Perceptual experiments
sir-skur-spur-stir

Amy Beeston & Guy Brown

19 May 2010

1. Introduction
2. Experiment 1: cutoff
    - Set up
    - Results
3. Experiment 2: reverse
    - Set up
    - Results
4. Discussion

introduction

Background

- Based closely on Watkins' sir-stir paradigm
- Gather human data for ASR comparison with/without constancy model
- Investigate effect of reverberation on stop consonants esp. place of articulation
- Replicate compensation for reverberation
    - in another lab
    - with naturalistic speech, not interpolated stimuli
    - with further unvoiced stop consonants {k,p,t}

Comparison with Watkins' sir-stir work
Similarities

Two experiments (works in progress)

- *cutoff*: frequency effects
  Watkins and Makin, JASA 2007 etc.
- *reverse*: time-direction effects
  Watkins, JASA 2005, experiment 5

Comparison with Watkins' sir-stir work
Differences

Listener data

- consonant confusions (not category boundary shifts)

		responses			
		sir	skur	spur	stir
stimuli	sir	■			
	skur	▒	■		
	spur	▒		■	
	stir	▒			■

- percentage correct
- relative information transferred
- something else?

% correct and relative information transferred (RIT)

$m =$

$$\begin{vmatrix} 20 & 0 & 0 & 0 \\ 0 & 20 & 0 & 0 \\ 0 & 0 & 20 & 0 \\ 0 & 0 & 0 & 20 \end{vmatrix}$$

$RIT(m) = 1$
$\%\ \text{correct}(m) = 100$

$m =$

$$\begin{vmatrix} 5 & 5 & 5 & 5 \\ 5 & 5 & 5 & 5 \\ 5 & 5 & 5 & 5 \\ 5 & 5 & 5 & 5 \end{vmatrix}$$

$RIT(m) = 0$
$\%\ \text{correct}(m) = 25$

- RIT reflects information about *pattern* of errors
- reflects complexity of task - useful for ASR - different sized vocabularies OK

$$RIT = H(X : Y)\ /\ H(X)$$

$H(X : Y)$ is the mutual information of X and Y
$H(X)$ is the self-information (entropy) of X

Ref: Smith (1990)

cutoff

experiment 1

cutoff experiment

Aim:

- find appropriate parameter set for future experiments
- should allow
  - effect of reverberation on test word
  - compensation due to reverberation on context

Prediction:

- Extreme low-pass filtering increases misclassification rate also blocks compensation for reverberation

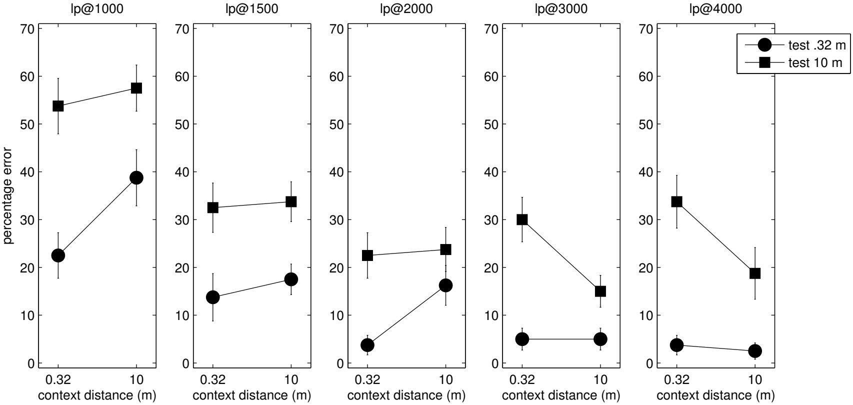
Stimuli (cutoff)

1600 stimuli = 20 talkers × 4 words × 4 distances × 5 cutoffs

- 80 Articulation Index Corpus utterances
  20 talkers, 4 test words {sir, skur, spur, stir}
- 4 reverberation conditions
  L-shaped room {near-near, near-far, far-near, far-far}
- 5 low-pass filter cutoff frequencies
  8th order Butterworth {1000, 1500, 2000, 3000, 4000} Hz

Each utterance once to each listener
1 group of 20 subjects

Results (cutoff) i. percentage error



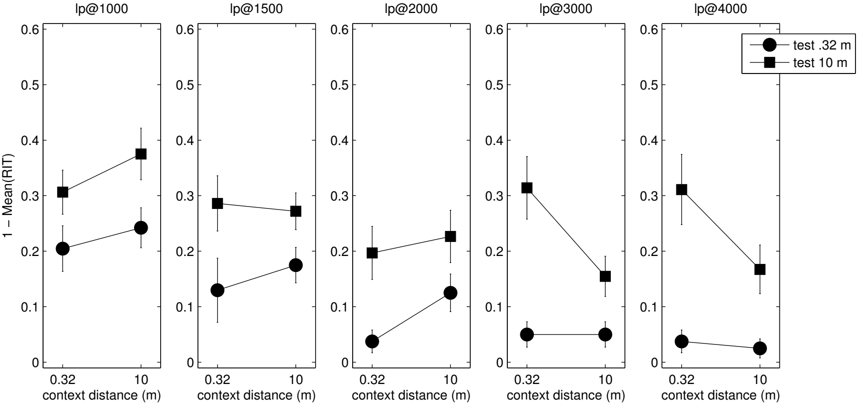
ANOVA (cutoff)
i. percentage correct

- 3-way repeated measures, all within-subject factors
- Independent variables
  - test-word distance (2 levels)
  - context distance (2 levels)
  - low-pass filter cutoff frequency (5 levels)
- Dependent variable
  - percentage correct

ANOVA (cutoff) results
i. percentage correct

- Significant main effects
  - test $F(1, 19) = 79.28, p < 0.001$
  - cutoff $F(4, 76) = 24.48, \epsilon_{HF} = 0.70, p < 0.001$
- Significant interactions
  - test × context $F(1, 19) = 8.47, p < 0.01$
  - context × cutoff $F(4, 76) = 4.227, \epsilon_{HF} = 0.90, p < 0.01$
- No other significant F-ratios

Results (cutoff) ii. relative information transferred



ANOVA (cutoff)
ii. relative information transferred

- 3-way repeated measures, all within-subject factors
- Independent variables
  - test-word distance (2 levels)
  - context distance (2 levels)
  - low-pass filter cutoff frequency (5 levels)
- Dependent variable
  - relative information transferred

ANOVA (cutoff) results
ii. relative information transferred

- Significant main effects
  - test $F(1, 19) = 59.27, p < 0.001$
  - cutoff $F(4, 76) = 9.19, \epsilon_{HF} = 0.96, p < 0.001$
- Significant interactions
  - context × cutoff $F(4, 76) = 2.593, \epsilon_{HF} = 1.0, p < 0.05$
- no other significant F-ratios
  - no significant interaction of test × context by this measure

Conclusion (cutoff)

Interim conclusion:

Compensation replicated best at 3 and 4 kHz cutoff conditions
Use 4 kHz cutoff frequency for future experiments

reverse

experiment 2

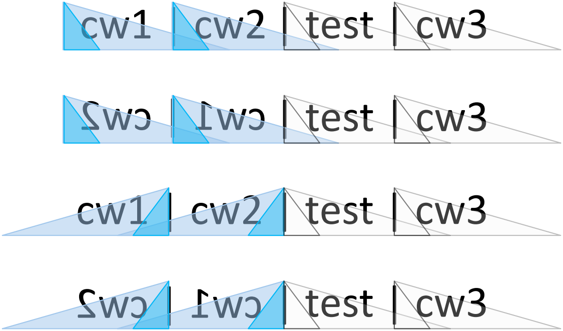
Stimuli (reverse)

1280 stimuli = 20 talkers × 4 words × 4 distances × 4 contexts

- Articulation Index Corpus
  20 talkers, 4 test words {sir, skur, spur, stir}
- Everything low-pass filtered
  8th order Butterworth, cutoff at 4 kHz
- 4 reverberation conditions
  L-shaped room {near-near, near-far, far-near, far-far}
- 4 preceding context conditions
  {forward, reverse} speech × {forward, reverse} reverb

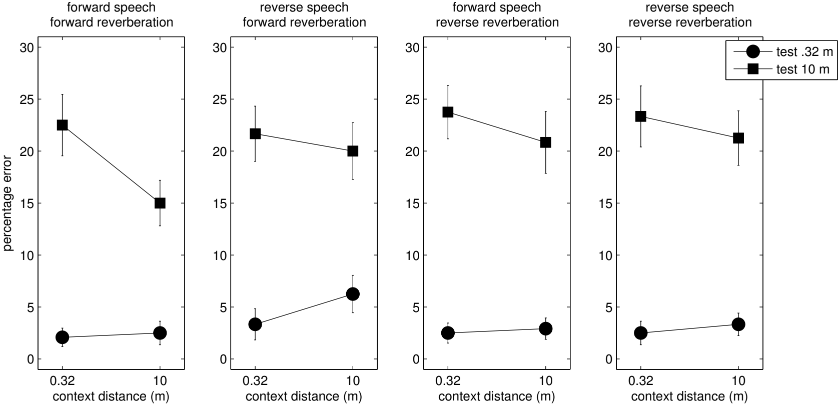
Each utterance once to each listener
48 subjects = 3 groups of 16

Stimuli (reverse)

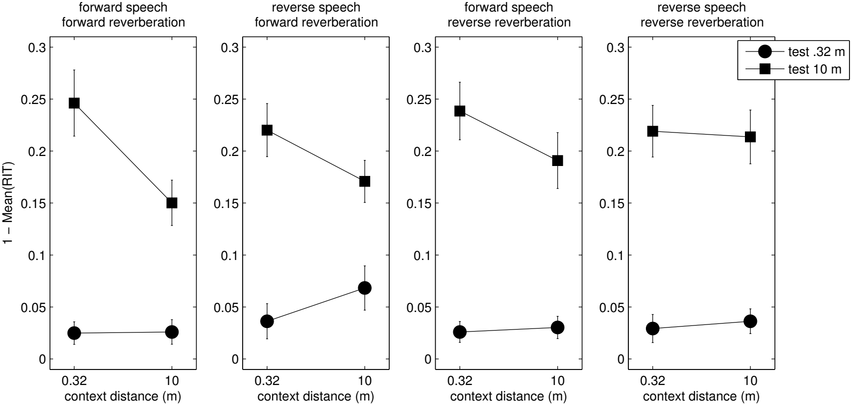


- Forward reverb cases: context reverb overlaps test word
- Reverse reverb cases: reverb during test word does not vary with context distance nn=fn, nf=ff

Results (reverse) i. percentage correct



Results (reverse) ii. relative information transferred



ANOVA (reverse)

- 4-way repeated measures, all within-subject factors
- Independent variables
  - test-word distance (2 levels)
  - context distance (2 levels)
  - speech direction (2 levels)
  - reverberation direction (2 levels)
- Dependent variable
  - i percentage correct
  - ii. relative information transferred

ANOVA (reverse) results

Significant main effects

- i. % correct: test $F(1, 47) = 240.0, p < 0.001$
- ii. RIT: test $F(1, 47) = 189.5, p < 0.001$
- ii. RIT: context $F(1, 47) = 5.7, p < 0.05$

Significant interactions

- i. % correct: test × context $F(1, 47) = 4.71, p < 0.05$
- ii. RIT: context × test $F(1, 47) = 7.9, p < 0.01$

No other significant F-ratios

ANOVA (reverse) significance per speech & reverb direction

	fwd	fwd	rev	fwd	fwd	rev	rev	rev
	speech	reverb	speech	reverb	speech	reverb	speech	reverb
	%	RIT	%	RIT	%	RIT	%	RIT
C	nearly	yes	no	no	no	no	no	no
T	yes	yes	yes	yes	yes	yes	yes	yes
C×T	yes	yes	no	nearly	no	no	no	no

Conclusion (reverse)

Interim conclusion:

- Fwd-fwd case shows typical compensation pattern
- Reverse reverberation seems to remove main effect of context-distance
- But...
  choice of dependent variable influences results considerably

discussion

Differentiating error patterns

$$m = \begin{vmatrix} 20 & 0 & 0 & 0 \\ 0 & 20 & 0 & 0 \\ 0 & 0 & 20 & 0 \\ 0 & 0 & 0 & 20 \end{vmatrix}$$

$RIT(m) = 1$
% correct$(m) = 100$

$$m = \begin{vmatrix} 5 & 5 & 5 & 5 \\ 5 & 5 & 5 & 5 \\ 5 & 5 & 5 & 5 \\ 5 & 5 & 5 & 5 \end{vmatrix}$$

$RIT(m) = 0$
% correct$(m) = 25$

$$m = \begin{vmatrix} 20 & 0 & 0 & 0 \\ 5 & 5 & 5 & 5 \\ 5 & 5 & 5 & 5 \\ 5 & 5 & 5 & 5 \end{vmatrix}$$

$RIT(m) = 0.190$
% correct$(m) = 43.75$

$$m = \begin{vmatrix} 20 & 0 & 0 & 0 \\ 15 & 5 & 0 & 0 \\ 15 & 0 & 5 & 0 \\ 15 & 0 & 0 & 5 \end{vmatrix}$$

$RIT(m) = 0.192$
% correct$(m) = 43.75$

Differentiating error patterns

$$m = \begin{vmatrix} 20 & 0 & 0 & 0 \\ 5 & 5 & 5 & 5 \\ 5 & 5 & 5 & 5 \\ 5 & 5 & 5 & 5 \end{vmatrix}$$

$RIT(m) = 0.190$
$\%\ \text{correct}(m) = 43.75$
$FP_{sir} = 15$

$$m = \begin{vmatrix} 20 & 0 & 0 & 0 \\ 15 & 5 & 0 & 0 \\ 15 & 0 & 5 & 0 \\ 15 & 0 & 0 & 5 \end{vmatrix}$$

$RIT(m) = 0.192$
$\%\ \text{correct}(m) = 43.75$
$FP_{sir} = 45$

Receiver operating characteristic (ROC)

$$\begin{array}{cccc|cccc|cccc|cccc}
\text{sir} = &&&& \text{skur} = &&&& \text{spur} = &&&& \text{stir} = &&& \\
\mathrm{TP} & \mathrm{FN} & \mathrm{FN} & \mathrm{FN} & \mathrm{TN} & \mathrm{FP} & \mathrm{TN} & \mathrm{TN} & \mathrm{TN} & \mathrm{TN} & \mathrm{FP} & \mathrm{TN} & \mathrm{TN} & \mathrm{TN} & \mathrm{TN} & \mathrm{FP} \\
\mathrm{FP} & \mathrm{TN} & \mathrm{TN} & \mathrm{TN} & \mathrm{FN} & \mathrm{TP} & \mathrm{FN} & \mathrm{FN} & \mathrm{TN} & \mathrm{TN} & \mathrm{FP} & \mathrm{TN} & \mathrm{TN} & \mathrm{TN} & \mathrm{TN} & \mathrm{FP} \\
\mathrm{FP} & \mathrm{TN} & \mathrm{TN} & \mathrm{TN} & \mathrm{TN} & \mathrm{FP} & \mathrm{TN} & \mathrm{TN} & \mathrm{FN} & \mathrm{FN} & \mathrm{TP} & \mathrm{FN} & \mathrm{TN} & \mathrm{TN} & \mathrm{TN} & \mathrm{FP} \\
\mathrm{FP} & \mathrm{TN} & \mathrm{TN} & \mathrm{TN} & \mathrm{TN} & \mathrm{FP} & \mathrm{TN} & \mathrm{TN} & \mathrm{TN} & \mathrm{TN} & \mathrm{FP} & \mathrm{TN} & \mathrm{FN} & \mathrm{FN} & \mathrm{FN} & \mathrm{TP}
\end{array}$$

Confusions (cutoff) lp@4000 Hz

@nf	sir	skur	spur	stir
sir	18	0	0	2
skur	3	15	0	2
spur	7	2	10	1
stir	8	1	1	10

@ff	sir	skur	spur	stir
sir	16	1	1	2
skur	0	16	0	4
spur	2	1	14	3
stir	1	0	0	19

@nn	sir	skur	spur	stir
sir	19	0	0	1
skur	0	20	0	0
spur	0	1	18	1
stir	0	0	0	20

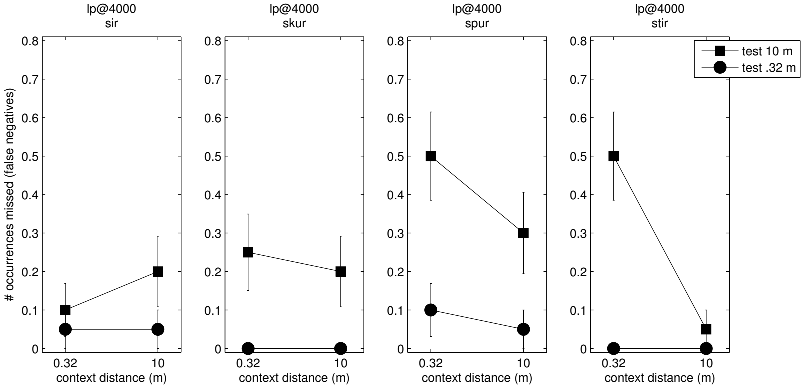
Confusions (reverse) fwd-fwd

@nf	sir	skur	spur	stir
sir	53	2	1	4
skur	11	47	2	0
spur	11	6	41	1
stir	13	2	0	45

@ff	sir	skur	spur	stir
sir	51	0	0	9
skur	2	52	1	5
spur	1	7	47	5
stir	4	2	0	54

@nn	sir	skur	spur	stir
sir	58	1	0	1
skur	1	59	0	0
spur	0	0	60	0
stir	0	2	0	58

Word-by-word (cutoff) lp@4000 Hz
i. False negatives

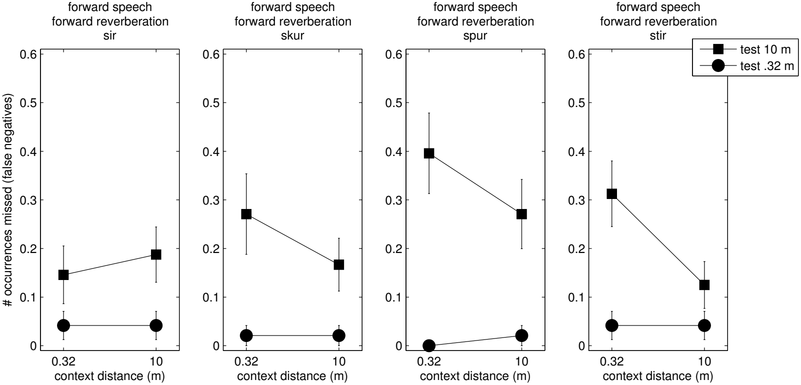


ANOVA (cutoff) lp@4000 Hz
i. False negatives

Independent variables (levels): context (2), test (2), word (4)
Dependent variable: # false negative responses

- Significant main effects
  - context $F(1, 47) = 9.67, p < 0.05$
  - test $F(1, 47) = 21.08, p < 0.001$
  - word $F(3, 141) = 42.17, \epsilon_{HF} = 0.44, p < 0.001$
- Significant interactions
  - context × test $F(1, 47) = 8.32, p < 0.01$
  - test × word $F(3, 141) = 2.82, \epsilon_{HF} = 0.81, p < 0.05$

Word-by-word (reverse) fwd-fwd
i. False negatives

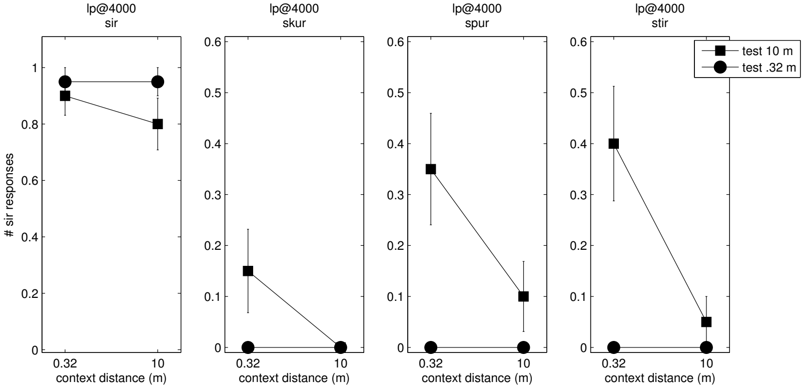


ANOVA (reverse) fwd-fwd
i. False negatives

Independent variables (levels): context (2), test (2), word (4)
Dependent variable: # false negative responses

- Significant main effect
  - test $F(1, 47) = 61.74, p < 0.001$
- Significant interactions
  - context × test $F(1, 47) = 4.14, p < 0.05$
  - test × word $F(3, 141) = 2.82, \epsilon_{HF} = 1.0, p < 0.05$

Word-by-word (cutoff) lp@4000 Hz
ii. Sir responses

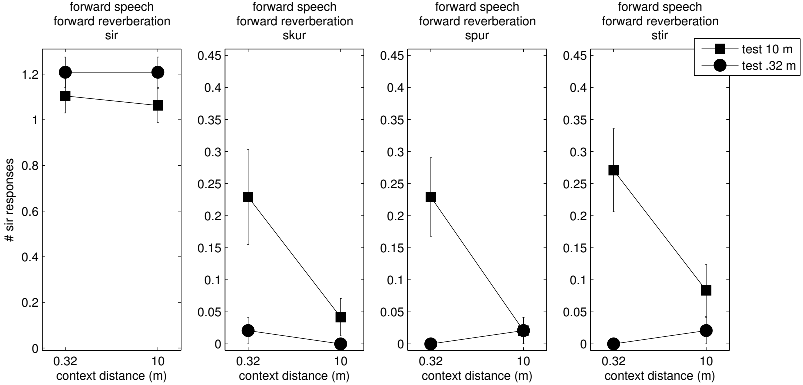


ANOVA (cutoff) lp@4000 Hz ii. Sir responses

Independent variables (levels): context (2), test (2), word (4)
Dependent variable: # sir responses

- Significant main effects
  - context $F(1,47) = 13.64, p < 0.01$
  - test $F(1,47) = 10.422, p < 0.01$
  - word $F(3,141) = 479.01, \epsilon_{HF} = 0.87, p < 0.001$
- Significant interactions
  - context × test $F(1,47) = 11.81, p < 0.01$
  - test × word $F(3,141) = 7.28, \epsilon_{HF} = 0.85, p < 0.01$

Word-by-word (reverse) fwd-fwd
ii. Sir responses



ANOVA (reverse), fwd-fwd
ii. Sir responses

Independent variables (levels): context (2), test (2), word (4)
Dependent variable: # sir responses

- Significant main effects
  - context $F(1, 47) = 7.96, p < 0.01$
  - test $F(1, 47) = 7.30, p < 0.05$
  - word $F(3, 141) = 704.64, \epsilon_{HF} = 0.99, p < 0.001$
- Significant interactions
  - context × test $F(1, 47) = 8.044, p < 0.01$
  - test × word $F(3, 141) = 7.09, \epsilon_{HF} = 0.70, p < 0.01$

Recap

Much work to do on analysis of current results

Future experiments to be designed with ASR experiments in mind (esp. to help tune constancy model)

Thanks...

extras

5 Appendix

- References
- Stimuli creation
- Stimuli partitioning details
- Additional results

Further reading

A.M. Smith. On the use of the relative information transmitted (RIT) measure for the assessment of performance in the evaluation of automated speech recognition (ASR) devices. In Australian International Conference on Speech Science and Technology, pages 368–373, 1990.

A.J. Watkins. Perceptual compensation for effects of reverberation in speech identification. J. Acoust. Soc. Am., 118(1):249–262, 2005.

A.J. Watkins and S.J. Makin. Steady-spectrum contexts and perceptual compensation for reverberation in speech identication. J. Acoust. Soc. Am., 121(1):257–266, 2007.

A.J. Watkins and S.J. Makin. Perceptual compensation for reverberation in speech identication: Effects of single-band, multiple-band and wideband noise contexts. Acta. Acust. United Ac., 93:403–410, 2007.

J. Wright, Articulation Index. Linguistic Data Consortium, Philadelphia, 2005.

Articulation Index Corpus

Talker	sir
f101	they recognize sir entirely
m102	anyone detect sir evenly
f103	you utter sir more
m104	noone see sir today
f105	you pronounce sir easily
f106	we notice sir sometime
m107	I echo sir today
f108	people watch sir clearly
f109	we show sir tenth
m110	you ponder sir first
m111	we notice sir seventh
m112	I echo sir happily
f113	noone suggest sir steadily
m114	everyone notice sir anyway
m115	I evoke sir precisely
m116	people study sir only
m117	everyone study sir sixth
m118	they read sir properly
f119	they see sir easily
m120	people note sir typically

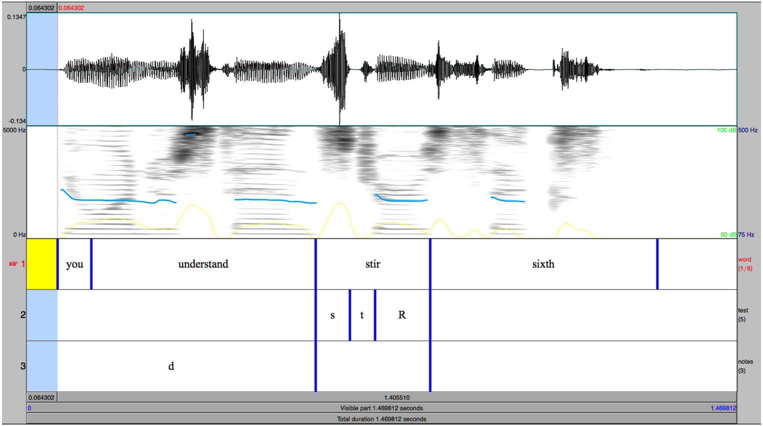
Talker	skur
f101	everyone attempt skur tenth
m102	someone record skur entirely
f103	everyone distinguish skur sometime
m104	noone remember skur third
f105	noone study skur neatly
f106	someone write skur precisely
m107	someone imagine skur precisely
f108	noone write skur second
f109	someone show skur fifth
m110	we imagine skur gladly
m111	I report skur nicely
m112	I think skur first
f113	you study skur daily
m114	everyone describe skur monthly
m115	noone echo skur today
m116	I repeat skur surely
m117	they distinguish skur wisely
m118	someone say skur fifth
f119	we sense skur twice
m120	people speak skur eighth

Articulation Index Corpus

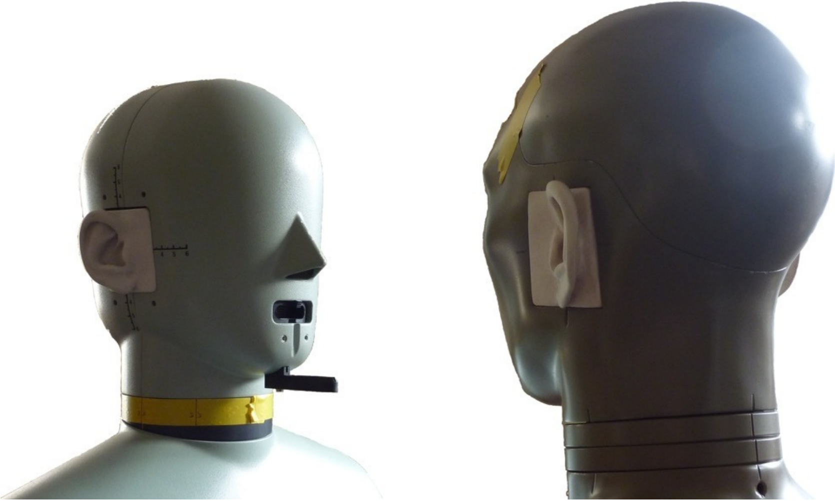
Talker	spur
f101	I use spur fluently
m102	everyone perceive spur properly
f103	we think spur fourth
m104	people ponder spur nicely
f105	people saw spur nicely
f106	we note spur properly
m107	they watch spur only
f108	I distinguish spur usually
f109	someone remember spur easily
m110	someone repeat spur anyway
m111	everyone propose spur happily
m112	they think spur entirely
f113	noone hear spur monthly
m114	we speak spur surely
m115	people echo spur ninth
m116	everyone thinks spur fluently
m117	anyone prompt spur easily
m118	they speak spur seventh
f119	someone witness spur now
m120	noone watch spur happily

Talker	stir
f101	noone check stir eighth
m102	people determine stir ninth
f103	they imagine stir surely
m104	we determine stir surely
f105	they review stir gladly
f106	people saw stir steadily
m107	I remember stir surely
f108	I use stir neatly
f109	I use stir wisely
m110	we view stir ninth
m111	people ponder stir second
m112	I evoke stir precisely
f113	I read stir second
m114	they said stir wisely
m115	I echo stir precisely
m116	noone report stir well
m117	everyone view stir neatly
m118	I imagine stir daily
f119	you understand stir sixth
m120	they sense stir gladly

Phonetic transcription



Convolution with Reading BRIRs for L-shaped room



Partitioning (cutoff)

- Each AIC utterance presented once only to each listener
- 20 conditions are tested
  4 reverb distances ×5 filter cutoffs
- 1600 stimuli partitioned between 20 listeners
- 1 listener gets 80 utterances
  4 utterances at each of 20 conditions
- Even partitioning
  1 word tested at each of 20 conditions

Presentation (cutoff)

- Listener seated in a sound-attenuating booth
- Monaural presentation (left ear)
- Familiarisation with interface
  4 buttons, labelled {sir, skur, spur, stir}
- Click one button for each trial heard
- 1 group of 20 listeners
  age 20-50, both native-English and non

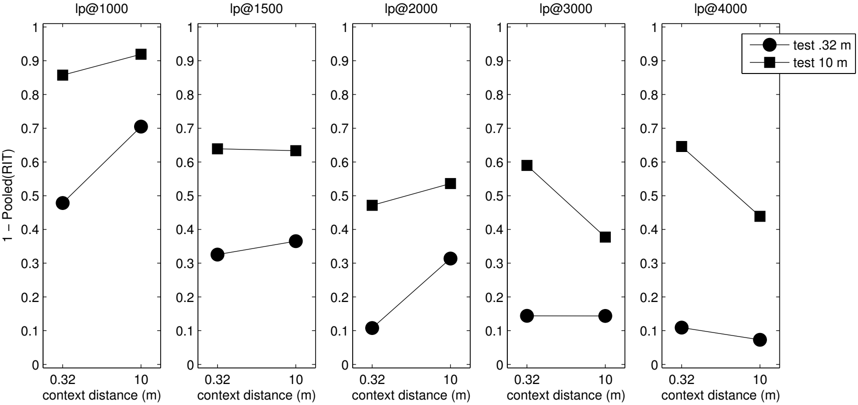
Partitioning (reverse)

- Each AIC utterance presented once only to each listener
- 16 conditions are tested
  4 reverb distances×4 preceding context directions
- 1280 stimuli partitioned between 16 listeners
- 1 listener gets 80 utterances
  5 utterances at each of 16 conditions
- Uneven partitioning
  3 words tested once in 16 conditions, 1 word tested twice

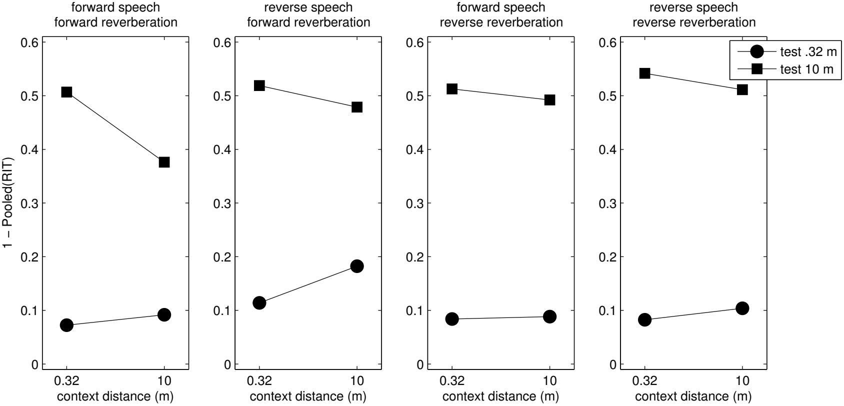
Presentation (reverse)

- Listener seated in a sound-attenuating booth
- Monaural presentation (left ear)
- Familiarisation with interface
  4 buttons, labelled {sir, skur, spur, stir}
- Click one button for each trial heard
- 48 subjects = 3 groups of 16 listeners
  age 20-50, both native-English and non

Cutoff results iii. Pooled (RIT)



Reverse results iii. Pooled (RIT)


forward speech
forward reverberation
reverse speech
forward reverberation
forward speech
reverse reverberation
reverse speech
reverse reverberation
1 – Pooled(RIT)
context distance (m)
test .32 m
test 10 m