Exploring the effects of masker spectro-temporal coherence on the informational masking of speech

Brian Roberts and Robert J. Summers
(Psychology, School of Life & Health Sciences, Aston University, Birmingham, UK)

b.roberts@aston.ac.uk, r.j.summers@aston.ac.uk

1. Introduction

The F2C paradigm – in which an extraneous formant acts as a competitor by providing an alternative candidate for F2 (F2C) – offers considerable scope for informational masking, owing to the high degree of unpredictability of the frequency contours of the target and competitor formants.

Recent research indicates that the extent of the informational masking caused by an extraneous formant on the intelligibility of a three-formant analogue of a target sentence is strongly influenced by the depth of variation in its frequency contour.

Specifically, scaling the extent of frequency variation around the geometric mean of the competitor's frequency contour from 100% (full scale) down to 0% (constant at the geometric mean) progressively reduces its impact on intelligibility (Roberts et al., 2014; Roberts & Summers, 2015).

The properties of stimuli in the F2C paradigm differ in important ways from those typically used in studies of informational masking. Research of this kind usually employs narrowband targets and broadband maskers, and increases in the extent of frequency variation across time are confounded with decreases in overall spectro-temporal coherence, arising from larger discontinuities between adjacent segments of the stimulus.

For example, maskers (and targets) are often constructed by concatenating a sequence of multiple tone bursts such that frequency variation across time is generated by making an independent constrained-random draw of frequencies for each successive burst (e.g., Kidd et al., 1994). In contrast, our previous studies involved changing the extent of frequency variation in the target and competitor formants while preserving formant-frequency contours with coherent trajectories.

This study explored whether masker impact depends not only on the extent of formant-frequency variation, but also on the spectro-temporal coherence of the interferer. While this coherence is clearly important for phoneme perception, it may not necessarily determine the extent of interference caused by a competitor.

For example, the spectro-temporal coherence of the extraneous formant may not be important if the interference it produces arises non-specifically from an increased cognitive load on the listener (cf. Mattys et al., 2012), rather than from the intrusion of specific competitor properties into the speech percept.

2. General Methods

Stimuli were derived from the first three formants of simple sentences spoken by a British male talker with Received Pronunciation English. The phonemes for each sentence comprised ≤25% involving closures or unvoiced frication.

The overall duration of the frequency and amplitude contours for these formants was scaled up/down by rounding to the nearest multiple of 200 ms. This allowed the contours to be subdivided equally into 100- or 200-ms-long segments, as required for some of the conditions (see Fig. 1). These adjusted contours were used to create three-formant analogues of the sentences using parallel synthesis and a monotonous buzz source (F0 = 140 Hz). The 3-dB bandwidths of F1, F2, and F3 were 50, 70, and 90 Hz, respectively.

For each sentence, a second-formant competitor (F2C) was created by inverting the frequency contour of F2 about its geometric mean and setting the amplitude to a constant RMS level (matching that of the target F2). F2C, when present, was delivered to the ear contralateral to the target, such that any interference caused by F2C must arise from informational masking.

In some conditions, the F2C amplitude contour was segmented such that each segment had a duration of 100 ms (C4 & C5) or 200 ms (C6 & C7). These values were informed by typical syllable durations. At each junction, the amplitude fell to zero and rose back to maximum over 20 ms (10-ms raised cosine envelopes).

In C5 & C7, the presentation order of the segments of the F2C frequency contour (defined by the segmented amplitude contour) was randomized subject to the constraint that no two consecutive segments were in their original order.

			F2C Properties		
	Target Ear	Other Ear	Amplitude Contour	Segment Order	Segment Duration (ms)
C1	F1+F3	-	-	-	-
C2	F1+F3	F2C	Constant	-	∞
C3	F1+F2+F3	F2C	Constant	-	∞
C4	F1+F2+F3	F2C	Segmented	In order	100
C5	F1+F2+F3	F2C	Segmented	Random	100
C6	F1+F2+F3	F2C	Segmented	In order	200
C7	F1+F2+F3	F2C	Segmented	Random	200
C8	F1+F2+F3	-	-	-	-

Table 1: Experimental conditions

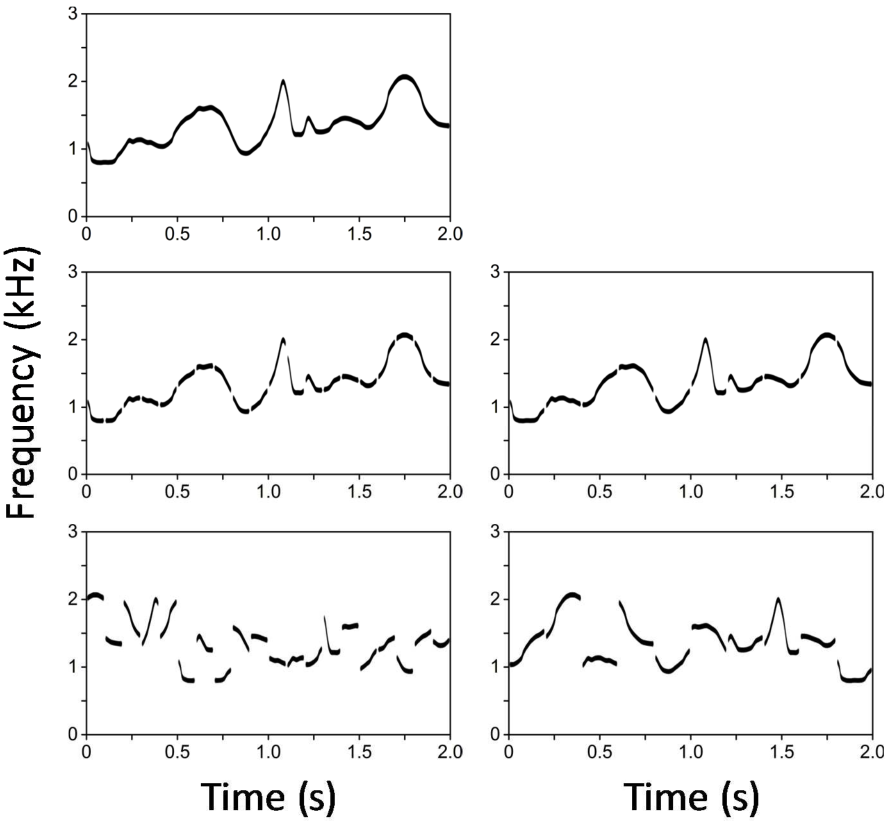

Figure 1: Schematic illustrating the segment durations and formant-frequency contours of the different F2Cs used in the experiment. All F2C frequency contours were derived from the inverted F2 frequency contour.

3. Procedure

For each listener, the 48 sentences were divided equally across conditions (i.e., 6 per condition) using an allocation that was counterbalanced by rotation across each set of 8 listeners tested.

Listeners sat in a sound-attenuating booth in front of a computer screen and keyboard. Stimuli were presented over headphones, in random order, at a reference level of 75 dB SPL. Listeners heard each stimulus once only before entering their transcription of the sentence. No feedback was given.

From trial to trial, the ear receiving the target formants was randomly assigned. Listeners first completed a training session with feedback intended to improve recognition of sentences rendered as this type of speech analogue.

Tight scoring was used to compute % keywords correct for each of the conditions.

4. Results

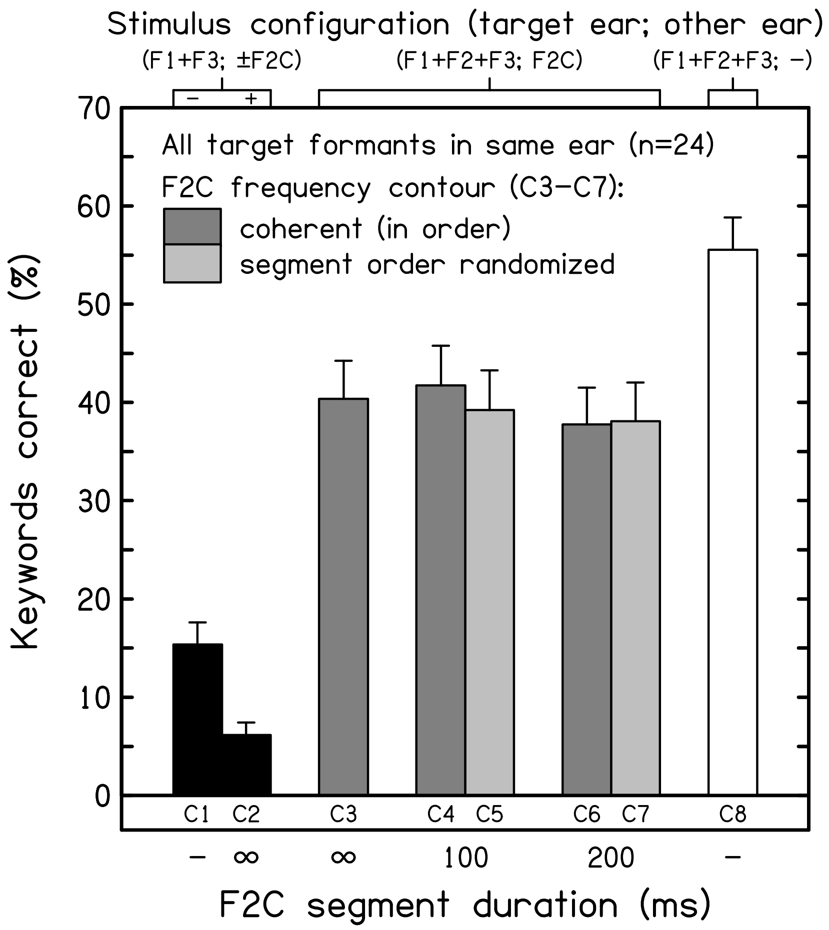

Figure 2: Experimental results.

24 listeners (2 males) successfully completed the experiment (mean age = 25.2 years). All listeners were native speakers of English.

As expected, intelligibility was low in the absence of the target F2 (black bars, C1 & C2) and was relatively high when the target formants were presented alone (white bar, C8).

Adding an F2C in the ear contralateral to the target formants (grey bars, C3-C7) led to a substantial reduction in keyword intelligibility (mean fall = 16.1%), presumably as a result of informational masking.

Critically, when F2C accompanied the target formants, the extent of its impact on intelligibility did not depend on either the segmentation of the amplitude contour (unbroken vs. divided into 100- or 200-ms-long segments) or the order randomization used for the frequency contour (coherent vs. incoherent).

5. Conclusions and Further Work

The results suggest that the impact of the extraneous formant on intelligibility depends critically on the overall extent of frequency variation in the interferer, but not on its spectro-temporal coherence.

Though not conclusive, this outcome is consistent with the notion that F2C acts as an interferer primarily by increasing the overall cognitive load on the listener.

While it is clear that introducing abrupt discontinuities in the F2C formant-frequency contour between adjacent segments does not change the extent of interference it produces, it remains unclear whether the within-segment changes in formant-frequency contour are important.

This issue might be explored in further research by replacing the time-varying frequency contour of each segment with a constant value set to the local geometric mean frequency for that segment.

Acknowledgments:

Supported by ESRC (UK) grant ES/K004905/1 (Roberts). We thank Quentin Summerfield for enunciating the sentences.

References

Kidd, G., Mason, C.R., Deliwala, P.S., Woods, W.S., Colburn, H.S. (1994). *J. Acoust. Soc. Am. 95*, 3475-3480.

Mattys, S.L., Davis, M.H., Bradlow, A.R., Scott, S.K. (2012). *Lang. Cognit. Process. 27*, 953-978.

Roberts, B., Summers, R.J. (2015). *J. Acoust. Soc. Am. 137*, 2726-2736.

Roberts, B., Summers, R.J., Bailey, P.J. (2014). *J. Exp. Psychol. Hum. Percept. Perform. 40*, 1507-1525.