

Abstract

The relationship between objective properties of speech and perceived pronunciation quality in read and spontaneous speech was examined.

Read and spontaneous speech of two groups of non-natives was scored for pronunciation quality by human raters.

The same material was analyzed by means of a continuous speech recognizer to calculate six temporal measures of speech quality.

The results show that temporal measures of speech are strongly related to pronunciation quality, in both read and spontaneous speech. Not all measures are as effective to predict pronunciation quality in spontaneous speech as they are in read speech.

1 Introduction

Recently, attempts have been made at developing automatic pronunciation tests by using continuous speech recognizers.

These studies have revealed that automatically obtained measures of speech quality are strongly correlated with scores assigned by human experts.

Most of these studies concern read non-native speech.

In this poster we explore whether this also holds for spontaneous non-native speech.

2 Goal

Exploring the relationship between automatic temporal measures and perceived pronunciation quality in read and spontaneous speech.

3 Method

Two independent experiments were conducted:

Experiment 1: read speech

Experiment 2: spontaneous speech

These experiments varied in:

- speakers
- speech mode: read versus spontaneous
- expert raters

4 Speakers

Experiment 1

60 non-native speakers (NNS)

speakers varied in:

- mother tongue
- gender

three proficiency levels (PLs):
PL1, PL2, PL3

Experiment 2

57 non-native speakers (NNS)

speakers varied in:

- mother tongue
- gender

two proficiency levels:
lower proficiency (LP)
higher proficiency (HP)

5 Speech material

Experiment 1

read speech

all speakers read the same sets of 10 phonetically rich sentences

Experiment 2

spontaneous speech

LP and HP answered two different sets of 8 questions
HP task was more cognitively demanding

An elaborated orthographic transcription, including disfluencies, was made for all speech material.

6 Expert raters

Experiment 1

three rater groups:
 3 phoneticians (ph)
 3 speech therapists (st1)
 3 speech therapists (st2)

speakers divided over the three raters in each group

raters did not receive any specific instructions

Experiment 2

ten Dutch teachers:
 5 for LP group
 5 for HP group

no overlap of LP and HP speech between rater groups

no specific instructions but knowledge about proficiency level to judge

7 Rating scales

The expert raters judged the speech material on the basis of the following four scales:

- Overall Pronunciation (OP) scale 1..10
- Segmental Quality (SQ) scale 1..10
- Fluency (FL) scale 1..10
- Speech Rate (SR) scale -5..+5

For each speaker one score on each of the four scales was calculated.

8 Automatic scores

An off-the-shelf CSR was used. A forced Viterbi alignment was applied to calculate the following scores:

art (articulation rate) = #phones / *tdur1*

ros (rate of speech) = #phones / *tdur2*

ptr (phonation / time ratio) = 100% * *tdur1* / *tdur2*

mlr (mean length of runs) = average #phones between pauses

#ps (# pauses (>.2 s) per second) = #pauses / *tdur2*

mlp (mean length of pauses (>.2 s))

tdur1 = total duration without pauses

tdur2 = total duration with pauses

9 Results: Inter-rater reliability [Cronbach's α]

raters ↓	OP	SQ	FL	SR	
ph	.89	.92	.96	.87	← scales read
st1	.89	.85	.88	.81	
st2	.87	.74	.83	.84	
RLP	.89	.82	.86	.89	spontaneous
RHP	.84	.81	.82	.80	

ph = phoneticians
 st1, st2 = speech therapists (two groups)
 RLP = raters of lower proficiency speakers
 RHP = raters of higher proficiency speakers

10 Results: Expert ratings [means]

scales ↓	read			spontaneous		← level
	PL1	PL2	PL3	LP	HP	
OP	4.32	4.22	5.30	5.79	4.72	
SQ	4.18	4.33	5.46	5.37	4.41	
FL	4.65	5.00	7.36	5.64	4.80	
SR	-1.37	-1.07	0.43	1.15	0.29	

Increase with proficiency level for read speech.
 Decrease with proficiency level for spontaneous speech.

11 Results: Objective scores [means]

autom. scores ↓	read				spontaneous			← level
	PL1	PL2	PL3	all NNS	LP&HP	LP	HP	
art	10.87	11.15	12.47	11.6	12.00	12.25	11.85	
ros	8.54	8.95	11.03	9.68	5.65	5.99	5.31	
ptr	77.97	79.62	88.28	82.7	47.10	49.32	44.92	
mlr	16.51	18.10	27.73	21.5	9.41	9.50	9.33	
#ps	0.37	0.34	0.17	0.28	0.52	0.52	0.52	
mlp	0.40	0.40	0.34	0.38	0.97	0.92	1.02	

increase decrease

12 Results: Correlations with fluency ratings

	art	ros	ptr	mlr	#ps	mlp	← autom. scale
RS	.83	.92	.86	.85	-.85	-.53	} read
SSLP	.06	.57	.46	.49	-.33	-.08	
SSHP	.05	.39	.39	.65	-.49	.00	} sponta- neous

Only correlation of automatic scores and fluency (FL) is shown.

RS = read speech
 SSLP = spontaneous speech, lower proficiency
 SSHP = spontaneous speech, higher proficiency

13 Discussion

Of the two factors that are important for pronunciation in read speech:

- the rate at which speakers articulate the sounds (*ros*)
- the frequency with which they pause (*mlr*)

the latter is most important for pronunciation in spontaneous speech.

mlr is a particularly good predictor of pronunciation in spontaneous speech.

mlr. contains not only pause frequency, but also distribution.

This suggests that pauses are tolerated, provided that sufficiently long uninterrupted stretches of speech are produced.

14 Conclusions

Most temporal measures of speech are strongly related to ratings of pronunciation quality. This is valid for both read and spontaneous non-native speech, but not all measures are as effective in spontaneous speech as they are in read speech to predict perceived pronunciation quality.

The degree to which temporal measures are useful in predicting pronunciation quality varies with speech style and the elicitation task.