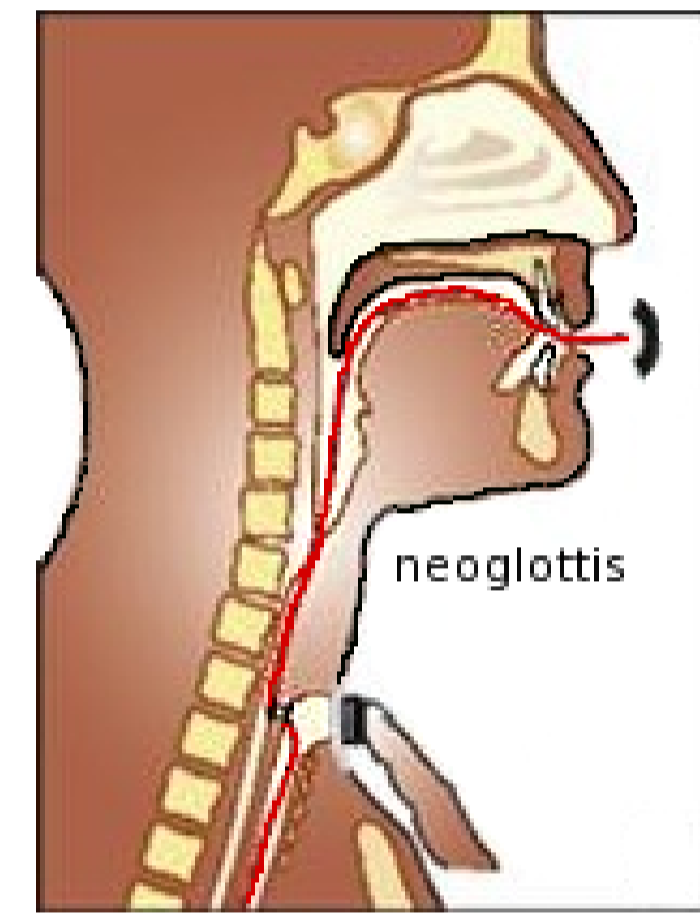


Abstract

- Use LPC analysis-synthesis to manipulate tracheoesophageal speech
- On-line experiment with expert judges
- Rate perceived intelligibility (7-point scale)
- Modeling the source amplitude improved speech most
- Regularizing pitch had no effect
- Using a fully synthetic voice source decreased intelligibility

Introduction

Tracheoesophageal speech (TE)



Pulmonary driven air passes from the trachea → prosthesis → pharyngoesophageal (PE) segment → oral cavity

“Voice” is generated by the neo-glottis in the PE.

- Intelligibility of TE speech often is low
- Lack of knowledge of the relation between intelligibility ⇔ underlying deficits
- Search for possibilities to improve therapy by modeling effects on speech

Speech manipulations

- Manipulate speech with LPC analysis-synthesis (*Linear Predictive Coding*)
- Compare LPC synthesis baseline to “improvements”

- Voice amplitude
- Pitch stability
- Source spectrum (pitch period shape)

Methods

Speech materials

- 16 TE speakers, Median age 58 (46-82)
- 30 recordings of sentence:
ook het weer heeft aan deze tocht meegewerkt
/ok at wer heft an dezə toxt meɣəwɛrkt/
(Eng: “The weather has also contributed to this trip”)

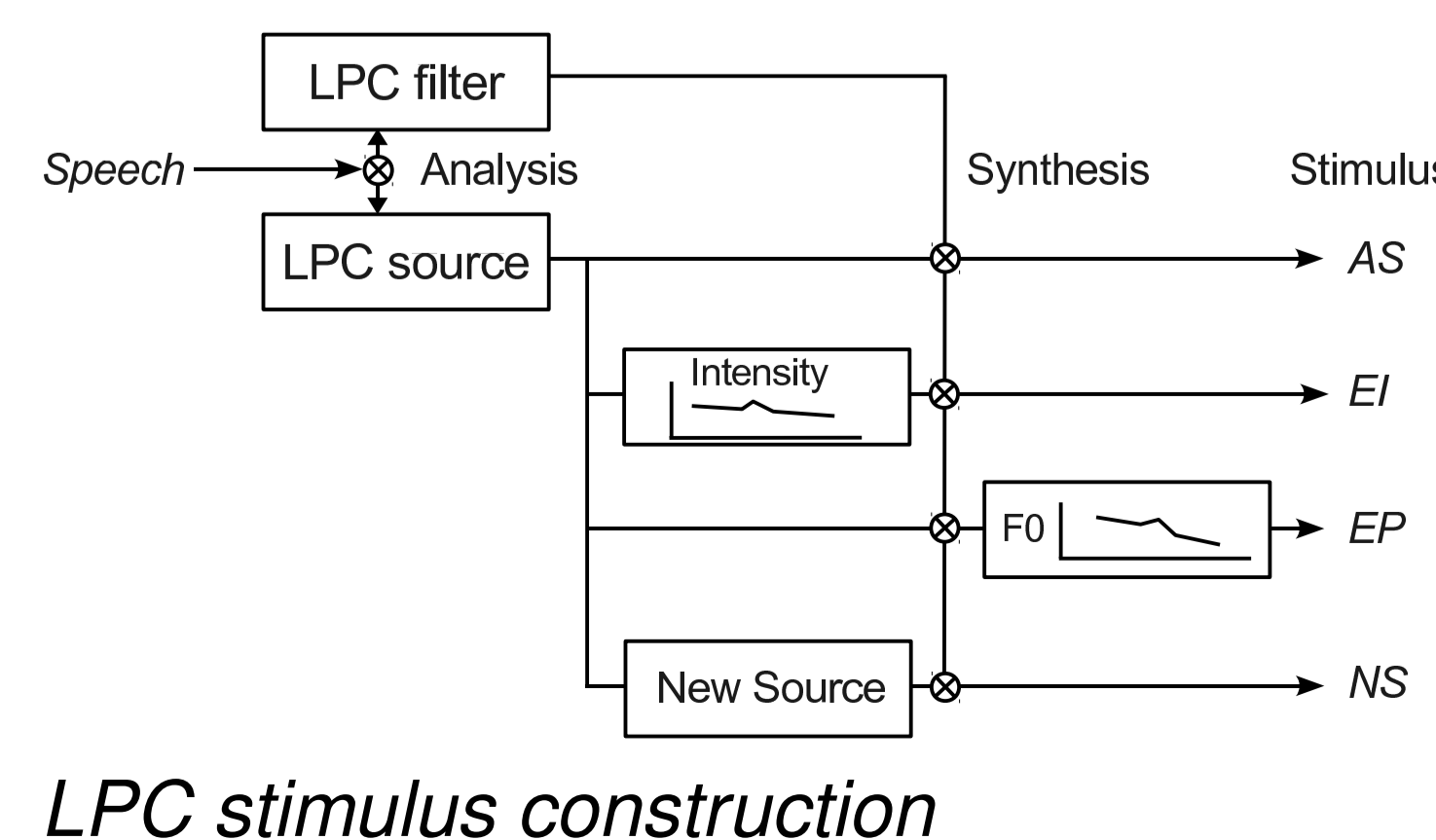
Subjects

- 6 Experienced speech therapists/foneticians
- On-line experiment
- Perceived intelligibility on a 7-point scale

Stimulus synthesis

Four types of stimuli:

- AS** baseline analysis-synthesis
- EI** regularized amplitude
- EP** regularized pitch
- NS** fully synthetic periods
- Synthesize only voiced parts (hand-labeled)*

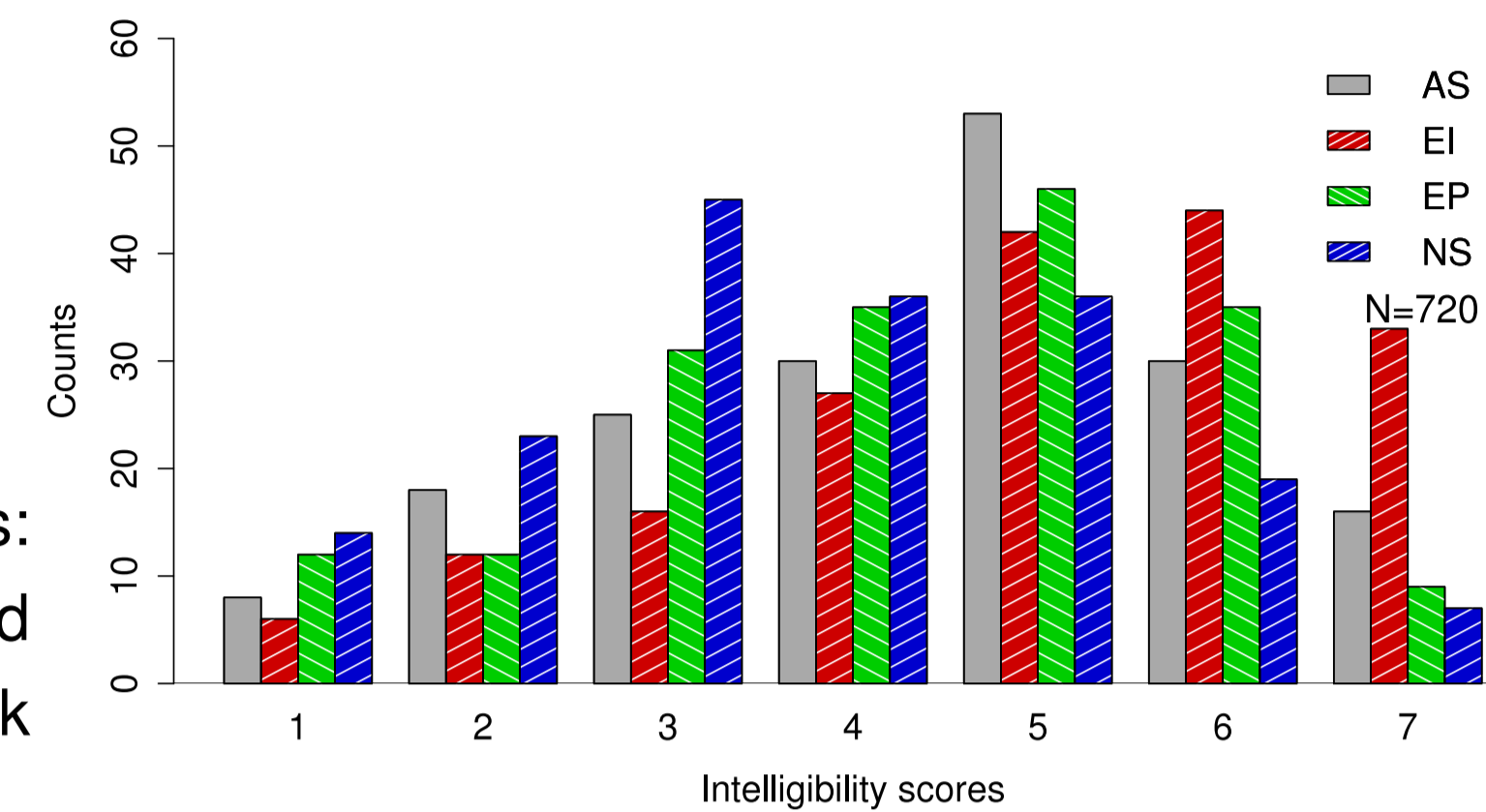


Results: Response Consistency

Distribution of responses

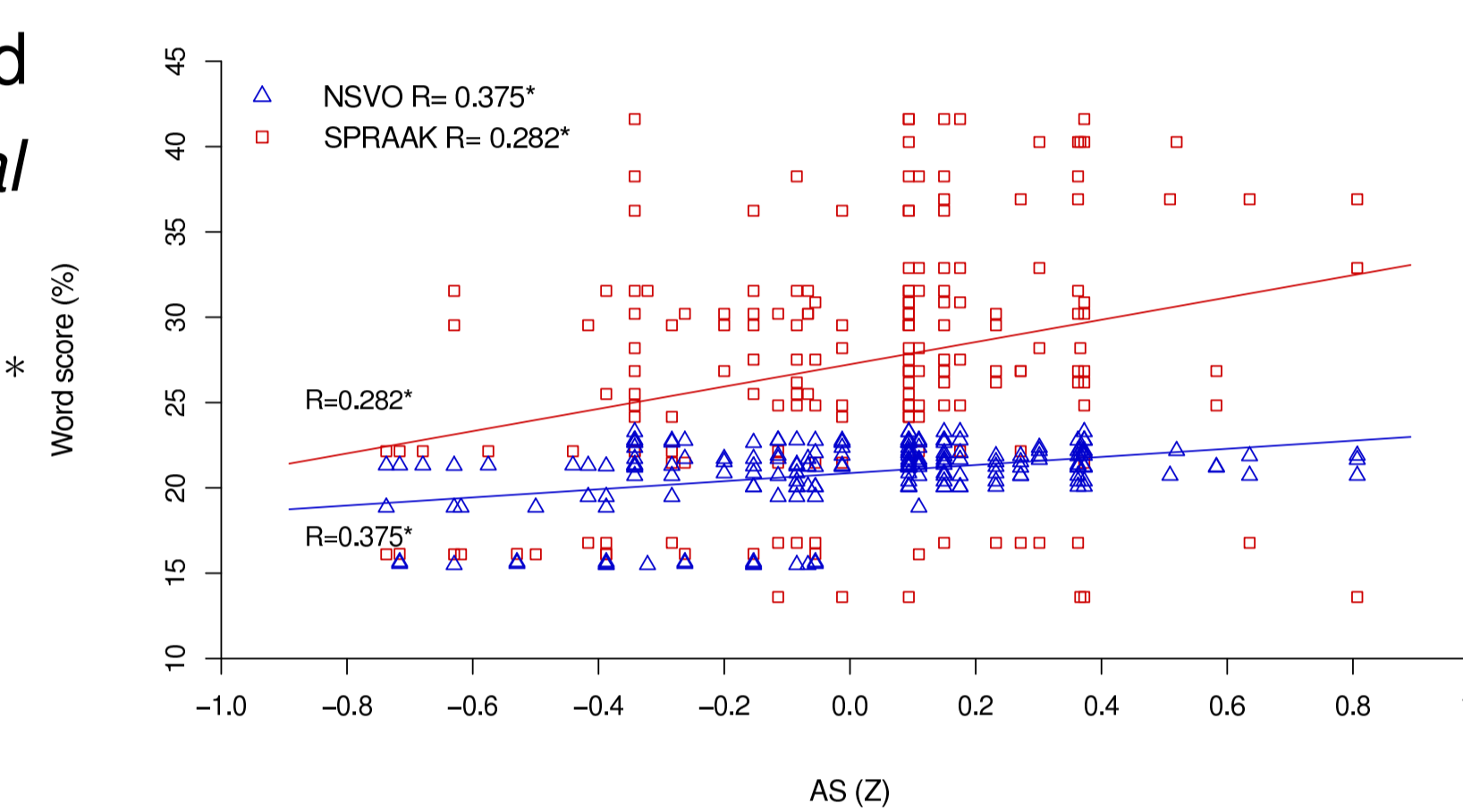
- 4 types x 180 responses
- Distributed over 1-7
- Ratings were consistent*
- Trend $EI > EP > AS > NS$

* Ratings were consistent over judges:
 $p < 0.001$ for each of AS, EI, EP, and NS; $\nu=29$, $\chi^2 > 99$, Friedman rank sum test



Intelligibility of original speech (V) versus baseline AS stimuli (H)

- Original and AS correlated
 - ASR scores (%) on original
 - AS responses (*Z-values*)
 - Weak correlation $R \leq 0.375^*$
- low quality synthesis from
1 Identification voiceless
2 LPC analysis-synthesis



* $p < 0.0002$

ASR scores

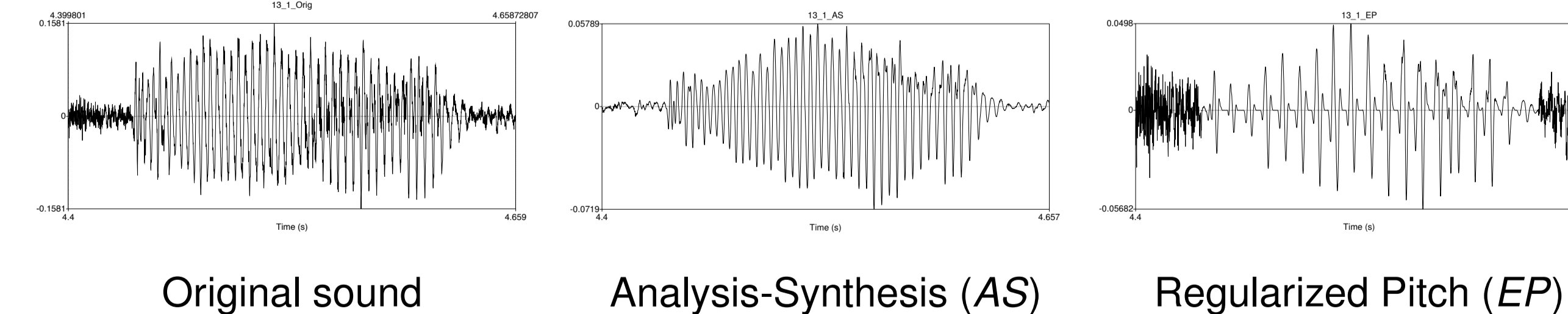
- 9 sentence, 149 word story read aloud by the TE speakers (carrier of stimulus sentence)
- NSVO: Phonological Features, average (%) recognition probability (ELIS Ghent Univ)
- SPRAAK: Word score (%) = number of words recognized (ESAT-PSI Leuven Univ)
Bag-of-Words model of sentences → No alignment

Stimulus quality and ratings

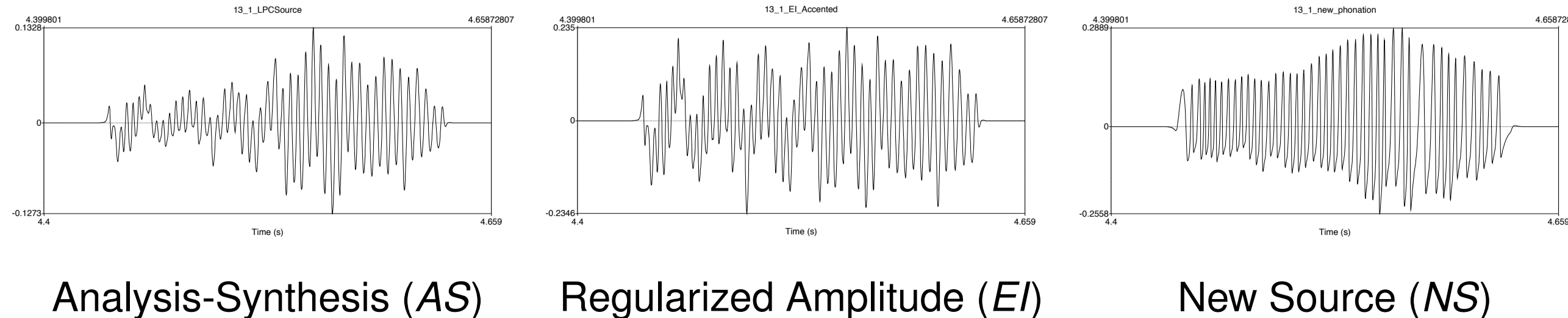
- Judges and ratings were consistent
- Rating task is feasible for speech therapists and phoneticians
- Original intelligibility differences were (somewhat) preserved in AS
- Synthesis quality is “fair” for low quality TE speech
- Synthesis quality is not good for normal speech

Examples of stimulus fragments (/ɣəwɛr/)

Sound waveforms



LPC source waveforms

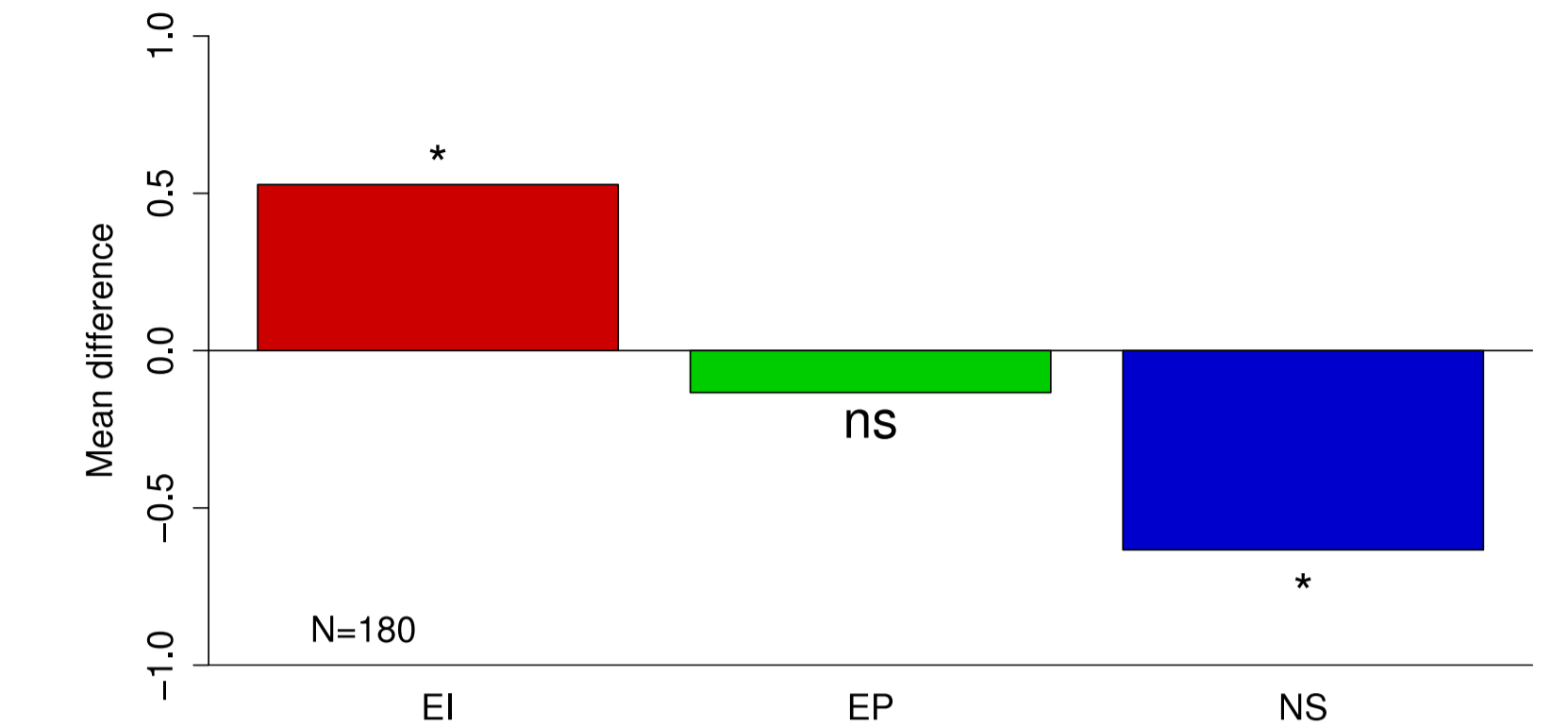


Results: Manipulation effects

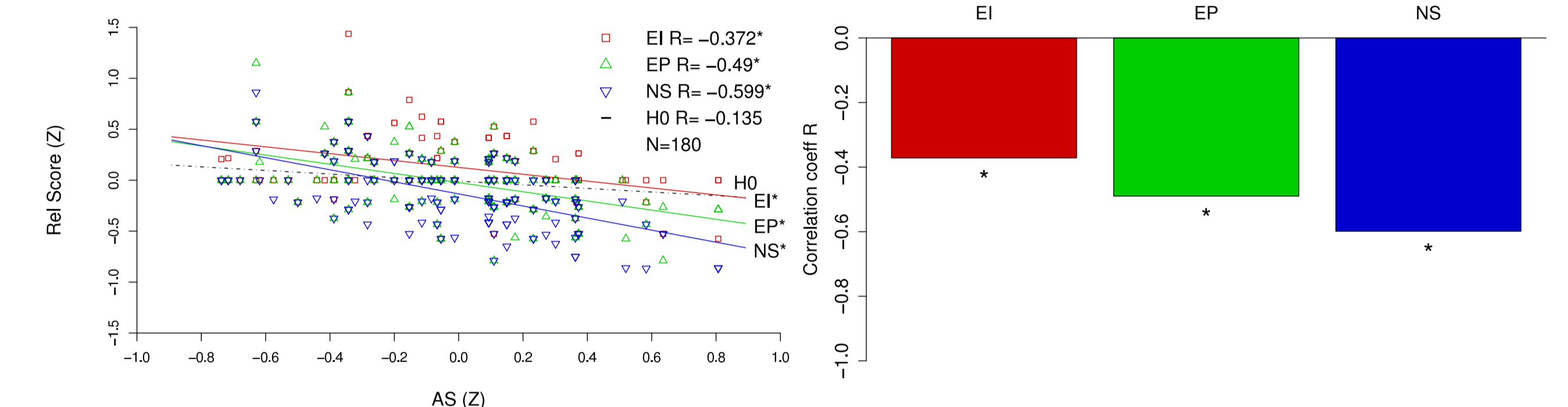
Mean effect of manipulation relative to baseline AS

- Perceived intelligibility
- EI** Improves significantly
- EP** Has no effect
- NS** Decreases significantly

* $p < 0.001$



Effect size versus baseline AS intelligibility



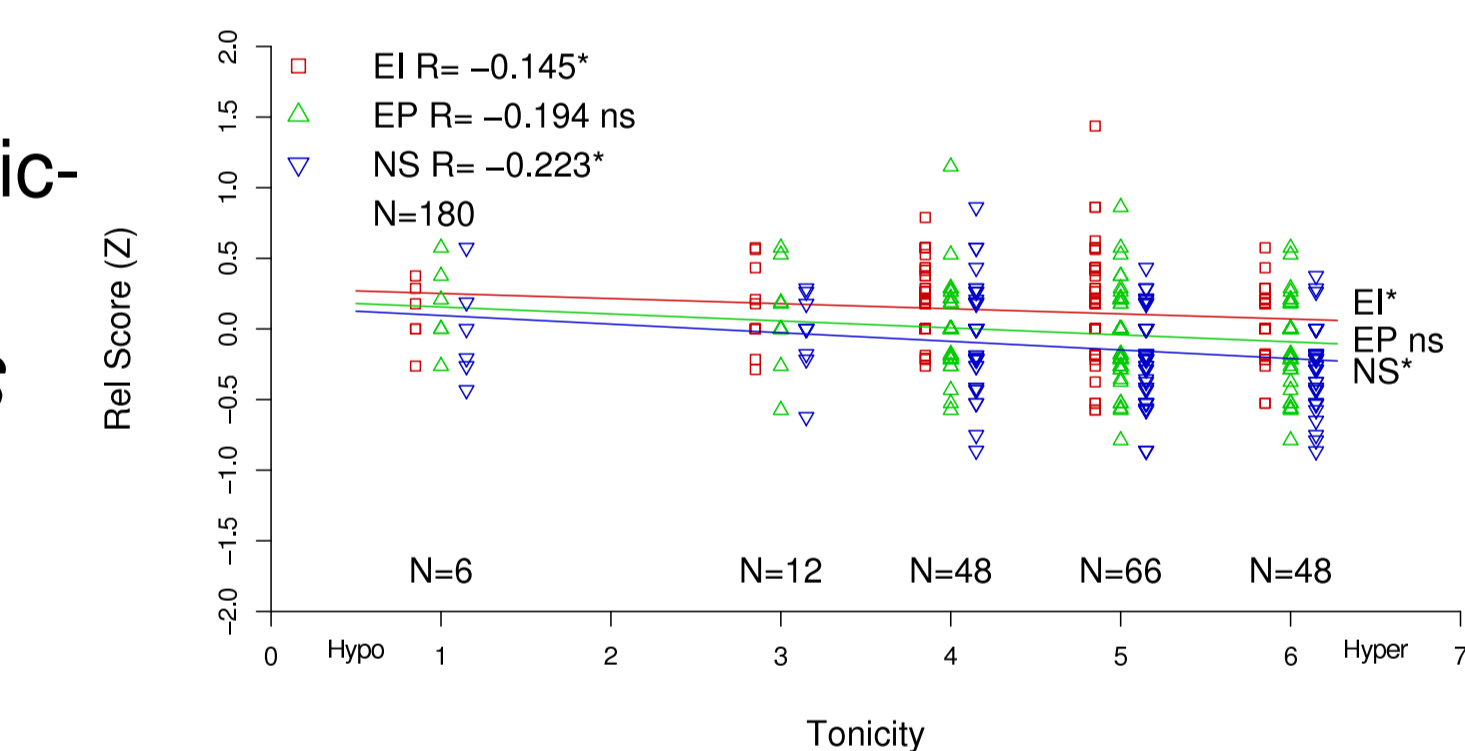
- Better baseline → Worse effect
- Low quality speech improves, high quality speech deteriorates

* $p < 0.001$

Hypo/hyper tonicity

- Expert rated Hypo/Hyper tonicity on 7-point scale
- Uncorrelated to original or AS
- Negative correlations

* $p < 0.005$



Discussion

- Perceived intelligibility of synthesized speech
- improves significantly with regularized source amplitude (EI)
- deteriorates significantly with synthetic source (NS)
- is unaffected by regularizing F_0 (EP)
- Low quality speech improves more than high quality speech
- Hypotonic improves more than Hypertonic speech

Conclusions

- Manipulating individual aspects of pathological speech is possible
- It can improve intelligibility
- Modeling a regular voice period amplitude was beneficial
- Replacing voice source with synthetic periods deteriorated quality
- Modeling a regular pitch period (stable F_0) had no effect
- Select speech features relevant to therapy
- ⇒ predictively synthesize speech after therapy?