

Automatic assessment of second language learners' fluency

Helmer Strik & Catia Cucchiarini
A²RT, Dept. of Language & Speech

University of Nijmegen
The Netherlands

Goal

to determine whether expert ratings of L2 learners' fluency can be predicted on the basis of automatically calculated measures

2 remarks:

- L2: Dutch as a Second Language
- automatic measures were calculated by means of ASR technology

2 experiments:

- 1. read speech
- 2. spontaneous speech

EXPERIMENT 1. READ SPEECH

Method

60 Non-Native Speakers (NNS)
various proficiency levels

2 sets of 5 phonetically rich utterances
read speech
orthographically transcribed

CSR: 38 monophones & lexicon
Viterbi alignment of speech signals &
orthographic transcriptions

→ segmentation on phone level

→ 7 automatic measures:
art, ros, ptr, mlr, #p, tdp, alp

Automatic measures

tdur1 = total duration of speech without pauses

tdur2 = total duration of speech with pauses

7 automatic measures:

1. art, articulation rate = # phones / tdur1

2. ros, rate of speech = # phones / tdur2

3. ptr, phon./time ratio = 100%*tdur1 / tdur2

4. mlr, mean length of runs =
mean(# phones between 2 pauses)

5. #ps, number of pauses per sec. = # pauses / tdur2

6. tdps, total dur. of pauses per sec. =
sum(duration of pauses) / tdur2

7. alp, average length of pauses =
mean(duration of pauses)

Human ratings

3 groups of 3 experts:

1. Phon : Phoneticians
2. ST1 : Speech Therapists 1
3. ST2 : Speech Therapists 2

scored the 10 sentences for fluency
on a scale ranging from 1 to 10

Results

interrater reliability coefficients (Cronbach's α)

raters	interrater reliability
Phon	0.96
ST1	0.88
ST2	0.83

Results

Correlations (corrected for attenuation)
between the fluency ratings
and the 7 automatic measures

	Phon	ST1	ST2	all
art	.82	.86	.79	.88
ros	.88	.93	.91	.97
ptr	.80	.86	.89	.91
mlr	.81	.84	.89	.90
#ps	-.82	-.89	-.90	-.90
tdps	-.79	-.86	-.87	-.89
alp	-.50	-.52	-.55	-.56

Conclusions

for read speech

automatic assessment of L2 learners' fluency
is feasible

what about spontaneous speech ?

EXPERIMENT 2. SPONTANEOUS SPEECH

Method

60 subjects, 2 groups

1. 30 LP, lower intermediate level
2. 30 HP, higher intermediate level

answers to 8 open questions

extemporaneous / spontaneous speech

orthographically transcribed

CSR: 38 monophones & lexicon

Viterbi alignment of speech signals &
orthographic transcriptions

→ segmentation on phone level

→ 7 automatic measures:

art, ros, ptr, mlr, #p, tdp, alp

Human ratings

2 groups of 5 teachers of Dutch:

1. RLP, for the 30 LP subjects
2. RHP, for the 30 HP subjects

scored the 8 sentences for fluency
on a scale ranging from 1 to 10

Results

interrater reliability coefficients (Cronbach's α)

raters	interrater reliability
RLP	0.86
RHP	0.82

Results

Means for read and spontaneous speech

	read speech	spontaneous speech
	60 NNS	30 LP & 30 HP
art	11.6	12.0
ros	9.68	5.65
ptr	82.7	47.1
mlr	21.5	9.41
#ps	0.28	0.52
tdps	0.12	0.49
alp	0.38	0.97

Results

Correlations (corrected for attenuation)
between the fluency ratings
and the 7 automatic measures

	read speech	spontaneous speech	
	all	RLP	RHP
art	.88	.07	.05
ros	.97	.62	.43
ptr	.91	.49	.43
mlr	.90	.53	.72
#ps	-.90	-.36	-.54
tdps	-.89	-.49	-.45
alp	-.56	-.09	-.01

Conclusions

- ❑ human ratings of fluency appear to be more dependent on the frequency of pauses than on the length of pauses

- ❑ some automatic measures that are suitable for read speech cannot be employed in spontaneous speech:

i.e. those automatic measures that do **NOT** contain information about the frequency of the pauses (art & alp)

- ❑ automatic assessment of L2 learners' fluency is feasible, for read and for spontaneous speech