Across-formant integration and speech intelligibility: Effects of acoustic source properties in the presence and absence of a contralateral interferer

Robert J. Summers and 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 ability to group and segregate formants across frequency using cues in the acoustic source characteristics is poorly understood. If the fundamental frequency (F0) of one formant is different from that of the others then it tends to segregate (Darwin, 1981; Summers et al., 2010).

Recently, it has been shown that radical differences in the source characteristics of individual formants can also affect segregation (Roberts et al., 2015). In that study, F1+F3 was either harmonic (H) or tonal (T) and there was little difference in intelligibility whether or not the acoustic source properties of F2 (presented contralaterally) matched F1+F3. In some conditions, a competitor for F2 (F2C) was presented in the same ear as F1+F3; the competitor must be rejected to optimize recognition. Intelligibility was lowest when F2C was harmonic and F2 was tonal, regardless of the acoustic source properties of F1+F3.

These findings suggest that it is the type of acoustic source that governs intelligibility rather than acoustic similarity between formants. As harmonic formants are wideband, they are louder than their narrowband tonal counterparts when matched for RMS power. Hence, a possible alternative account is that louder formants dominate in contributing phonetic information.

Here, we present three experiments designed to explore the effects of source characteristics, and loudness differences, in a configuration where F2 is presented in the same ear as F1+F3 and F2C is presented in the contralateral ear.

2. General Methods

Stimuli were derived from the first 3 formants of 96 BKB sentences (Bench et al., 1979) spoken by a British male talker with Received Pronunciation English. Each sentence comprised ≤25% phonemes involving closures or unvoiced frication. The frequency and amplitude contours of these formants were used to create three-formant (F1+F2+F3) analogues of the sentences using parallel synthesis.

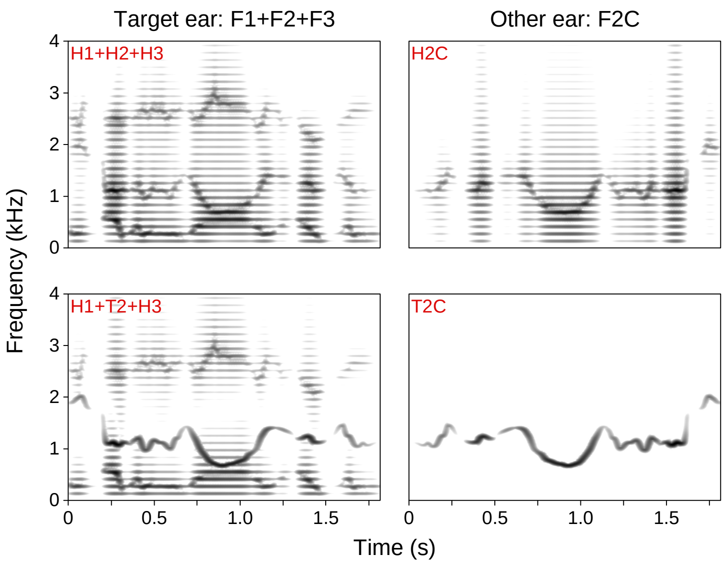

Figure 1: Example configurations of the target formants only (F1+F2+F3, left column) and the interfering formant (F2C, right column) when F2/F2C is either harmonic (top row) or tonal (bottom row), for the utterance "*The oven door was open*" when F1+F3 is harmonic.

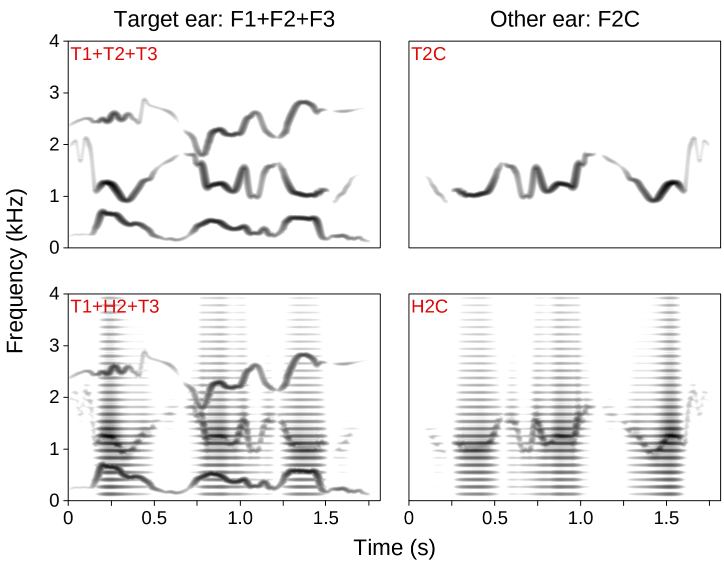

Figure 2: Example configurations of the target formants only (F1+F2+F3, left column) and the interfering formant (F2C, right column) when F2/F2C is either tonal (top row) or harmonic (bottom row), for the utterance "*Flowers grow in the garden*" when F1+F3 is tonal.

Each formant was either generated by passing a buzz excitation source through a second-order resonator (harmonic, H) or was a sine-wave analogue (tonal, T). The glottal source was monotonous (F0 = 140 Hz), and the 3-dB bandwidths of H1, H2, and H3 were 50, 70, and 90 Hz, respectively. F1 and F3 were always derived from the same source (H or T). F2 was presented in the same ear as F1+F3 and either had the same or different source characteristics as F1+F3.

For each sentence, competitors (H2C & T2C) were created using the time-reversed frequency and amplitude contours of F2. F2C was always presented in the contralateral ear to F1+F2+F3. For stimuli with an H1+H3 frame, the level of T2/T2C was *boosted* so as to match the RMS power of H2. For stimuli with a T1+T3 frame, the level of H2/H2C was *reduced* so as to match the RMS power of T2.

3. Procedure

For each listener, the 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. Tight scoring was used to calculate percent keywords correctly identified for each of the conditions.

4. Conditions: Experiment 1 and 2

Condition	Target ear	Other ear
C1	T1+T3	-
C2	T1+T3	T2C
C3	T1+T2+T3	T2C
C4	T1+T2+T3	H2C
C5	T1+H2+T3	T2C
C6	T1+H2+T3	H2C
C7	T1+T2+T3	-
C8	T1+H2+T3	-

Condition	Target ear	Other ear
C1	H1+H3	-
C2	H1+H3	H2C
C3	H1+H2+H3	H2C
C4	H1+H2+H3	T2C
C5	H1+T2+H3	H2C
C6	H1+T2+H3	T2C
C7	H1+H2+H3	-
C8	H1+T2+H3	-

Tables 1 and 2: Experimental conditions for Experiment 1 (left) and Experiment 2 (right)

There were eight conditions in experiment 1 and in experiment 2.

C1 and C2 were the F2-absent conditions. The stimuli for C1 comprised the F1+F3 frame alone; C2 differed only in that F2C (same acoustic source as F1+F3) was present in the contralateral ear.

The stimuli for C3-C6 comprised all three target formants plus the contralateral competitor. This set represents all four combinations of acoustic properties for F2 and F2C.

The stimuli for the remaining conditions (C7-C8) comprised only the target formants.

5. Experiment 1: Harmonic F1+F3 frame

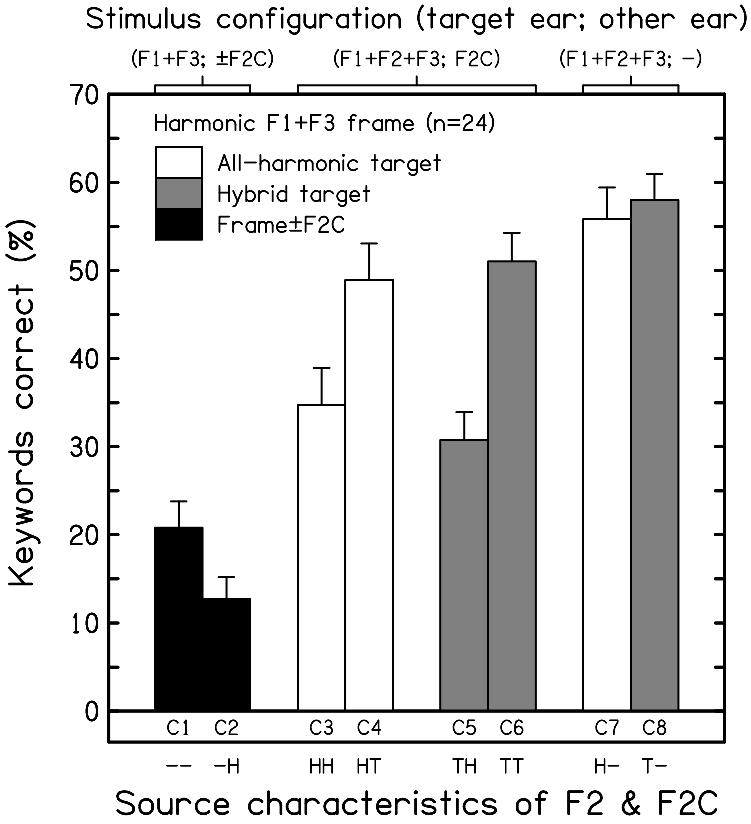

Figure 3: Results from Experiment 1.

Twenty-four listeners (4 males) successfully completed the experiment (mean age = 19.6 years, range = 18.2 – 34.8).

Results largely confirm those of Roberts et al. (2015). Adding F2 to H1+H3 substantially improved intelligibility regardless of whether F2 was harmonic (matched; C1 vs C7) or tonal (mismatched; C1 vs C8). Harmonic competitors were ***highly effective*** regardless of whether F2 was harmonic (matched; C3 vs C7) or tonal (mismatched; C5 vs C8). Tonal competitors were ***not effective*** regardless of whether F2 was harmonic (matched; C4 vs C7) or tonal (mismatched; C6 vs C8).

6. Experiment 2: Tonal F1+F3 frame

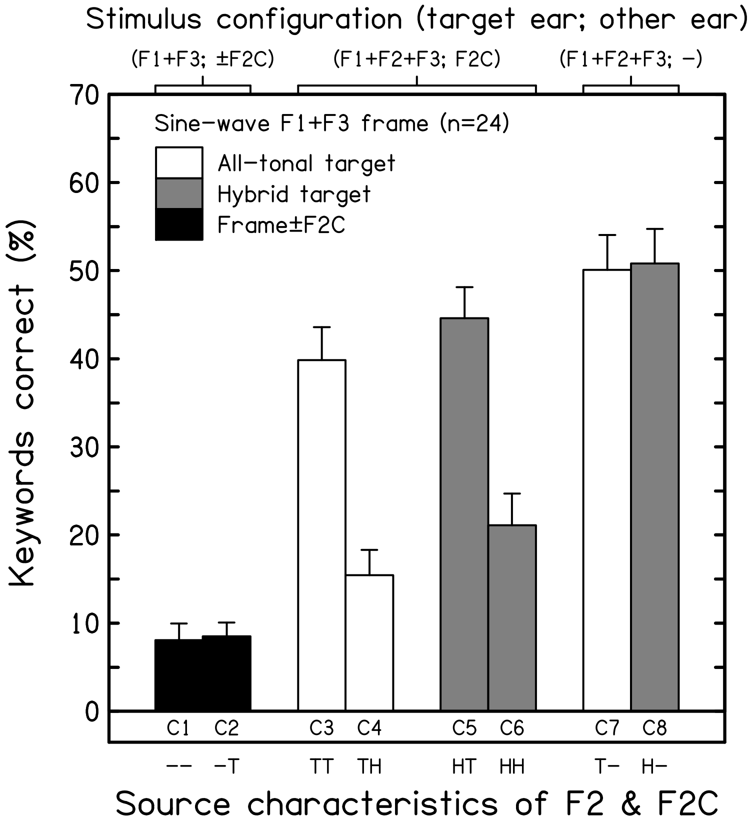

Figure 4: Results from experiment 2.

Twenty-four listeners (9 males) successfully completed the experiment (mean age = 25.3 years, range = 18.9 – 48.8).

Results largely confirm those of Roberts et al. (2015). Adding F2 to T1+T3 substantially improved intelligibility regardless of whether F2 was tonal (matched; C1 vs C7) or harmonic (mismatched; C1 vs C8). Tonal competitors were ***not effective*** regardless of whether F2 was tonal (C3 vs C7) or harmonic (C5 vs C8). Harmonic competitors were ***highly effective*** regardless of whether F2 was tonal (C4 vs C7) or harmonic (C6 vs C8).

7. Experiment 3: F2C Loudness control

Do harmonic competitors have a greater impact on intelligibility than their RMS-matched tonal counterparts because they are louder? In this experiment, all target stimuli were rendered as mixed-source signals. Stimuli were derived from the 24 most intelligible sentences in experiment 1 (harmonic frame) and the 24 most intelligible in experiment 2 (tonal frame).

Level adjustments to match loudness between F2 & F2C were restricted to F2C only; these adjustments have no effect on energetic masking (F2C is in the contralateral ear to the target formants). We used the software for the time-varying loudness model (Glasberg and Moore, 2002) to estimate the loudness of the isolated F2 & F2C in C2 and C7.

Assuming that loudness-level differences between tonal and harmonic formants (mean ≈ 9 phon) are approx. equivalent to a difference in dB, initially the level of H2C (in C2) was reduced by ~9 dB and T2C (in C7) was boosted by ~9 dB. Remaining loudness-level differences between F2 & F2C were computed so that final dB adjustments could be made. Matches were close. After correction, the mean residual difference between F2 & F2C in C2 and C7 was ~0.1 phon.

Condition	**Target ear**	**Other ear**
C1	H1+H3	-
C2	H1+T2+H3	H2C
C3	H1+T2+H3	T2C
C4	H1+T2+H3	-
C5	T1+T3	-
C6	T1+H2+T3	H2C
C7	T1+H2+T3	T2C
C8	T1+H2+T3	-

Table 3: Experimental conditions for Experiment 3.

8. Experiment 3: Results

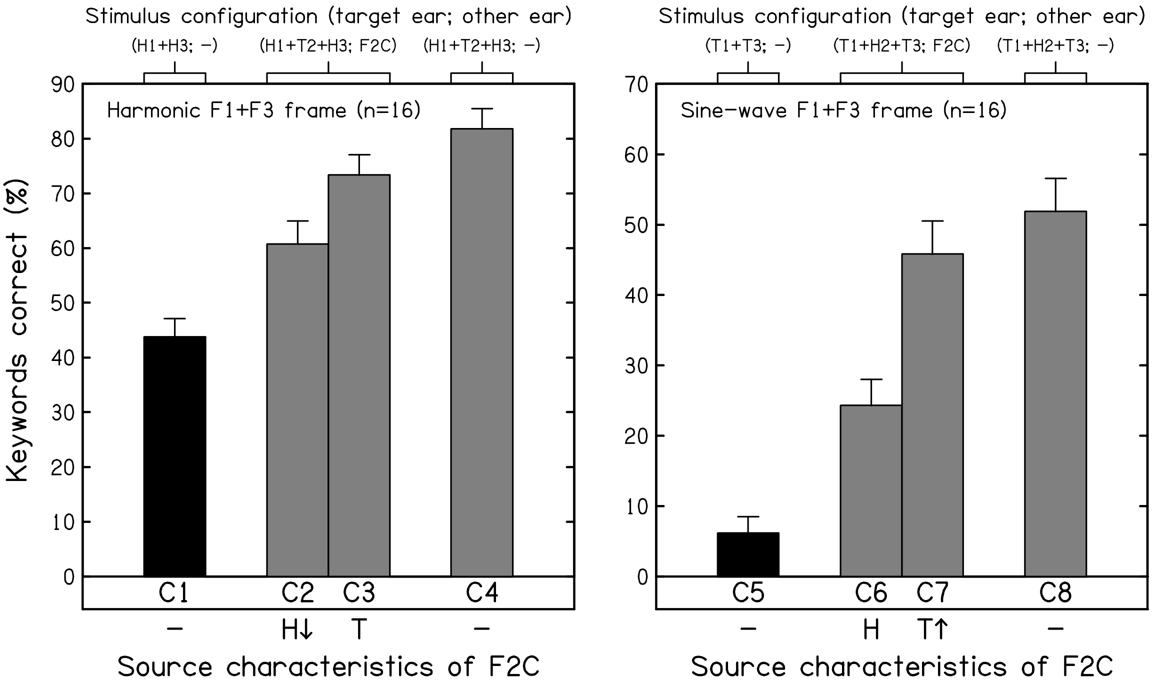

Figure 5: Results from Experiment 3.

Sixteen listeners (three males) successfully completed the experiment (mean age = 20.8 years, range = 18.2 – 37.9).

Results confirm those of experiments 1 and 2. Regardless of the acoustic source properties of F1+F3 or F2, H2C with ~9 dB less energy than T2C was significantly more effective as a competitor (C2 vs C3; C6 vs C7).

9. Conclusions

The impact of adding F2C was modest when it was tonal, but large when it was harmonic, regardless of whether the source for F2C matched that for F1+F3. Consistent with Roberts et al. (2015), this outcome contradicts the idea that target-masker similarity governs across-formant grouping.

Even when F2 and F2C were matched for loudness, H2C was still more effective than T2C regardless of the source properties of F1+F3.

An important difference from the results for dichotic targets (Roberts et al., 2015) is that here (monaural targets) H2C was no more effective at interfering with the phonetic contribution of T2 than with that of H2. This may reflect the need in that study for formant integration across ears, as well as across frequency.

Acknowledgments

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

References

Bench, J., Kowal, A., Bamford, J. (1979). Brit. J. Audiol. 13, 108-112.

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

Glasberg, B.R., & Moore, B.C.J. (2002). J. Audio Eng. Soc. 50, 331–342.

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.