Speech recognition in *N*-talker babble: Patterns of performance with increasing *N* vary across types of speech material

Tim Green and Stuart Rosen

Speech, Hearing & Phonetic Sciences, UCL

Funded by the MRC

Speech recognition amongst competing voices

Speech must often be perceived against a background of other voices

- particularly challenging for hearing-impaired listeners

A clearer understanding of what contributes to normal-hearing performance is likely to benefit technological developments to help the hearing-impaired

Includes isolation of contribution of features such as periodicity (P3 Steinmetzger & Rosen)

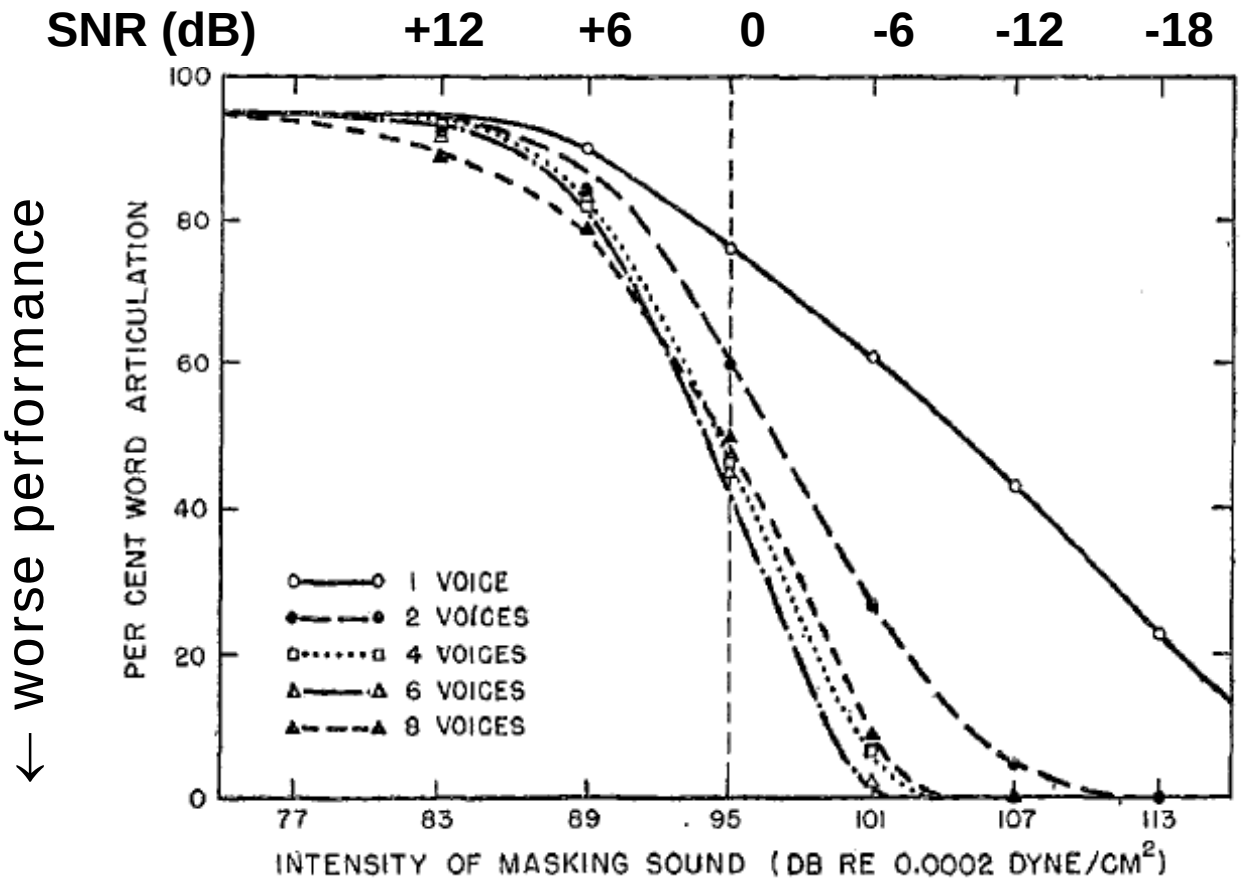
Focus here on how performance changes with different numbers of competing talkers (*N*)

- expected to provide basis for testing ideas about factors contributing to masker effectiveness

Miller (1947)

Recognition of single words from male target talkers at various SNRs

Maskers: equal numbers of male and female voices (**1 VOICE** is male)



'It is relatively easy for a listener to distinguish between two voices, but as the number of rival voices is increased the desired speech is lost in the general jabber.'

Types of masking

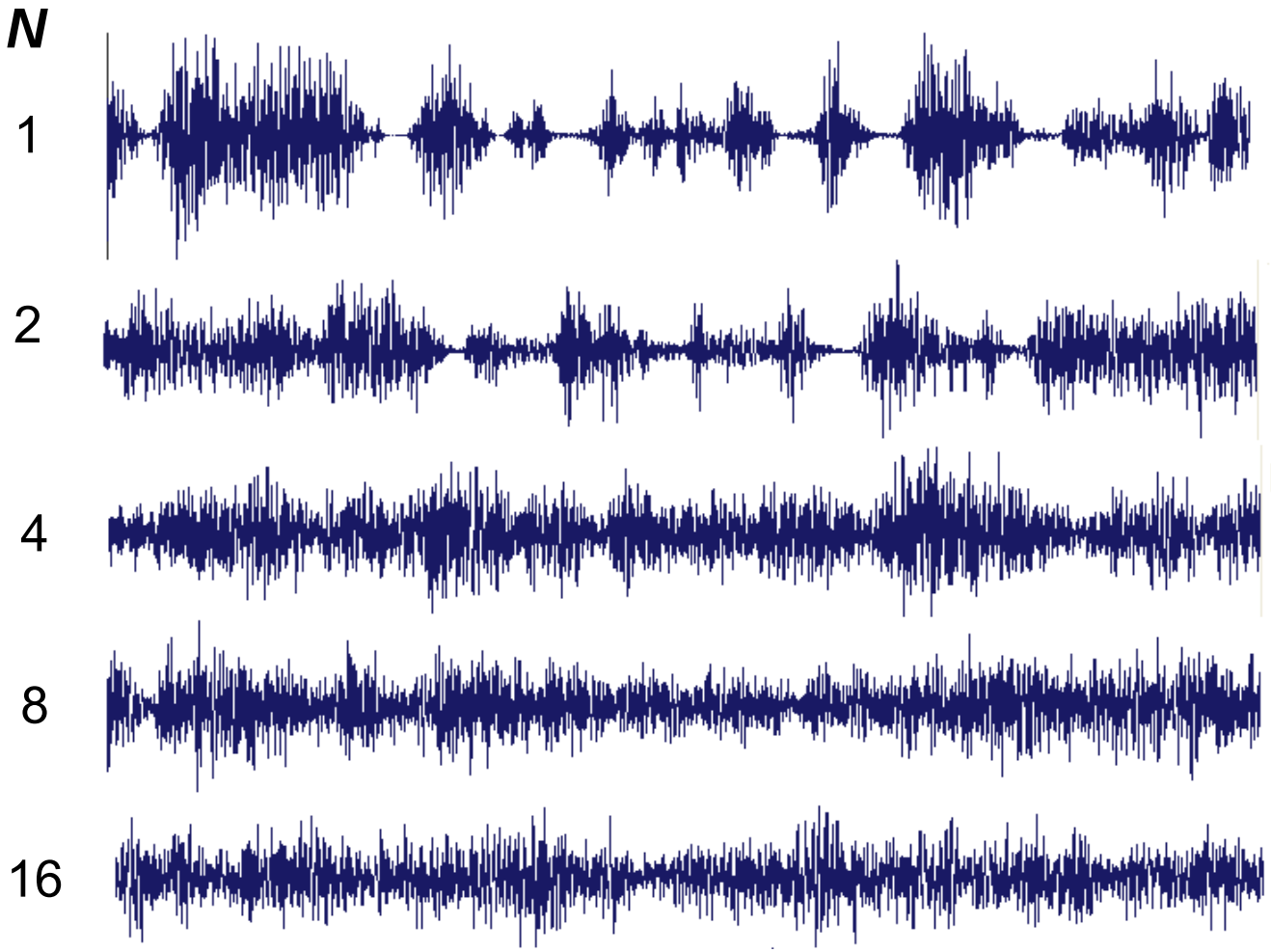
Energetic masking

- primarily reflects interaction of target and masker in the cochlea
- maskers interfere with target to the extent that they have energy in the same time/frequency regions
- modulations in the masker can be beneficial
- temporal and/or spectral ‘dips’ in the masker allow ‘glimpses’ of the target speech

Informational masking

- interactions between target and masker at central levels
- similarity between target and masker is important
- problems in object formation (auditory scene analysis) and/or
- object selection (attention and distraction) (Shinn-Cunningham, 2008)

Masking effectiveness and number of talkers



As *N* increases …

Spectral and temporal dips in the babble are filled in

Opportunities for glimpsing reduced

Energetic masking expected to increase monotonically with *N*

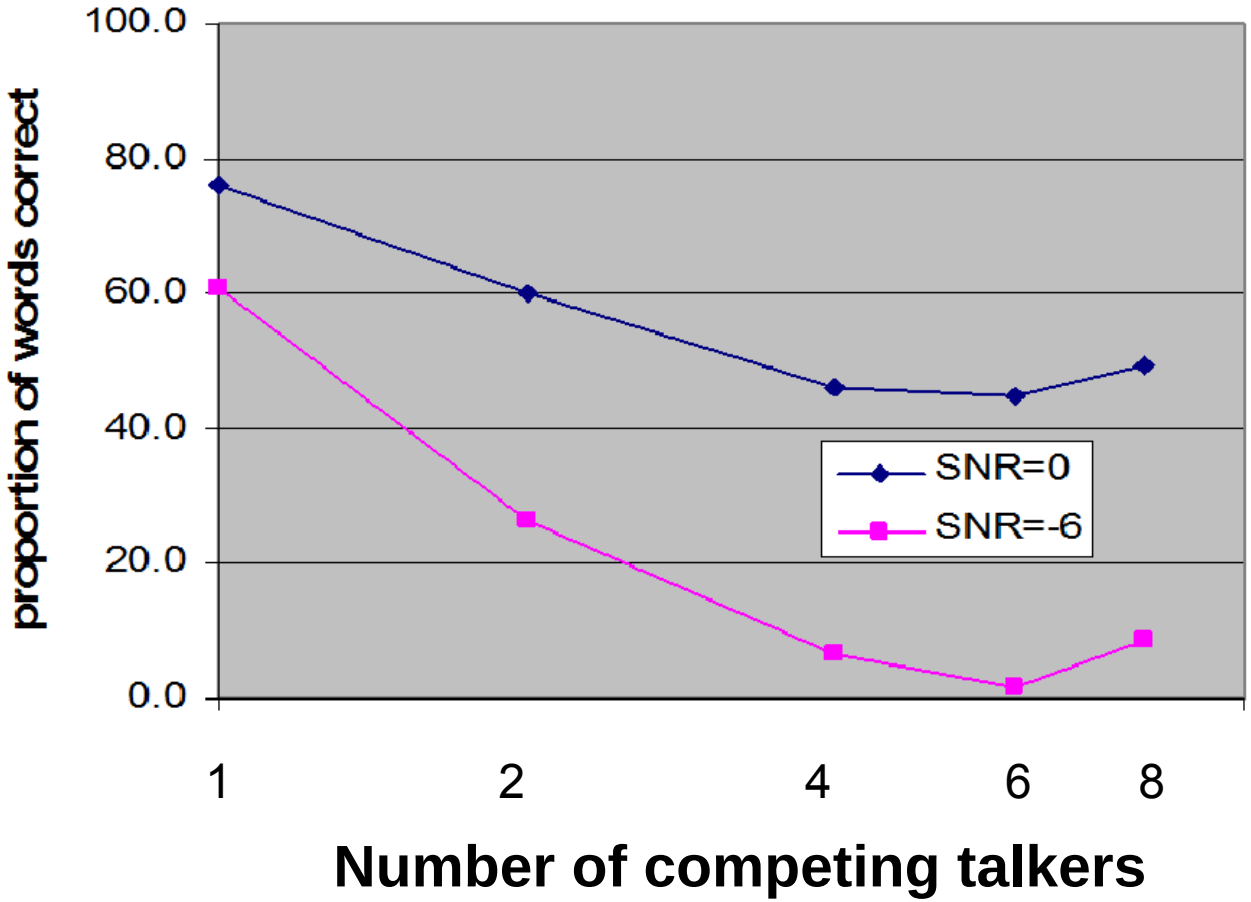
Masking effectiveness and number of talkers

Informational masking

- Initial increase beyond *N* = 1 may exacerbate difficulties in segregating target from masker and selecting target speech
- But at larger values of *N*, individual competing speech streams will become imperceptible…
- similarity between target and masker will be reduced
- distracting effect of masker will be reduced
- **Informational masking** will change in a non-monotonic way as *N* increases

Previous evidence of performance change with *N*

Miller's (1947) word recognition data re-plotted to show change with *N* for fixed SNR

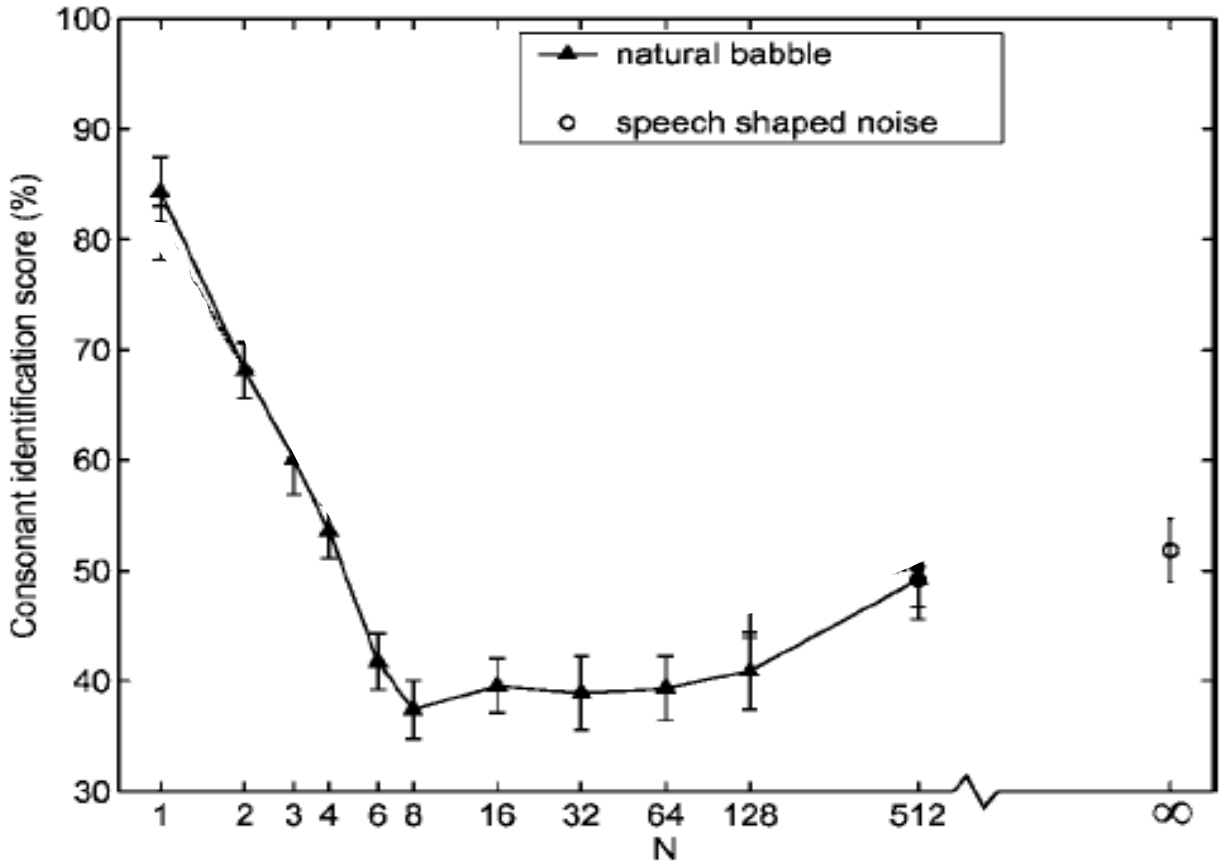


Male target talker

Male and female competitors

Simpson & Cooke (2005) - Consonant Identification

VCVs Male target and maskers SNR -6 dB



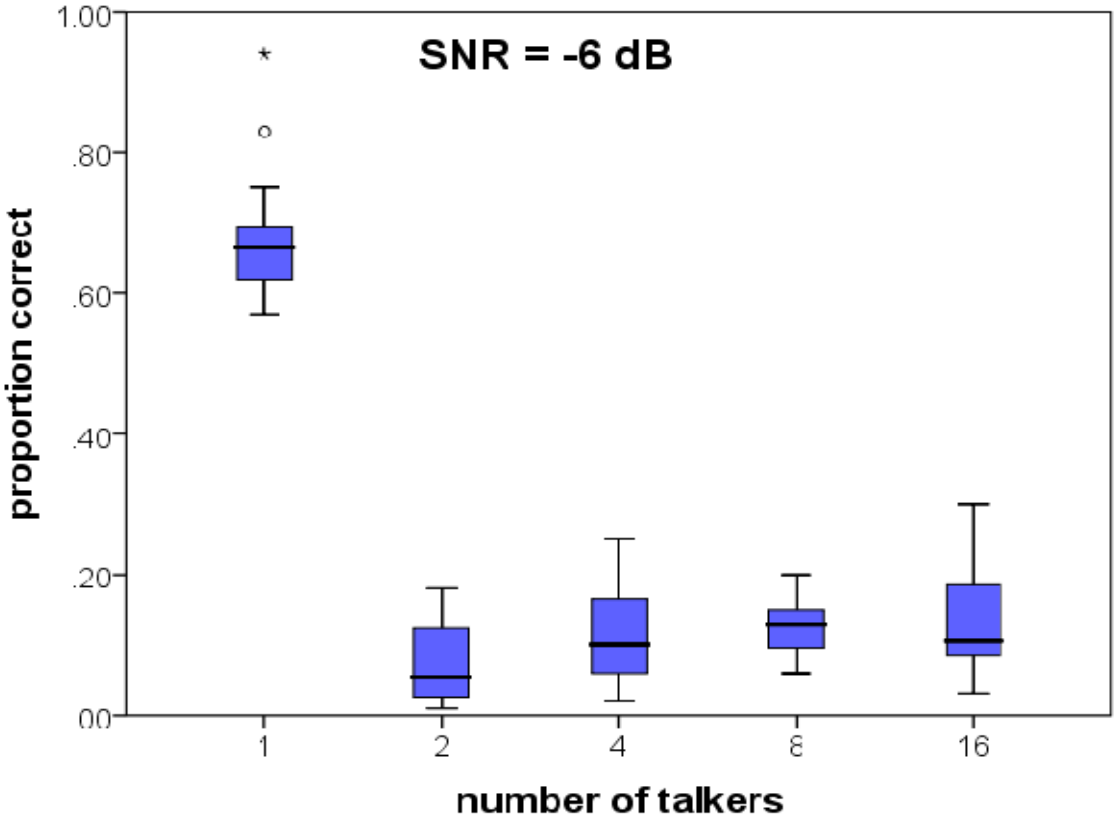
Relatively steep initial decline in performance

Minimum performance at $N = 8$

Shallow increase in performance as N increases further

Rosen et al (2013) – IEEE sentences

Male target and maskers SNR -6 dB

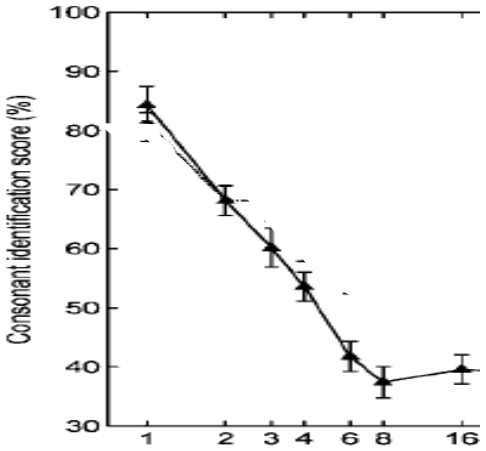


Very steep decline in performance as *N* increases from 1 to 2

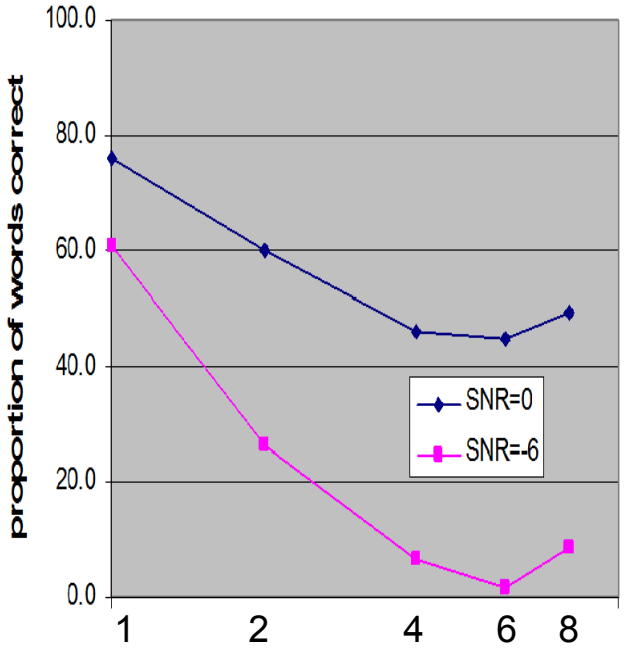
Minimum performance at *N* = 2

Slight increase for higher *N*

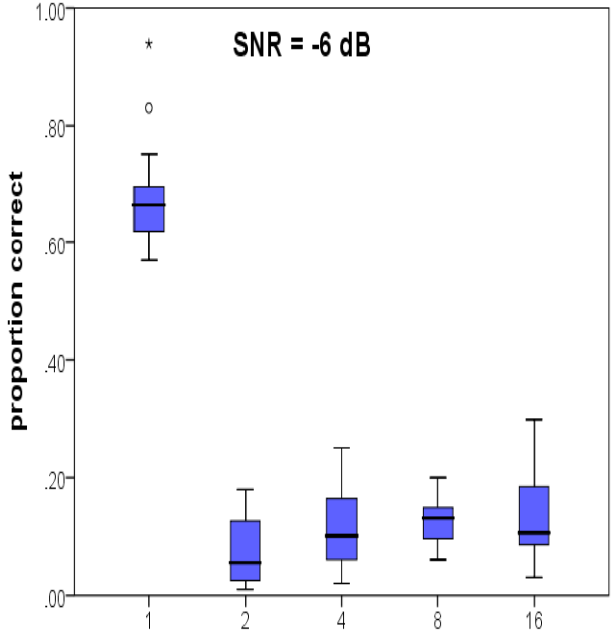
VCV



Word



IEEE



In all cases single talker is least effective masker

Performance initially declines with increasing *N*, then plateaus or increases slightly

Value of *N* at which performance reaches minimum differs across studies

The previous data suggest that patterns of performance with increasing *N* vary according to target material

- briefer glimpses available with larger *N* may be sufficient for VCV identification, while not supporting sentence recognition

- difficulties in segregation may be greater when both target and masker are connected speech

However, other differences across studies make it difficult to evaluate the effect of target material

In addition, effects of sex differences between target and masker have not been examined

The present study

Compare how performance changes with *N* for different types of speech material recorded from the same talkers (one male, one female)

How much do patterns vary according to whether the sex of the maskers matches the target?

- energetic masking only slightly affected by target-masker sex differences (Brungart et al 2009)
- informational masking may be substantially affected

Methods

Target speech:

BKB sentences - 28 per condition

IEEE sentences - 20 per condition

Monosyllabic Words - phonetically balanced lists, 40 per condition

VCVs - 16 consonants in 3 different vowel contexts

Competing speech:

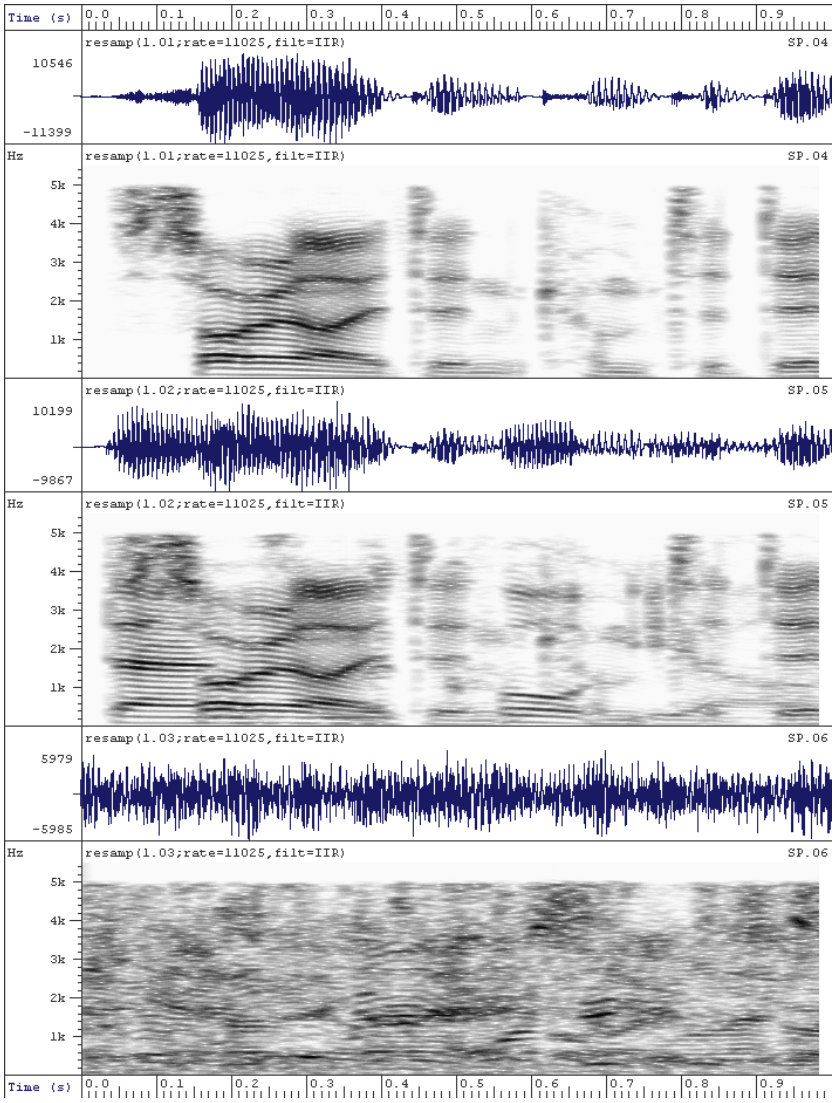
Derived from EUROM database of connected passages (pauses>100 ms removed)

N = 1, 2, 4, 8, 16 (all male or all female)

Example maskers – male talkers

waveform

spectrogram



N = 1

N = 2

N = 16

Methods

Listeners

8 different groups of 10 normally-hearing young adults

One group for each combination of target talker and target material

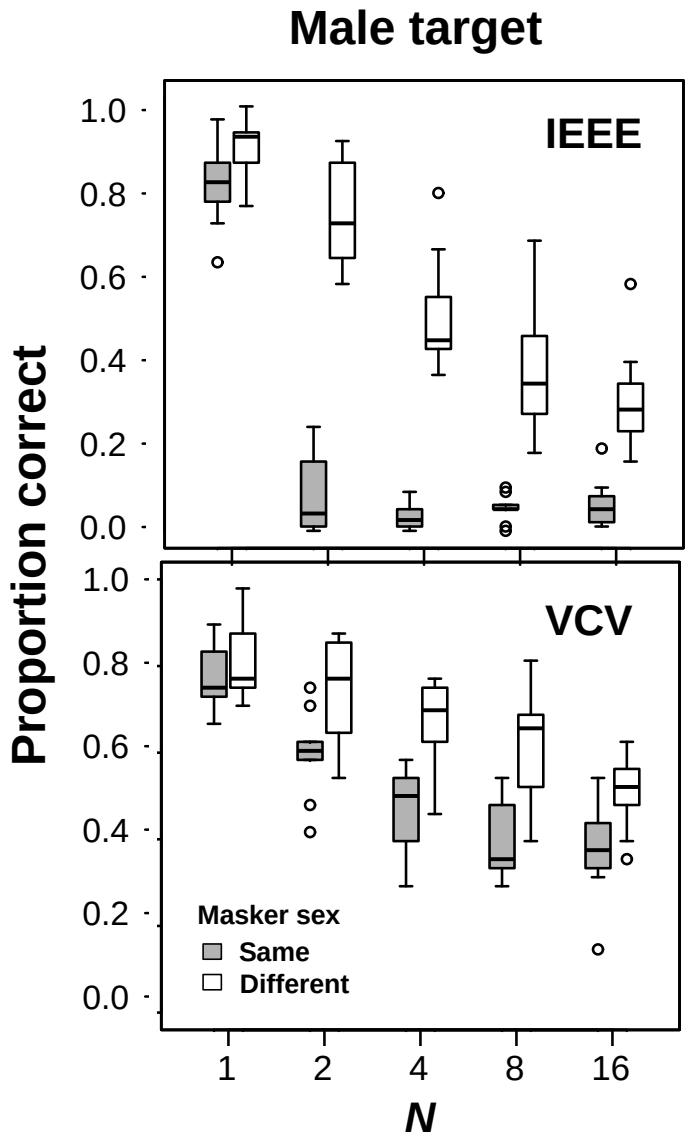
10 conditions per listener (5 values of *N*, both M and F maskers)

Order of conditions counterbalanced using Randomised Latin Squares

SNRs

Male Talker	Sentences	-6 dB	Words & VCVs	-4 dB
Female Talker	Sentences	-4 dB	Words & VCVs	-2 dB

Data analysis

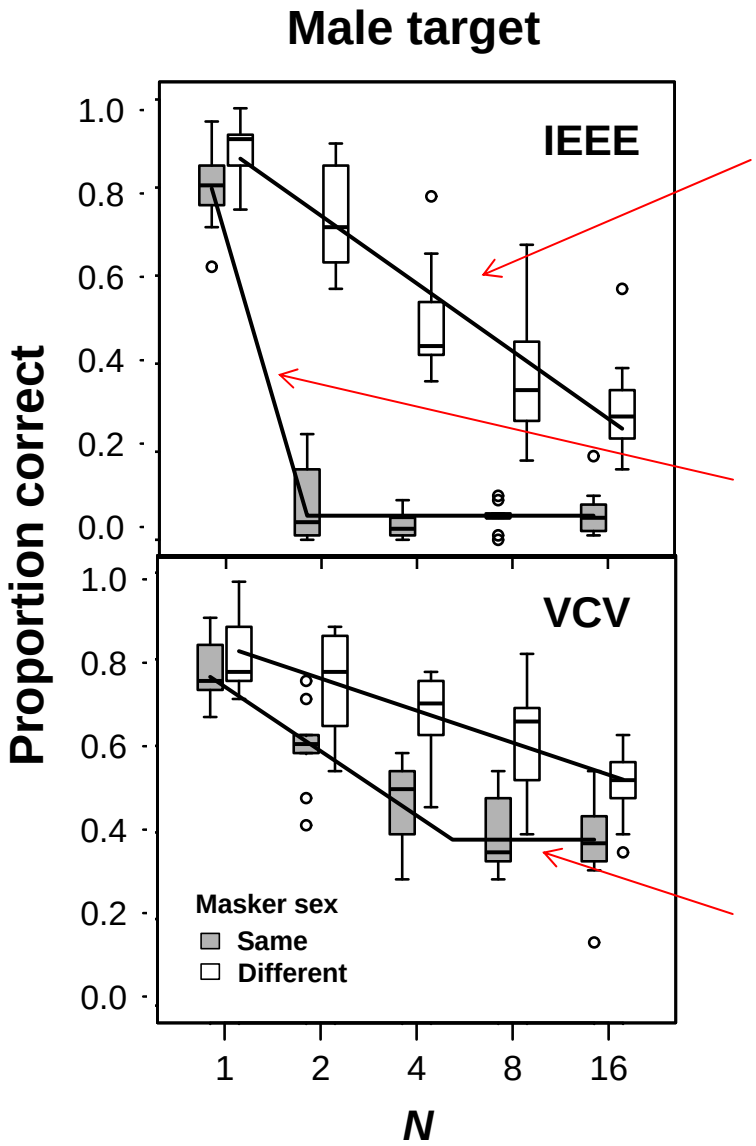


Patterns characterised using segmented regression (Ritz and Streibig, 2008)

Fits data with 2 straight lines of arbitrary slope that meet at a 'breakpoint'

NB. Boxes offset along x-axis for clarity

Data analysis



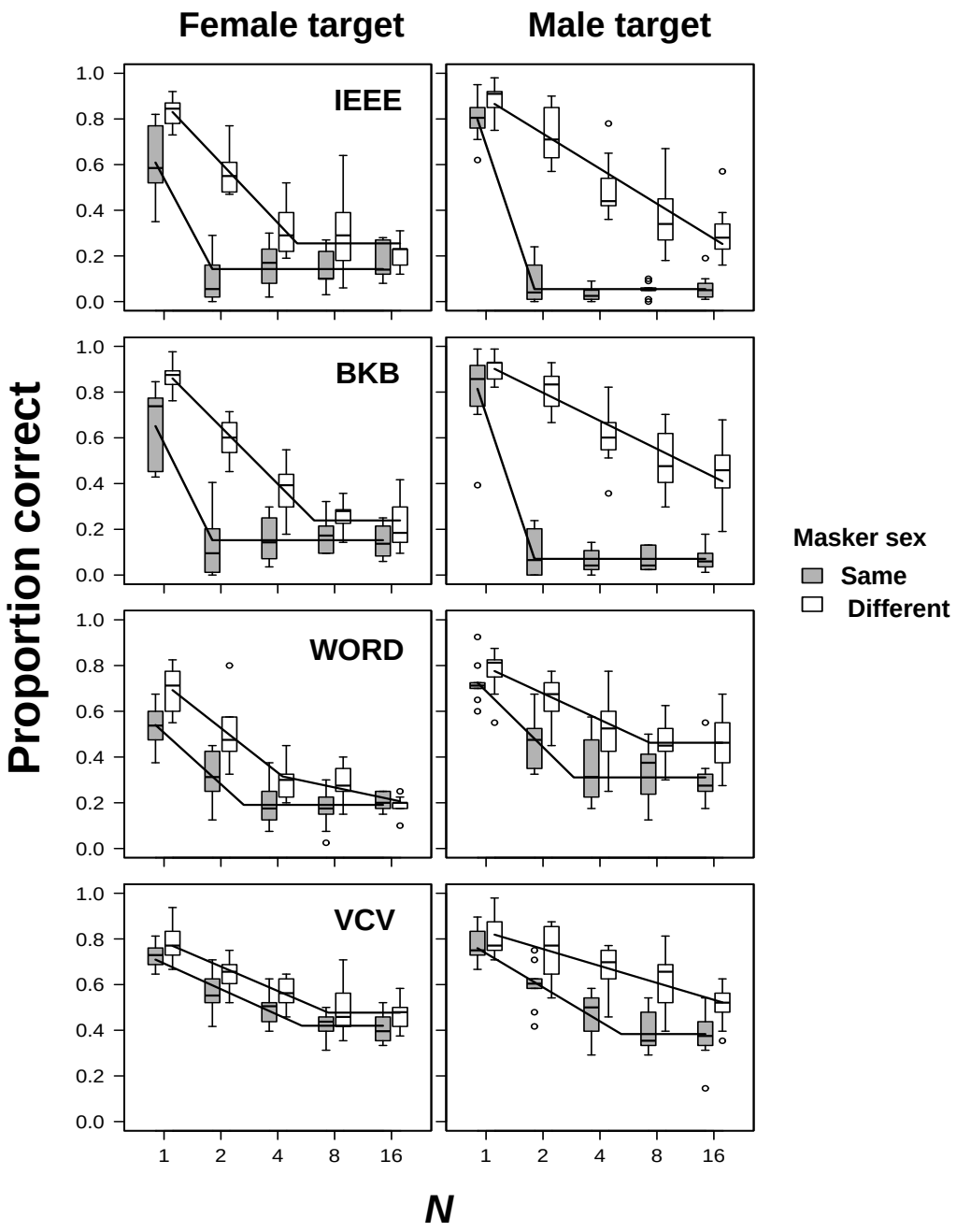
In some case the data were fit equally as well by a single line

Where the initial decline is very steep the fit may be ‘degenerate’

-any breakpoint between 1 and 2 provides an equally good fit

-here, breakpoint is set to 2

If slope after the breakpoint did not differ significantly from zero, it is set to zero



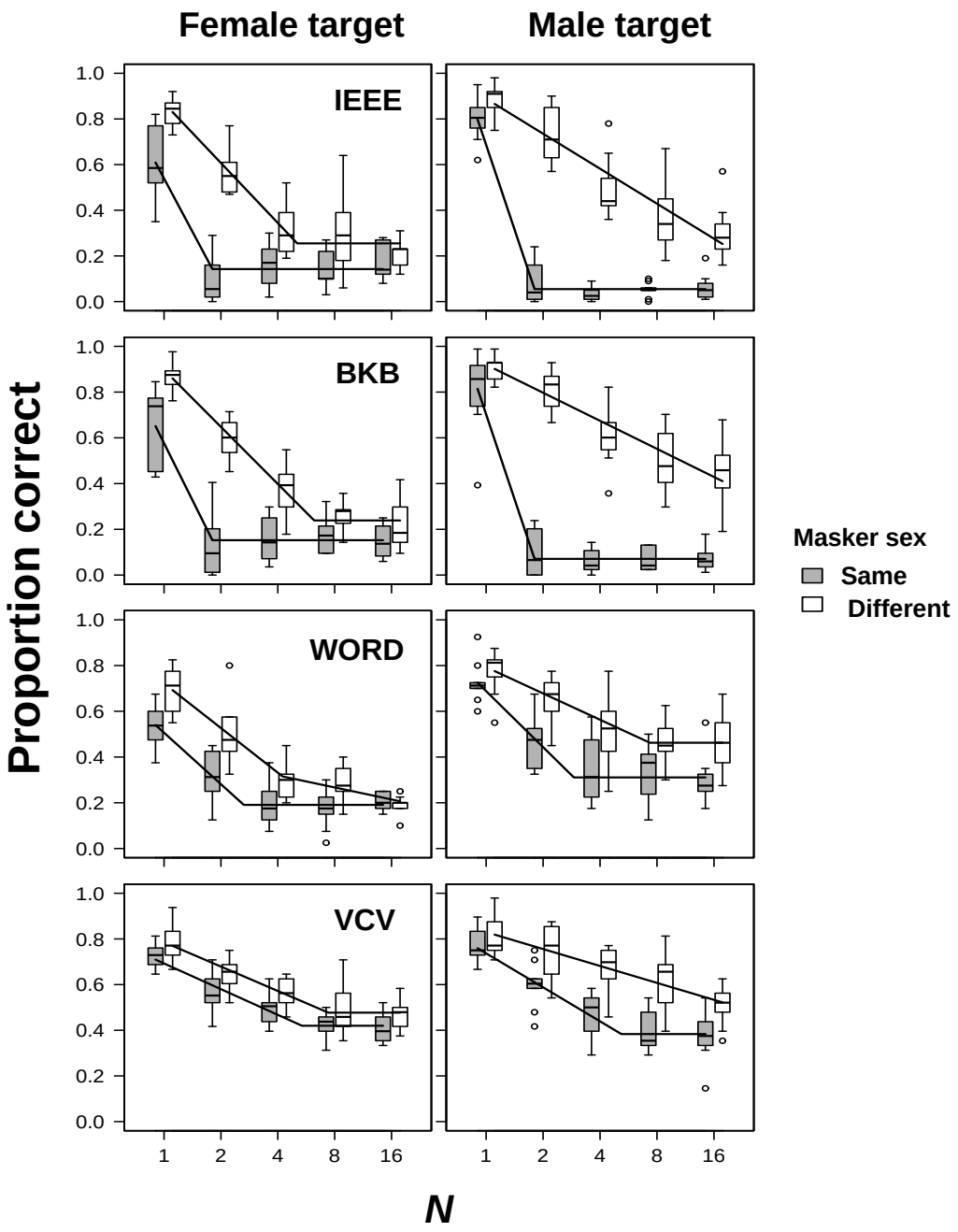
Typically single line fits for male targets with female maskers

In other cases, the slope after the breakpoint is typically zero

Substantial differences in location of breakpoints, particularly for same sex maskers

- steeper decline and plateau at smaller N for sentences than VCV and WORD

Differences across material less pronounced for different sex maskers



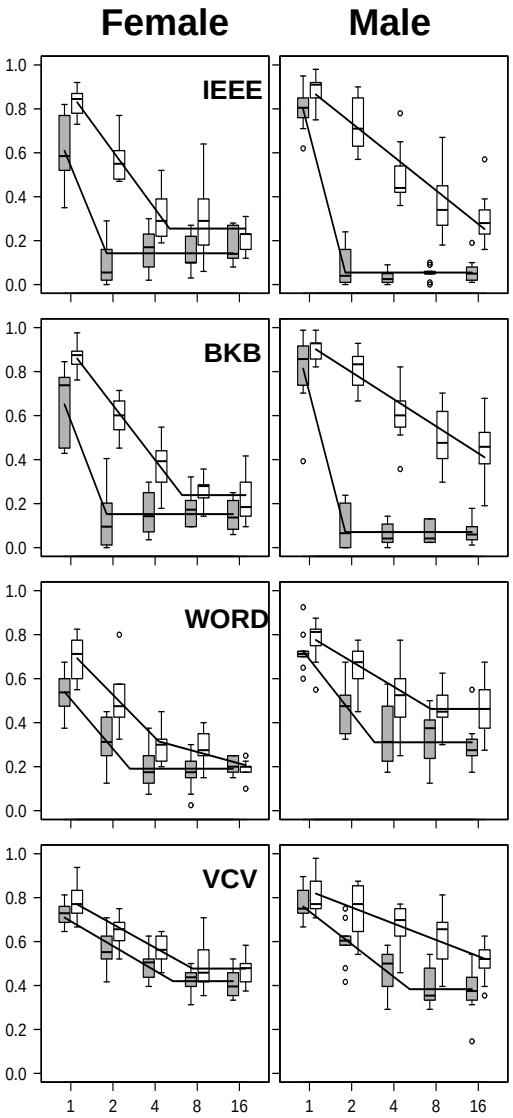
Very similar patterns for IEEE and BKB

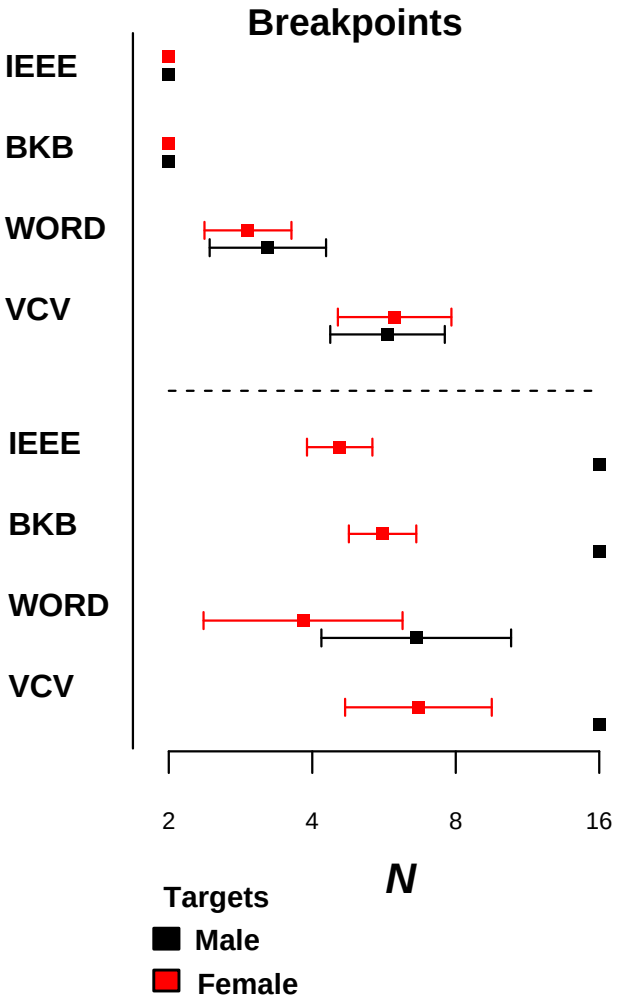
Performance generally very similar for male and female target talkers

Performance always better with different-sex maskers, but size of the advantage varies considerably across *N* and target material

Differences due to sex of masker particularly striking for sentences

Effect of target materials on breakpoints





Same sex maskers

Very similar pattern for male and female targets

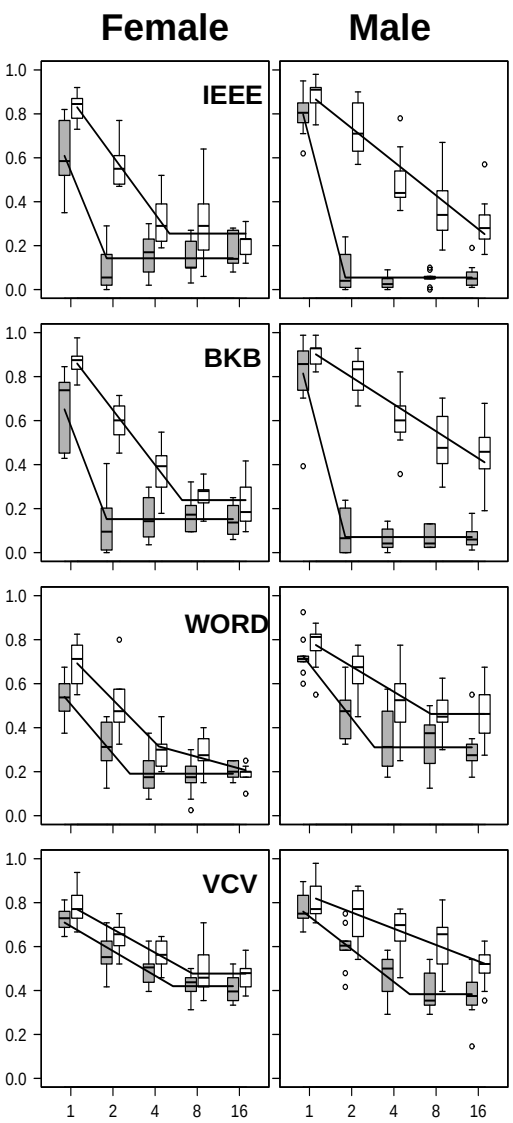
Strong effect of test material on breakpoint

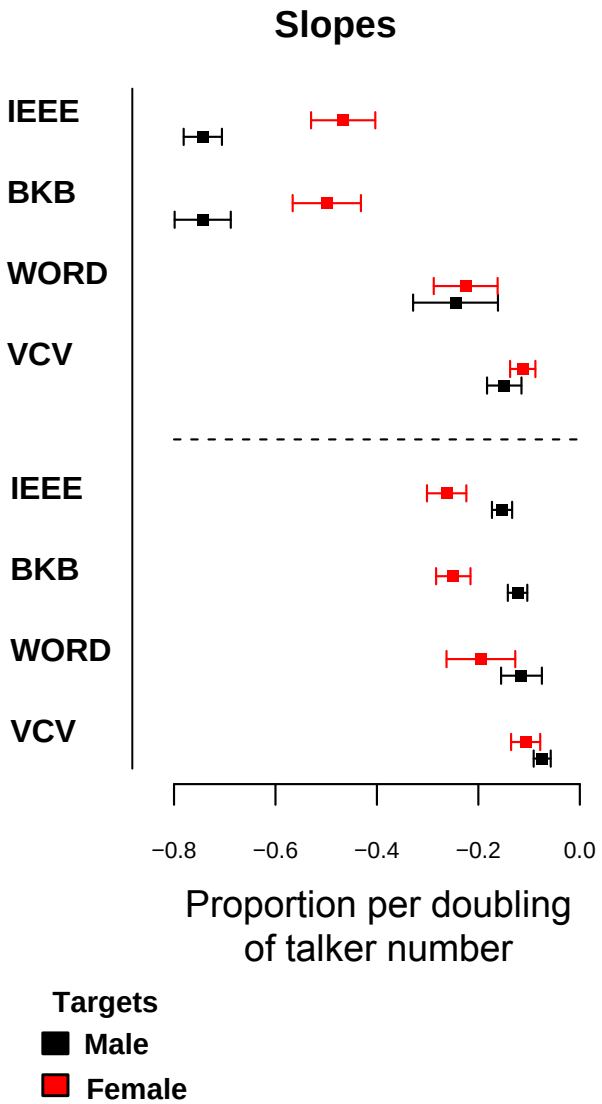
Different sex maskers

Only weak effects of target material

Differences between target talkers

Effect of target materials on slopes





Strong effects of material on initial slopes for same sex maskers

Little effect for different sex maskers

Summary

Clear effect of target material on pattern of change with increasing *N*

Similar to previous studies, with same-sex maskers sentence recognition declined precipitously when *N* changed from 1 to 2, but word and VCV identification began to plateau only at higher values of *N*

With different-sex maskers, performance was more similar across target material, with breakpoints always greater than 4

Suggests a particularly prominent role for informational masking in sentence recognition with same sex maskers when *N* increases beyond 1

Sentence recognition with one male and one female competitor

Large differences in sentence recognition with N = 2 according to whether competitors were same- or different-sex

A follow-up experiment compared these situations with a masker comprising one male and one female voice

Methods

IEEE sentences

Different groups of 10 normally hearing adults tested with either male or female target talker

5 masking conditions:

1 same-sex talker (1S), 2 same-sex talkers (2S),

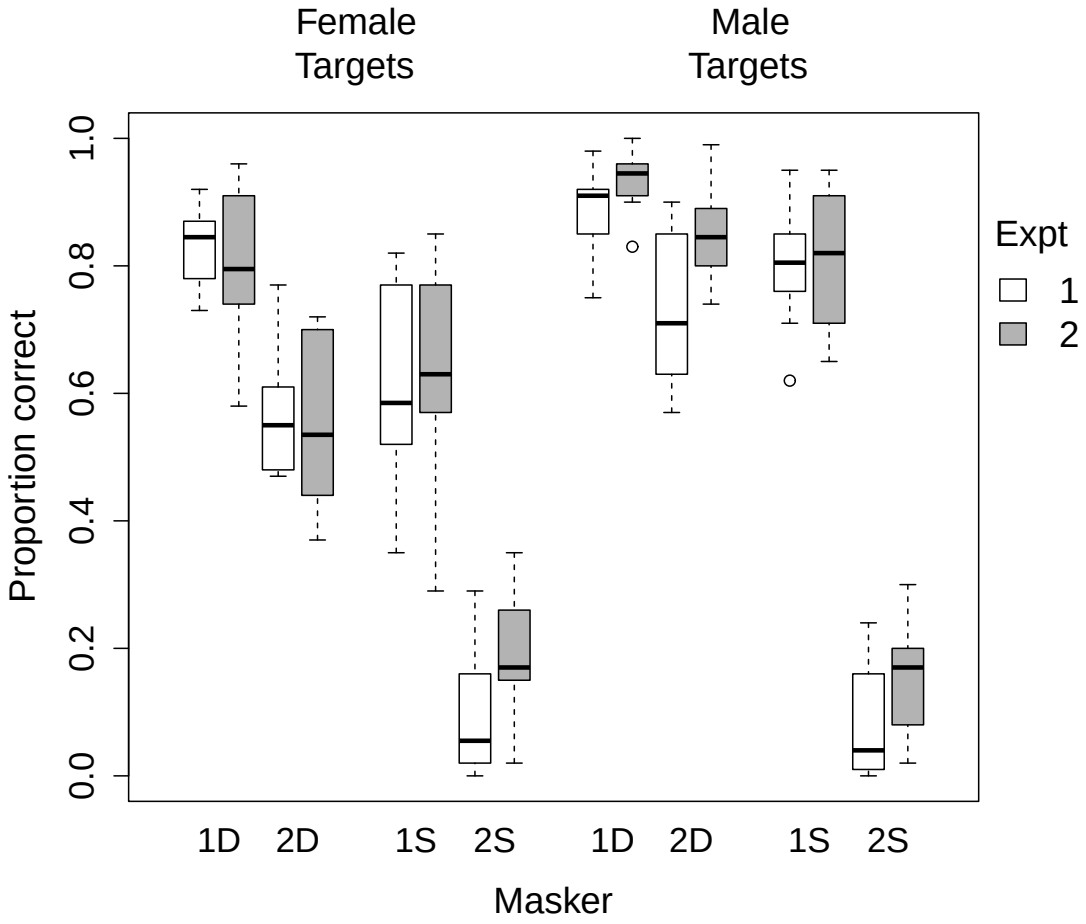
1 different-sex talker (1D), 2 different-sex talkers (2D)

1 same 1 different (SD)

SNRs: Male target -4 dB Female target -2 dB

Unlike in first expt, competing talkers were selected at random, without replacement, on each trial

Comparison of similar conditions across experiments

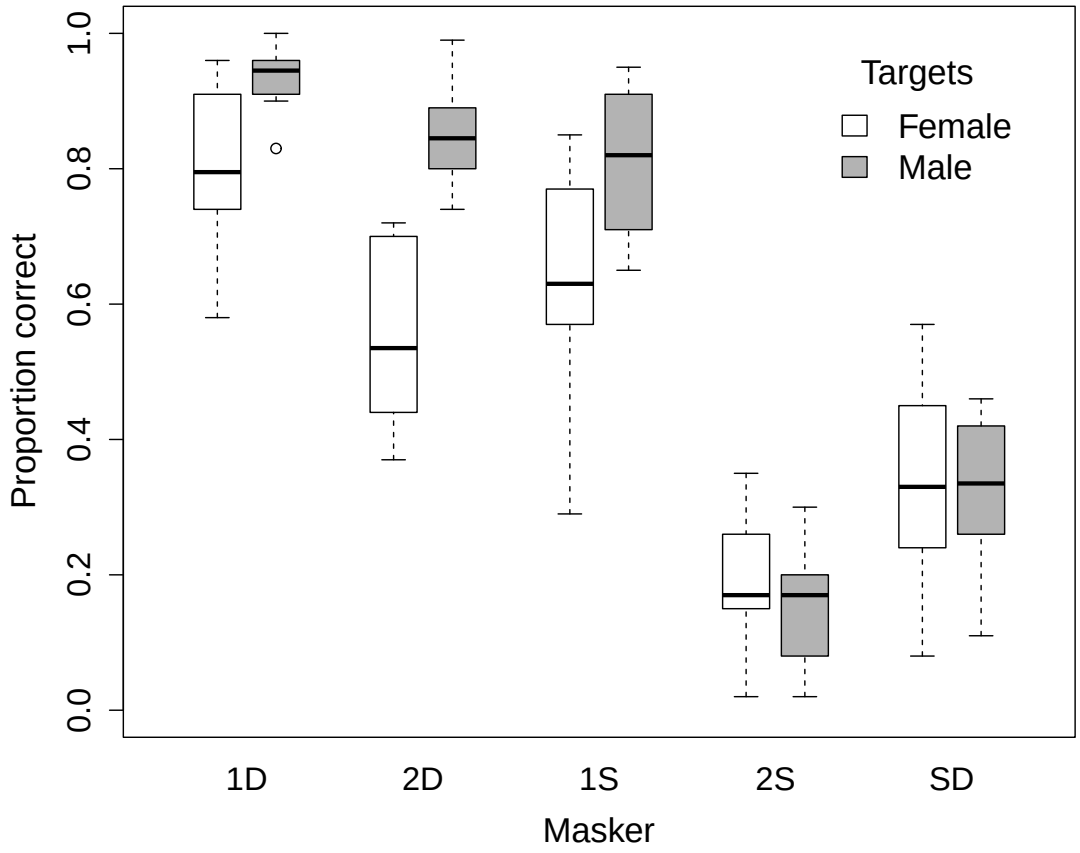


Performance generally slightly higher in expt 2, consistent with small increase in SNR

Pattern of results similar across expts

Suggests little effect of trial-to-trial variation of competing talkers

Comparison with Same-Different condition



Performance in the SD condition falls in between that in the 2D and 2S conditions and is significantly different to both

Conclusions

Type of target material and target-masker similarity interact in determining the pattern of change in speech recognition with increasing numbers of competing talkers

Likely to primarily reflect differences in informational masking

Sentence recognition severely affected by presence of two similar competing talkers

May indicate difficulties in segregating talkers when multiple similar F0 contours are present

Example test stimuli

Male target talker

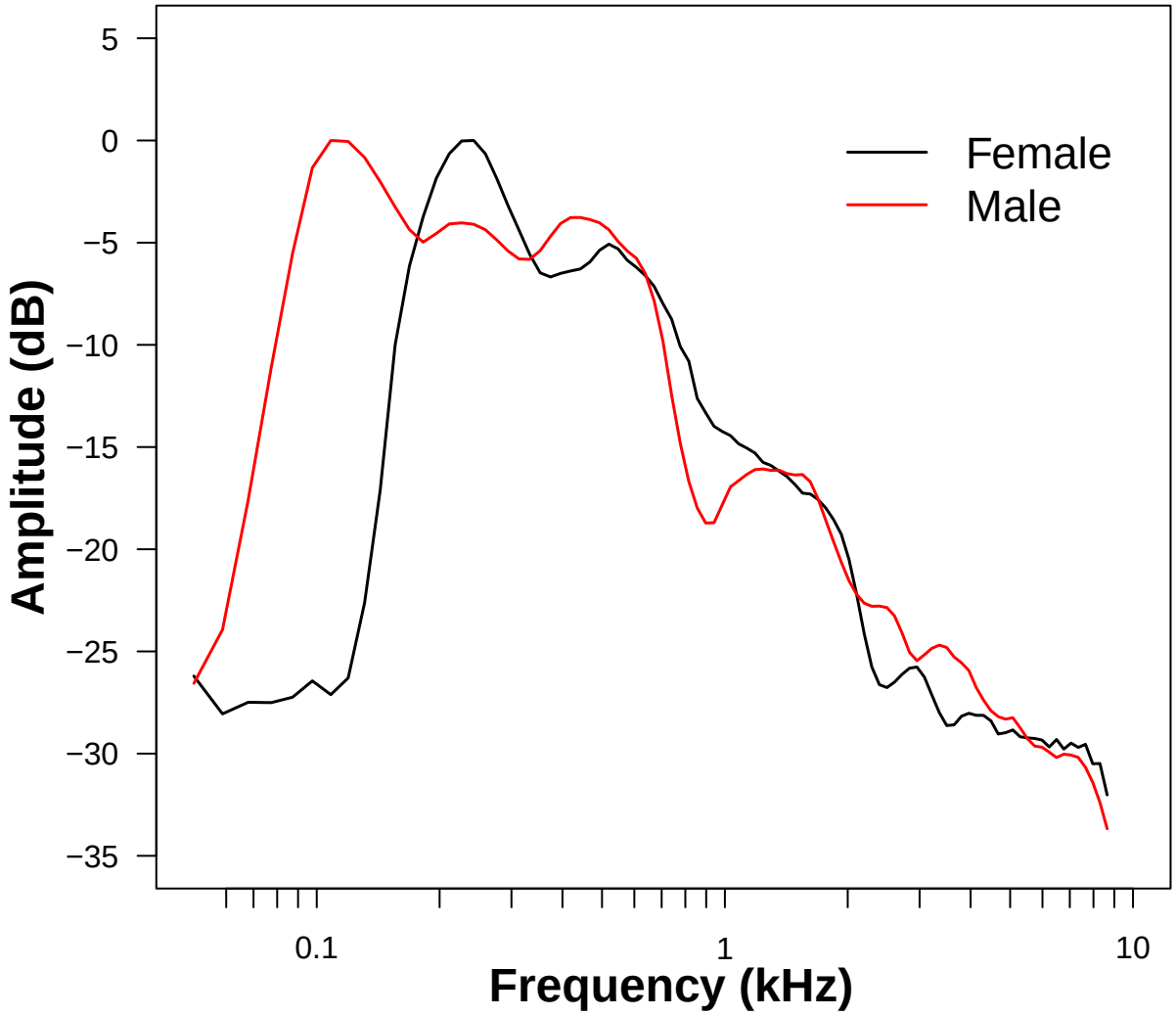
	IEEE		VCV	
Masker Sex	Same	Different	Same	Different
N = 1				
N = 2				
N =16				

Example test stimuli

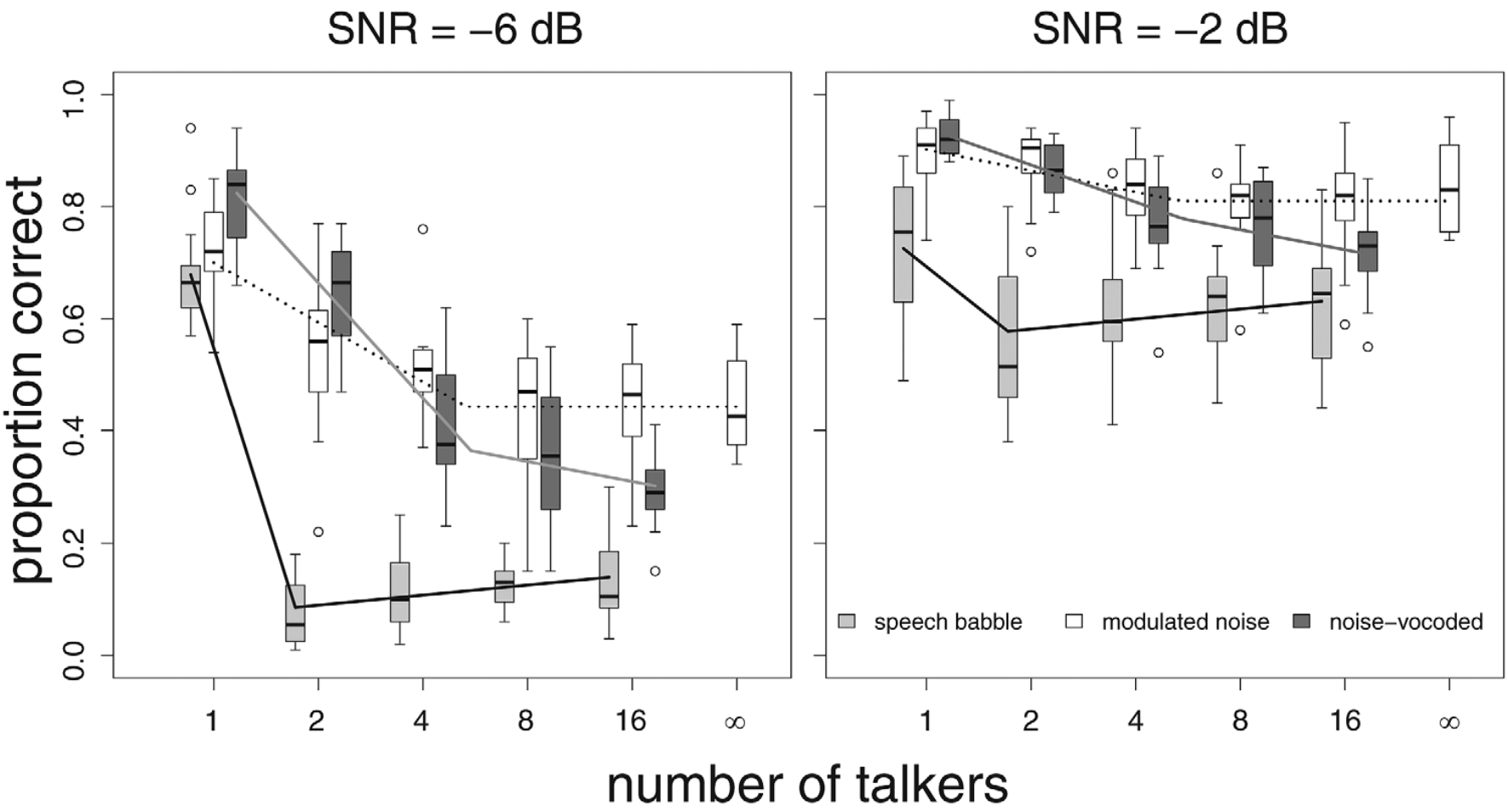
Female target talker

	IEEE		VCV	
Masker Sex	Same	Different	Same	Different
N = 1				
N = 2				
N =16				

Average long-term spectra of male and female 16-talker babbles



Rosen et al (2013) – Effect of SNR on initial slope



Rosen et al (2013) aggregate psychometric functions