



1. Introduction

Humans are remarkably good at listening to a single speaker in a multispeaker environment. This ability has been named the “cocktail party effect” (Cherry, 1953).

Both spatial cues (Arbogast *et al.*, 2002; Hawley *et al.*, 2004) and speaker size cues (Brungart, 2001) have been shown to have a significant effect on the recognition of one speaker in a multispeaker environments.

This study investigates the interaction between spatial cues [expressed as interaural time differences (ITDs)] and speaker size cues [expressed as speaker size differences] in multispeaker environments.

2.Experimental procedure

Participants were presented with two phrases of concurrent speech syllables and asked to recognize syllables from one of the phrases. The task was made more or less difficult by changing the ITD, the size difference and/or the level between the two talkers.

Stimuli

The speech syllables were taken from a large speech database (Ives *et al*, 2005).

The speech was recorded from one speaker, pitch normalised, perceptual-centre corrected, level normalised. Speaker size was controlled using STRAIGHT (Kawahara and Irino, 2004).

	b	d	f	g	h	k	l	m	n	p	r	s	sh	t	v	w	y	z	~
a fa, la	ba	da	fa	ga	ha	ka	la	ma	na	pa	ra	sa	sha	ta	va	wa	ya	za	
e re	be	de	fe	ge	he	ke	le	me	ne	pe	re	se	she	te	ve	we	ye	ze	
i mi, ti	bi	di	fi	gi	hi	ki	li	mi	ni	pi	ri	si	shi	ti	vi	wi	yi	zi	
o do, so	bo	do	fo	go	ho	ko	lo	mo	no	po	ro	so	sho	to	vo	wo	yo	zo	
u tofu	bu	du	fu	gu	hu	ku	lu	mu	nu	pu	ru	su	shu	tu	vu	wu	yu	zu	
a fa, la	ab	ad	af	ag	ah	ak	al	am	an	ap	ar	as	ash	at	av	aw	ay	az	
e re	eb	ed	ef	eg	eh	ek	el	em	en	ep	er	es	esh	et	ev	ew	ey	ez	
i mi, ti	ib	id	if	ig	ih	ik	il	im	in	ip	ir	is	ish	it	iv	iw	iy	iz	
o do, so	ob	od	of	og	oh	ok	ol	om	on	op	or	os	osh	ot	ov	ow	oy	oz	
u tofu	ub	ud	uf	ug	uh	uk	ul	um	un	up	ur	us	ush	ut	uv	uw	uy	uz	

Syllable phrases

There are two phrases: the **target** and the **masker**.

The target phrase contains three syllables and the masker phrase contains two syllables. The masker syllables coincide with the second and third target syllables.

Task: identify either the second or third target syllable (chosen randomly).

Envelope matching

Coincident syllables across the two phrases have matched temporal envelopes, the masker is drawn from the same syllable group but contains a different consonant and vowel.

Parameters measured

Target to masker ratio (TMR): 15, 9, 3, 0, -3, -9 and -15 dB;

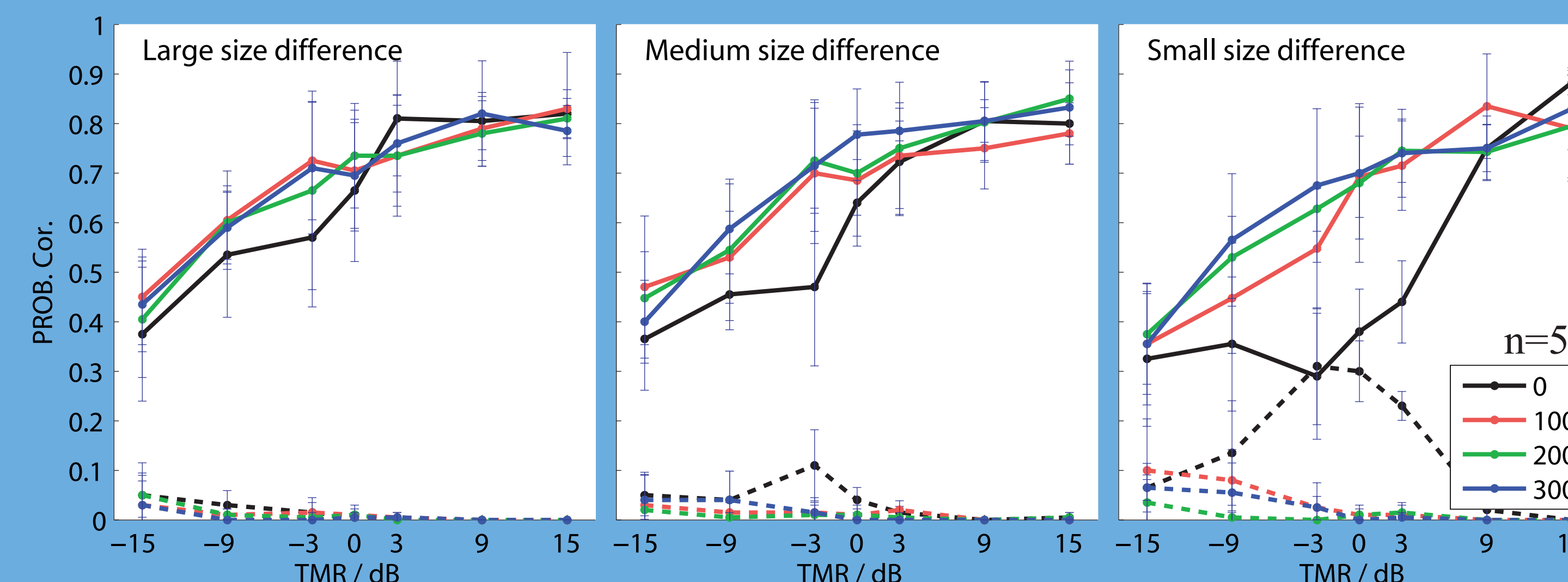
Interaural time differences (ITD): 0, 100, 200 and 300 μ s;

Speaker size for masker voice: small, medium and large difference;



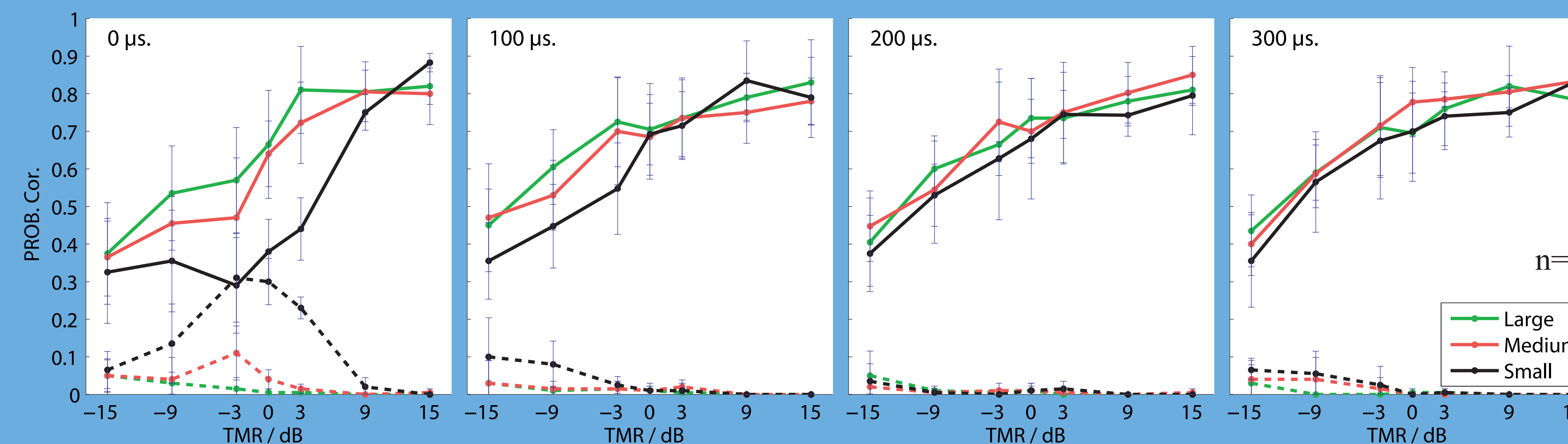
3.Results

The effect of ITD on performance for different size differences, showing ITD affects performance more for smaller size differences.



Solid lines show the probability of a correct response. Dashed lines show the probability of choosing the masker syllable.

The effect of speaker size differences on performance for different ITDs, showing speaker size difference affects performance more for smaller ITDs.



4. Conclusions

ITD differences and size differences do interact in the production of a release from masking.

This interaction occurs when ITD and speaker size cues are small, such that their individual contribution would be insufficient to produce a large masking release on their own.

Increasing either the ITD or speaker size, reduces the interaction, as the increased cue comes to dominate.

The effect of size difference and ITD on performance, and the interaction between them, is strongly dependent on TMR.

Acknowledgements

Research supported by the U.K. Medical Research Council (G0500221, G9900369).

References

- Arbogast, T. L., Mason, C. R. and Kidd, G., Jr. (2002) “The effect of spatial separation on informational and energetic masking of speech,” J. Acoust. Soc. Am. 107, 1074-1088.
- Brungart, D. S. (2001) “Informational and energetic masking effects in the perception of two simultaneous talkers,” J. Acoust. Soc. Am. 109, 1101-1109.
- Cherry, E. C. (1953) “Some experiments on the recognition of speech, with one and with two ears,” J. Acoust. Soc. Am. 25, 975-979.
- Hawley, M. L., Litovsky, T. Y., and Culling, J. F. (2004) “The benefit of binaural hearing in a cocktail party: Effect of location and type of masker,” J. Acoust. Soc. Am. 115, 833-843.
- Ives, D.T., Smith., D.R.R., and Patterson, R.D. (2005) “Discrimination of speaker size from syllable phrases” J. Acoust. Soc. Am. 118, 3816-3822.
- Kawahara, H., and Irino, T. (2004). “Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation,” in Speech Separation by Humans and Machines, edited by P. Divenyi, Kluwer Academic, Massachusetts, 167–180.