Informational masking and the effects of differences in fundamental frequency and fundamental-frequency contour on phonetic integration in a formant ensemble

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

Peter J. Bailey
(Department of Psychology, University of York, York, UK)

r.j.summers@aston.ac.uk, peter.bailey@york.ac.uk, b.roberts@aston.ac.uk

1. Introduction

The intelligibility of dichotically presented target speech (left ear = F1+F3; right ear = F2) is typically impaired when it is accompanied, in the same ear as F1+F3, by a single extraneous formant intended to provide an alternative possibility for F2 (the competitor, termed F2C). For this stimulus configuration, the effect of F2C arises primarily from informational masking (e.g., Roberts et al., 2015).

Using sentence-length speech analogues presented dichotically on a monotonic F0, Summers et al. (2010) found that applying a difference in F0 (ΔF0) to F2C relative to the target formants led to a significant but relatively modest reduction in interference, which was attributed to grouping by common F0 (cf. Darwin, 1981). The reduction in interference was also influenced by absolute F0.

Summers et al. (2010) neither explored the effect of applying a ΔF0 to the target F2 (rather than to the competitor) nor examined the role of natural F0 contours in the integration of acoustic-phonetic information. The first limitation is relevant because the target F2 is spatially isolated from the other formants and so may be particularly susceptible to perceptual exclusion. The second limitation is relevant because, although it is known that F0 contour can affect overall speech intelligibility (e.g., Laures & Weismer, 1999; Binns & Culling, 2007), little attention has been paid to whether across-formant differences in F0 contour per se influence the grouping and segregation of formants. Specifically, are there direct effects of differences in F0 contour, over and above those arising from the ΔF0 that inevitably results from any mismatch in F0 contour?

The two experiments reported here address these limitations by comparing the effects of applying differences either in constant F0 or in F0 contour to the target F2, in the presence and absence of F2C. When F2C was present, its F0 contour always matched that of F1+F3. For this stimulus configuration, note that there are two grouping cues (ear of presentation and common F0) favouring the fusion of the extraneous formant with the other target formants.

2. General Methods

Stimuli were derived from the contours for F0 and the first three formants extracted from sentences with almost continuous voicing, spoken by a British male talker with Received Pronunciation English.

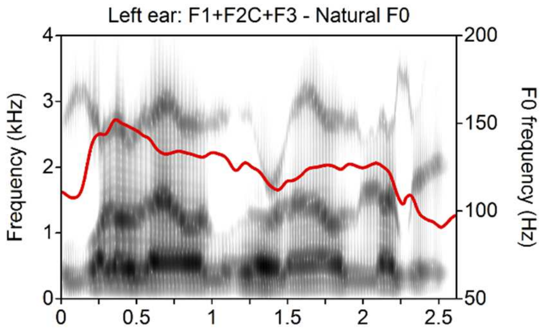
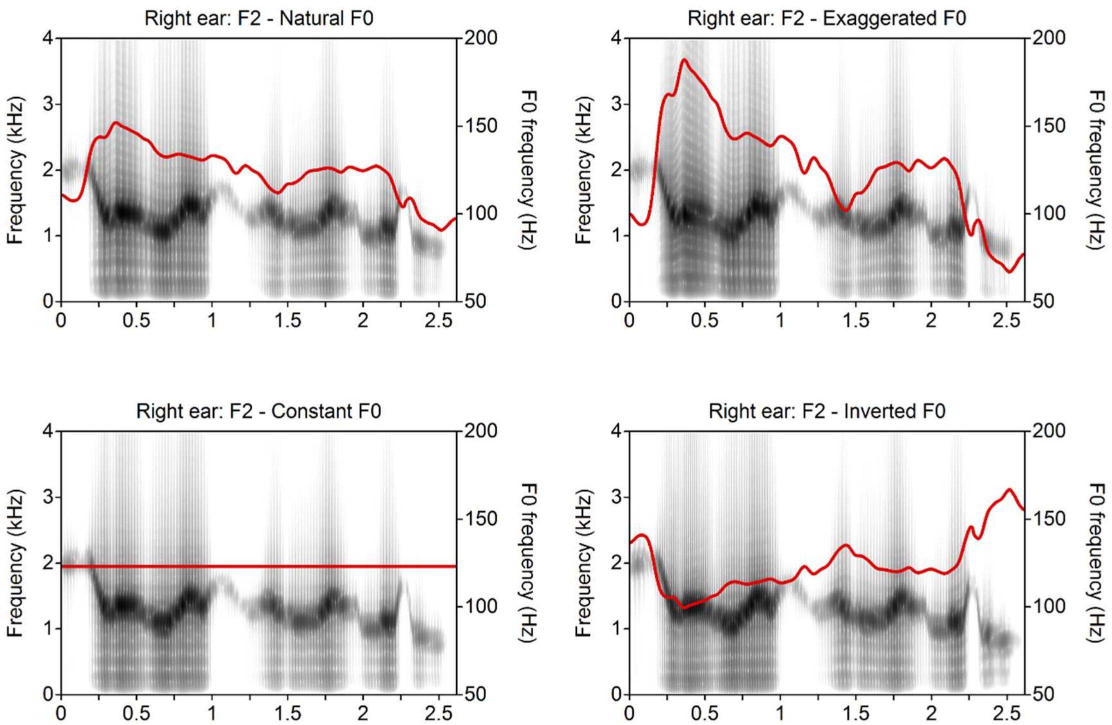


Figure 1: Stimulus spectrograms for the utterance ‘The yellow lion wore an iron muzzle.’ Each panel also shows the F0 contour for the corresponding stimulus (red line). The F1+F3 frame with F2C (top panel) and the 4 versions of the target F2 (other panels) are shown.



The frequency and amplitude contours of these formants were used to create three-formant analogues of the sentences using parallel synthesis and a simulated glottal source.

The 3-dB bandwidths of F1, F2, and F3 were 50, 70, and 90 Hz, respectively. These stimuli were presented in a dichotic configuration (F1+F3; F2).

For each sentence, a second-formant competitor (F2C) was created by inverting the formant-frequency contour of F2 about its geometric mean. When present, F2C was always in the same ear as F1+F3. Stimuli were selected such that the centre frequency of F2C was always ≥80 Hz from F1 and F3.

F1+F2C+F3 always shared the same F0 – either monotonous at 140 Hz (Experiment 1) or the natural contour (Experiment 2).

The F0 contour of F2 ($F0_{F2}$) could be the same as, or different from, that of the other formants.

	Left Ear	**Right Ear**	**ΔF0 on F2 (STs)**
C1	F1+F3	-	-
C2	F1+F2C+F3	-	-
C3	F1+F2C+F3	F2	-4
C4	F1+F2C+F3	F2	0
C5	F1+F2C+F3	F2	+4
C6	F1+F3	F2	-4
C7	F1+F3	F2	0
C8	F1+F3	F2	+4

Table 1: Conditions for Experiment 1.

	Left Ear	**Right Ear**	**F0 contour for F2**
C1	F1+F3	-	-
C2	F1+F2C+F3	-	-
C3	F1+F2C+F3	F2	Inverted (I)
C4	F1+F2C+F3	F2	Constant (C)
C5	F1+F2C+F3	F2	Exaggerated (E)
C6	F1+F2C+F3	F2	Natural (N)
C7	F1+F3	F2	Inverted (I)
C8	F1+F3	F2	Constant (C)
C9	F1+F3	F2	Exaggerated (E)
C10	F1+F3	F2	Natural (N)

Table 2: Conditions for Experiment 2.

In Experiment 1, $F0_{F2}$ was monotonic at either 111.1 Hz, 140 Hz, or 176.4 Hz (i.e., -4, 0, or +4 semitones w.r.t. 140 Hz).

In Experiment 2, $F0_{F2}$ was set to one of the following: the natural contour (N), the geometric mean frequency (constant, C), twice the depth of variation about the geometric mean F0 (exaggerated, E) or the sign of the contour's variation about the geometric mean was reversed (inverted, I). See Fig. 1.

In Experiment 2, the mean absolute ΔF0 between the other formants (N) and the target F2 was: 0 (N), 1.65 (C and E), or 3.29 semitones (I), respectively, when averaged across all sentences. Note that these F0 contours have the same mean F0, which allows the effects of ΔF0 to be isolated from those of absolute F0.

All listeners were native speakers of English.

3. Procedure

For each listener, the sentences were divided equally across conditions (6 per condition) using an allocation that was counterbalanced by rotation across each set of 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 up to three times (Experiment 1) or once only (Experiment 2) before entering their transcription of the sentence. No feedback was given.

Listeners first completed a training session with feedback intended to improve recognition of the speech analogues used.

Tight and Loose Scoring were used to compute % keywords correct for each of the conditions.

4. Experiment 1: Results

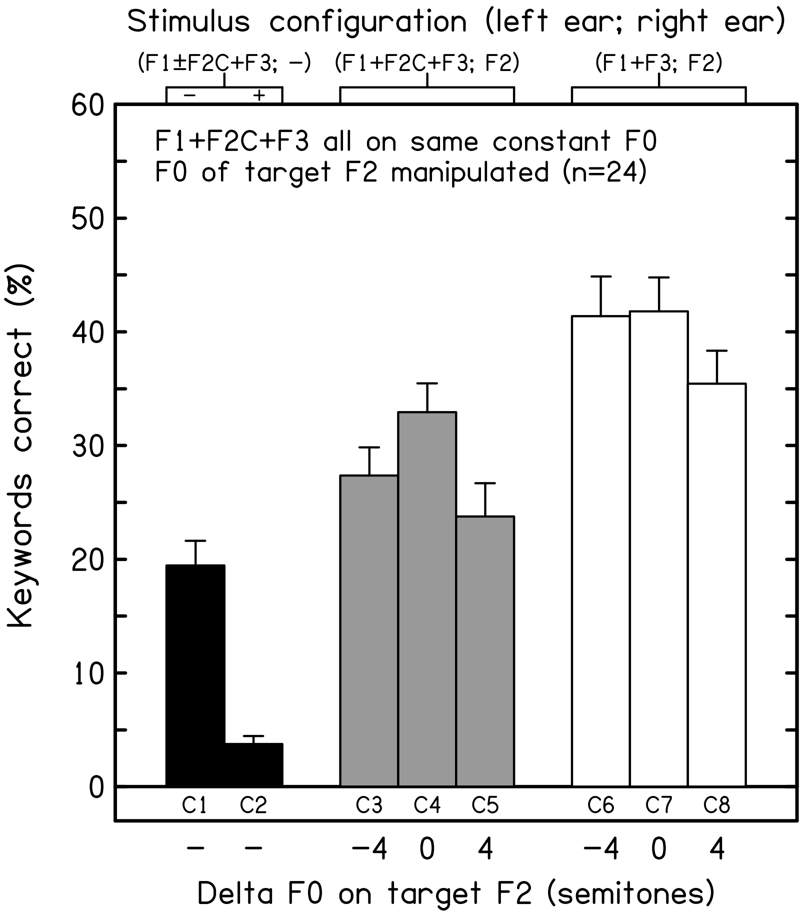


Figure 2: Results for Experiment 1.

24 listeners (3 males) successfully completed the experiment (mean age = 20.5 years).

As expected, intelligibility was low in the absence of F2 (black bars, C1 & C2). When the target formants were presented alone, intelligibility was relatively high (white bars, C6-C8). Adding F2C in the same ear as F1+F3 (grey bars, C3-5) led to a clear reduction in keyword intelligibility, presumably mainly as a result of informational masking (IM, mean fall = 11.5%).

There was no significant effect of applying ΔF0 to the target F2 when the target formants were presented alone.

When F2C accompanied the target formants, the extent of its impact was dependent on ΔF0. Intelligibility was highest when ΔF0 = 0 ST and lowest when ΔF0 = +4 ST.

If expressed as a difference score w.r.t. the corresponding target-only case, F2C impact for the F0s tested was 8.9% (ΔF0 = 0 ST), 14.0% (ΔF0 = -4 ST), and 11.6% (ΔF0 = +4 ST).

Critically, however, there was no significant interaction between F2C status (present/absent) and F0 for target F2. There are indications that absolute F0 may have influenced this outcome.

5. Experiment 2: Results

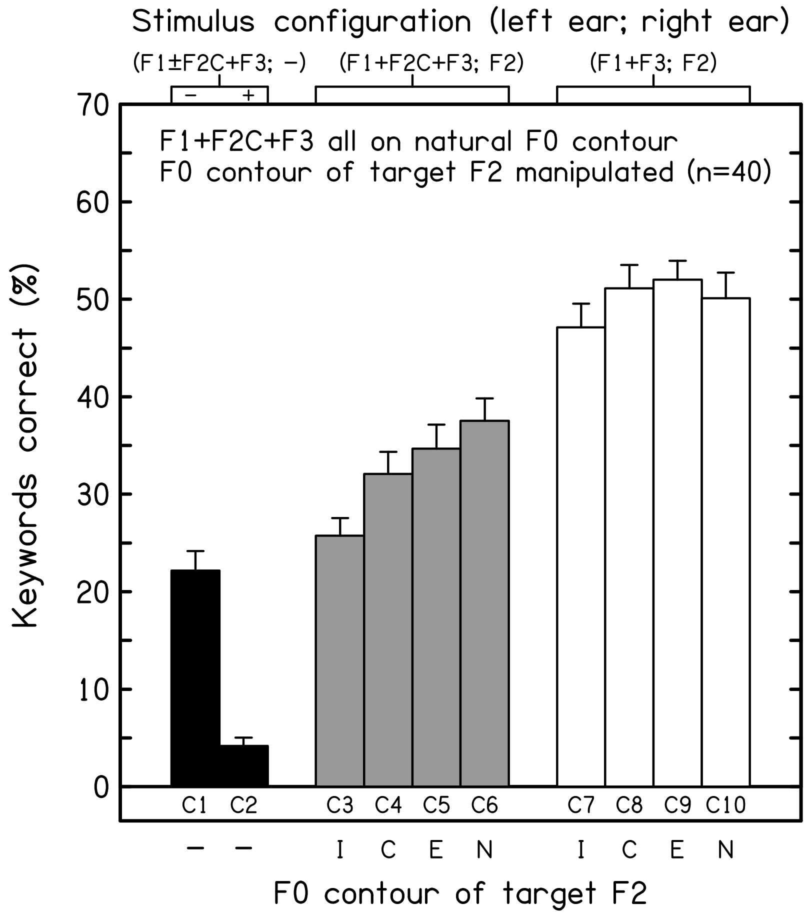


Figure 3: Results for Experiment 2.

Adding an F2C in the same ear as F1+F3 led to a clear reduction in keyword intelligibility (mean fall = 17.0%), presumably mainly as a result of IM.

When the target formants were presented alone (white bars, C7-C10), there was no intelligibility cost of applying a different F0 contour to F2, despite the consequent $\Delta F0$ between F2 and F1+F3 when the $F0_{F2}$ contour was inverted, constant, or exaggerated.

When F2C accompanied the target formants (grey bars, C3-C6), the extent of its impact on intelligibility was dependent on the F0 contour applied to the target F2.

If expressed as a difference score with respect to the corresponding target-only case, competitor impact for the $F0_{F2}$ contours tested was 21.5% (I), 18.5% (E), 17.3% (C), and 10.7% (N). In this case, for which there were no differences in mean F0 across formants, there was a significant interaction between F2C status (present/absent) and F0 contour for the target F2.

Competitor impact was greatest when F2 had an inverted F0 contour, for which the mean $\Delta F0$ w.r.t. F1+F3 was largest. F2C impact was about double for the I vs. N case. There was no indication that the shape/pattern of the F0 contour of F2 per se affected F2C impact (cf. C vs. E contours, which share the same mean ΔF0).

6. Conclusions

The literature on informational masking for non-speech stimuli suggests an important role for target-masker similarity in determining the extent of interference (e.g., Neff, 1995; Lee & Richards, 2011).

The experiments reported here indicate that differences in F0 between target formants can influence their grouping and segregation, when accompanied by a competitor, in a manner consistent with grouping by target-masker similarity.

This result contrasts with that obtained when more radical differences in source type are involved. If some formants are rendered as tonal (sine-wave source) and others as harmonic analogues (buzz source, as here), the tonal analogues always lose out to the harmonic analogues under competitive conditions, irrespective of target-masker similarity (Roberts et al., 2015; Summers et al., 2016).

Although it is known that the F0 contour of a sentence can affect its overall intelligibility, the experiments reported here suggest that any effects of differences between formants in F0 contour are indirect, arising from the consequent ΔF0.

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

References

Binns, C., Culling, J.F. (2007). *J. Acoust. Soc. Am. 122*, 1765-1776.

Darwin, C.J. (1981). *Q. J. Exp. Psychol. 33A*, 185-207.

Laures, J.S., Weismer, G. (1999). *J. Speech Lang. Hear. Res. 42*, 1148-1156.

Lee, T.Y., Richards, V.M. (2011). *J. Acoust. Soc. Am. 129*, EL280-EL285.

Neff, D.L. (1995). *J. Acoust. Soc. Am. 98*, 1909-1920.

Roberts, B., Summers, R.J., & Bailey, P.J. (2015). *JEP:HPP 41*, 680-691.

Summers R.J., Bailey P.J., Roberts B. (2010). *J. Acoust. Soc. Am. 128*, 3667-3677.

Summers R.J., Bailey P.J., Roberts B. (2016). *J. Acoust. Soc. Am.*, accepted subject to minor revision.