oral / pharyngeal resonance:
      - sonorousness, observed by auditory evaluation and evaluated on a five-point scale;
      - formant spectrum clarity, observed by visual evaluation and evaluated on five-point scale;
      - cul-de-sac resonance, observed by auditory evaluation and evaluated as absent (0) or present (1);
    - nasal resonance: degree of nasality, observed by auditory evaluation and evaluated as absent (0) or present (1);

- b320: variables concerning articulation functions at the segmental level: the quality of articulation of words and the sum total of mistakes;
- b330 Fluency and rhythm of speech, or speech variables concerning speech flow:
  - b3300 fluency of speech
    - fluency of speech, evaluated on five-point scale;
    - syllabic fragmentation - irregular breaks in speech I., evaluated as absent (0) or present (1);
    - phoneme fragmentation – irregular breaks in speech II., evaluated as absent (0) or present (1);
    - interposition of sounds/phonemes between phonemes, evaluated as absent (0) or present (1);
    - interposition of sounds/phonemes between syllables, evaluated as absent (0) or present (1);
    - finger spelling, evaluated as absent (0) or present (1);

- b3301 rhythm of speech:
  - lengthening of  syllables, evaluated as absent (0) or present (1);
  - contrast reduction, evaluated as absent (0) or present (1);
  - scanned speech, evaluated as absent (0) or present (1);
- b3302: speed of speech, evaluated on five-point scale;
- b3303: melody of speech:
  - inappropriate prosody or intonation variability, evaluated as monotonous (1), normal (0) or too variable (2);

- Intelligibility of speech using a closed set of words, evaluated on five-point scale where each grade indicates that 20% of the words were pronounced intelligibly. According to some studies, the degree of oral speech communication skills of hearing-impaired children can be tested by means of speech intelligibility (Metz, 1985).

# procedures

- We recorded the subjects' speech with a Sony (DAT) digital recording device, model number TCD-D8, using a Sony microphone, model number ECM-719, and digital tapes (in 1999). We placed the microphone at a distance of 30 cm in front of the subject's mouth.
- The sound samples consisted of 5-10 minutes of reading/naming a closed set of words/pictures from the Slovenian three-position test for evaluating articulation of first-grade students.

# Editing and evaluation of the data

- The CoolEdit96/CoolEdit2000 program was used to prepare the recorded material and Praat and Speech Analyzer programs were used for analysing the resulting sound files.
- For data processing, we used Excel for Windows XP and the SPSS 13.0 statistical package for Windows.
- The data were evaluated twice, using the test-retest method (concordance values were 87%) and re-evaluation. The data were normalized.

# Outcomes & Results
# Basic statistics

In the speech of hearing-impaired speakers we observed

- a 38.5% occurrence of nasality,
- 53.8% occurrence of cul-de-sac resonance,
- 38.5% of hoarse voice,
- 20.7% of interposition of sounds between syllables and
- 19.8% of interposition of sounds between phonemes.
- 29.7% occurrence of syllabic fragmentation,
- 9.9% of phonemic fragmentation,

- 5.5% of finger spelling,
- 36.3% of lengthening of syllables,
- 46.2% of scanned speech and
- 38.5% of contrast reduction.
- 27.5% of all speech was monotone and
- 8.8% exhibited abnormal variation.

- The sonorousness of speech was
  - very poor in 12.1% of the subjects,
  - poor in 16.5%,
  - moderately poor in 33.0%,
  - slightly poor in 11.0% and normal in 27.5%.
- Formants spectrum clarity was
  - very poor in 19.8% if the subjects' speech, poor in 16.5%,
  - moderately poor in 31.9%,
  - slightly poor in 9.9% and
  - normal in 22.0%.
- Fluency of speech was
  - very poor in 17.6% of the subjects,
  - poor in 12.1%,
  - moderately poor in 11.0%,
  - slightly poor in 47.3% and
  - normal in 12.1%.

- The speed of speech was
  - very slow in 2.2% of the subjects,
  - slow in 13.2%,
  - moderately slow in 34.1%,
  - normal in 33.0% and
  - high in 17.6%.
- The intelligibility of speech was evaluated as
  - profoundly impaired in 9.9% of the subjects,
  - severely impaired in 22.0%,
  - moderately impaired in 14.3%,
  - slightly impaired in 25.3% and
  - normal in 28.6%.

# Factor analysis - identifing speech areas

- Factor analysis followed the KMO criteria (0.849 > 0.5), and Bartlett's test (aprox. Chi square = 797.097, df = 136, sig. 0.000 < 0.05).
- The extraction was done according to the method of principal components, resulting in four factors with the eigenvalue > 1.
- In Table 1, communalities are displayed from the highest (variable contrast reduction: 0.818) to the lowest (variable formant spectrum clarity: 0.483).

# Table 1: Communalities

| Variables | initial | extraction |
|---|---|---|
| Contrast reduction | 1.000 | .818 |
| Interposition of phonemes. sounds between phonemes | 1.000 | .763 |
| Nasality | 1.000 | .746 |
| Sum of articulation mistakes | 1.000 | .742 |
| Syllabic fragmentation | 1.000 | .737 |
| Lengthening of the syllables | 1.000 | .726 |
| Sonorousness | 1.000 | .716 |
| Cul de sac resonance | 1.000 | .696 |
| Fluency of speech | 1.000 | .650 |
| Hoarse voice | 1.000 | .613 |
| Scanned speech | 1.000 | .607 |
| Interposition of phonemes. sounds between syllables | 1.000 | .585 |
| Speed of speech | 1.000 | .575 |
| Phonemic fragmentation | 1.000 | .571 |
| Finger spelling | 1.000 | .566 |
| Prosody – intonation variability | 1.000 | .565 |
| Formant spectrum clarity | 1.000 | .483 |

- The factors (Table 2) explain 65.650% of the total variance,  or about two-thirds.
- Since the structure was not sufficiently clear, rotation of the matrix was performed using the Oblimin method with Kaiser normalization.
- The first factor explained 40.036% of total variance,
- the second 11.430%,
- the third 7.970% and
- the fourth 6.214% of total variance.

# Table 2: Total Variance Explained

| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings(a) |
|---|---|---|---|---|---|---|---|
| | Total | % of Variance | Cum.% | Total | % of Variance | Cum.% | Total |
| 1 | 6.806 | 40.036 | 40.036 | 6.806 | 40.036 | 40.036 | 5.598 |
| 2 | 1.943 | 11.430 | 51.466 | 1.943 | 11.430 | 51.466 | 3.732 |
| 3 | 1.355 | 7.970 | 59.436 | 1.355 | 7.970 | 59.436 | 3.875 |
| 4 | 1.056 | 6.214 | 65.650 | 1.056 | 6.214 | 65.650 | 1.910 |
| 5 - 17 | omitted | | | | | | |

# Table 3: Pattern matrix

| | | | Component | | | |
|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 |
| 1 | b3300 | syllabic fragmentation | .885 | | | |
| | b3301 | lengthening of the syllables | .769 | | | |
| | b3302 | speed of speech | -.346 | | | |
| | b3300 | fluency of speech | -.760 | | | |
| | b3301 | contrast reduction | .895 | | | |
| | b3301 | scanned speech | .736 | | | |
| | b3303 | prosody | .667 | | | |
| 2 | b3101 | breathy speech / hoarse voice | | -.792 | | |
| | b3101 | Cul de sac resonance | | -.787 | | |
| | b3101 | sonorousness | | .745 | | |
| | b3101 | formant spectrum clarity | | .608 | | |
| | b320 | sum of articulation mistakes | | -.517 | .324 | .353 |
| 3 | b3300 | Interposition of phonemes between phonemes | | | .783 | |
| | b3300 | Finger - spelling | | | .781 | |
| | b3300 | Interposition of phonemes between syllables | | | .657 | |
| | b3300 | phoneme fragmentation | .377 | | .539 | |
| 4 | b3101 | nasality | | | | .843 |

# The first factor

explains 40.036% of the total variance and includes those variables that refer to code b330 and its subclasses of fluency and rhythm of speech.

- It is a matter of fluency and of the temporal organization of speech: contrast reduction, syllabic fragmentation, syllabic lengthening, fluency of speech, scanned speech, prosody, speed of speech.
- The first factor is projected on the variables of larger segments than the phoneme: the syllable, word, and sentence.
- Thus, we see a general fluency factor of macrochaining at the general macro-rhythmical/temporal level.
- This is an overall dynamic factor of chaining, more specific a prosodic factor.

# The second factor

## is a more static factor and explains 11.430% of total variance.

- This factor includes voice and resonance variables from codes b3101 and b320 (hoarse voice, cul-de-sac resonance, sonorousness, formant spectrum clarity) and articulation (total number of articulation mistakes).
- This factor determines overall quality of speech production (good phonation – F0, articulation and resonance, or formants). It also involves sound production within the larynx, the pharynx and other speech organs and refers to sonorants, obstruents, and their spectra (generator, phonator, resonator, and articulator).
- The variable of the sum of articulation mistakes also is projected on the third and fourth factors, meaning that nasality and a degree of poor microchaining can affect articulation.

# The third factor
## explains 7.970% of total variance

- and is projected on the variables reflecting insertion of pauses and of phonemes or sounds between the borders of segments - such as interpositions of phonemes between phonemes and syllables, finger spelling, where interposition of the schwa occurs, and phoneme fragmentation (b3300).
- This factor reflects phoneme fluency on the micro-temporal/rhythmical level and on the segmental phonemic/phoneme level. It is a dynamic factor of microchaining.

# The fourth factor

## explains 6,214% of total variance

- and reflects accompanying nasal sounds, or nasality (code b3101).
- It reflects velopharyngeal inadequacy and it speaks to the quality of resonance (a static characteristic) during speech production.

# Correlations between the speech intelligibility and the 4 factors

- The four factors are in statistically significant correlation with the intelligibility of speech in HI subjects ($p < 0.01$) (Table 4).
- The correlation coefficients show a negative correlation, from -0.587 to -0.401, with the first, third and fourth factor, and a positive correlation (0.698) with the second one.
- The highest correlations appear in the second and first factor, showing that voice characteristics, resonance, good articulation and fluent prosodic chaining lead to intelligible speech.

# Table 4: Correlations between speech intelligibility and the derived factors

| | Prosodic, macrochaining | Voice/timbre/articulation | microchaining | nasality |
|---|---|---|---|---|
| **Intelligibility of speech** | **REGR factor score 1** | **REGR factor score 2** | **REGR factor score 3** | **REGR factor score 4** |
| **Spearman's rho** | -.587(**) | .698(**) | -.380(**) | -.401(**) |
| **Sig. (2-tailed)** | .000 | .000 | .000 | .000 |
| **N** | 91 | 91 | 91 | 91 |

# Regression analysis

We were interested in how the four factors predict speech intelligibility in HI speakers.

The values of a regression analysis are displayed, where the four factors are the predictors of the dependent variable: speech intelligibility in HI speakers.

The adjusted R square is 0.717, so our model explains 71% of the observed variance.

The ANOVA shows that there are significant differences at $p < 0.01$ and that the reliability of prediction is high.

- The regression equation is

*speech intelligibility = constant – 0.213 • factor 1 + 0.537 • factor 2 – 0.247 • factor 3 – 0.263 • factor 4.*

The second factor of articulation and voice quality has the highest coefficient (0.537), followed by the others.

# Discussion

The scope of the four factors covers the phenomenon of speech production rather exhaustively and reflects all of the components of speech production.

The criteria of the factor conceptualization are:

- A. dynamic versus static,
- B. micro versus macro, and
- C. oral versus nasal.

# A.Dynamic / static

- There are two factors reflecting the dynamic aspects of speech: the prosodic factor of macro speech fluency, and that of micro speech fluency. In other words, the first factor operates at the prosodic and syllabic macro level and the third at the phonemic level.
- There are two static factors: a voice-articulation-resonance oral factor (the second factor) and a nasal resonance factor (the fourth). The spectra are connected with the static dimension of speech and consequently with resonance, sonority, and formants.

- This axis shows the duality of speech production: the quality of speech units and how they are chained. The second and the fourth factor are factors of basic speech units; the first and third are factors showing how they are implemented.
- The first and third factors are dynamic factors and reflect the chaining of units, whereas the second and the fourth factor are static factors reflecting the quality of units themselves (quality of articulation and spectra).

# B. Units / prosodic

- The third factor and, to a degree, the second are factors in the production of the units of speech (phonemes and chaining of phonemes); the first, the second (partially), and
- the fourth factor are factors relating to the production of units larger than phonemes and reflect the prosodic elements of speech.

# C. Oral / nasal

- Factor analysis gives three oral factors (first, second and third): the two factors of speech chaining (first and third) and the second factor, reflecting quality of voice, resonance and articulation.
- In addition, there is also a nasal factor (the fourth).

- Our factor analysis mirrors speech production, covering the macro-temporal level of the chaining of syllables and the overall sequencing and chaining of words and the micro-temporal level where phoneme fluency affects the chaining of sounds.
- It also reflects voice and articulation characteristics and the role of velopharyngeal control, or nasality. It is interesting that nasality, which is the primary factor in the (in)appropriate functioning of resonators--velopharyngeal occlusion--and of functions connected to them, rather than directly to the voice organ, forms an additional dimension as a separate factor. It stands in opposition to the other oral factors.

- The ICF proposes a three-dimensional model of speech (b310, b320 and b330).

- The first and the third factor correspond to the b330 class (the fluency and rhythm of speech functions: b3300 fluency of speech; b3301 rhythm of speech; b3302 speed of speech; and b3303 melody of speech), while
- the second and fourth match the b310 category of voice functions (b3101 quality of voice) and the b320 classes (articulation).
- The fourth factor is hidden in b3101.

# Our first hypothesis can be partially refuted:

- that regarding the distribution of the factors of fluency and rhythm, articulation, and voice quality, and thus to the equality of ICF classes and derived factors, while we can accept the similarity between our and the ICF categorisations.
- In our research, there are four dimensions of speech, whereas there are three in the ICF; in our findings there is a distinction between macro and micro chaining, between dynamic and static characteristics of speech, and between oral and nasal dimensions.

- The distinction between the dynamic and static characteristics of speech corresponds to the studies (Metz et al. 1985, 1990; Tobey, 1995) and agrees with the theories of neurolinguistics, where the left hemisphere monitors the dynamic parts of speech (temporal information), and the right hemisphere monitors the static parts (spectral information) (Gandour, 1998, 213, Hellige 1998, 407)

# Hypothesis 2 can be partially accepted. We can accept the third hypothesis.

- Hypothesis 2 can be partially accepted. In our research, the correlation coefficients between intelligibility and the four factors are statistically significant. The highest correlation with the intelligibility of speech is that of the second factor (0.698), then the first one (- 0.587), the fourth one (0.401) and finally the third one (-0.380). Voice quality and articulation are the most important factors, followed by the prosodic factor.
- We can accept the third hypothesis, concerning the prediction of speech intelligibility, based on the predictor values of the four derived factors, at sig. $p < 0.01$

- The ICF and speech factor distributions are similar although not identical. Speech can be seen not only in terms of 3D factors of fluency, voice, and articulation, but also in terms of 4D factors on the micro- and macro-segmental levels: prosodic factors, voice, resonance and articulation quality and nasality.
- According to the correlation between the four factors, and to the results of the regression analysis, we can state that speech is really "a product of a series of interactive processes such as articulation, phonation, resonance and prosody" (De Bodt et al. 2002).

# Conclusions & implications

- These results could be useful in speech therapy, and not only for hearing-impaired speakers: the training in one element of several elements of speech can bring improvements in overall speech production.
- In speech therapy, articulation is not the only goal. Correct breathing, good phonation, overall good chaining of segments, velopharyngeal function and good control of the larynx can lead to intelligible speech. Knowledge of the latent space of verbal production represents a suitable starting point for rehabilitation at the vocal, speech segmental and suprasegmental level.

- By developing and improving all the elements of speech production, the goal of sufficiently intelligible speech in deaf or hard-of-hearing subjects can be achieved.
- Speech does not only mean good articulation, but rather a coordinated pneumo-phono-articulation-hearing system.

# Acknowledgements


- The author wishes to express her deep appreciation to the hearing-impaired pupils from the three Slovenian hearing centres, ZGN Ljubljana, CSG Maribor and CKSG Portorož, who gave their time and effort for this research.

- I am also grateful for the contribution of the speech and language pathologist during the investigation.

*Martina Ozbič*