Hearing Research xxx (2016) 1–9

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

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

Robert J. Summers [a], Peter J. Bailey [b], Brian Roberts [a, *]

[a] Psychology, School of Life and Health Sciences, Aston University, Birmingham B4 7ET, UK
[b] Department of Psychology, University of York, Heslington, York YO10 5DD, UK

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Speech intelligibility
Formant integration
Informational masking
Competition
Speech segregation
Target-masker similarity

ABSTRACT

This study explored the effects on speech intelligibility of across-formant differences in fundamental frequency (ΔF0) and F0 contour. Sentence-length speech analogues were presented dichotically (left = F1+F3; right = F2), either alone or—because competition usually reveals grouping cues most clearly—accompanied in the left ear by a competitor for F2 (F2C) that listeners must reject to optimize recognition. F2C was created by inverting the F2 frequency contour. In experiment 1, all left-ear formants shared the same constant F0 and $\Delta F0_{F2}$ was 0 or ±4 semitones. In experiment 2, all left-ear formants shared the natural F0 contour and that for F2 was natural, constant, exaggerated, or inverted. Adding F2C lowered keyword scores, presumably because of informational masking. The results for experiment 1 were complicated by effects associated with the direction of $\Delta F0_{F2}$; this problem was avoided in experiment 2 because all four F0 contours had the same geometric mean frequency. When the target formants were presented alone, scores were relatively high and did not depend on the $F0_{F2}$ contour. F2C impact was greater when F2 had a different F0 contour from the other formants. This effect was a direct consequence of the associated ΔF0; the $F0_{F2}$ contour *per se* did not influence competitor impact.

1. Introduction

When more than one talker is speaking at once, successful communication depends on the ability of the listener to separate the formant ensemble reaching their ears into a figure (target) and background (interferer). There are a number of ways in which the interferer may lower the intelligibility of the target speech; these can be categorized broadly into energetic masking, in which the auditory-nerve response to the target is swamped by the response to the masker, modulation masking, in which masker amplitude variation lowers sensitivity to similar rates of variation in the target (e.g., Stone and Moore, 2014; Stone and Canavan, 2016), and informational masking, which is of central origin and may be considered as encompassing all other forms of interference (e.g., Durlach et al., 2003; Kidd et al., 2008). The study reported here is concerned with informational masking, in which the interference may arise from the disruption of auditory object formation or selection, or from an increase in the cognitive load on the listener (see, e.g., Shinn-Cunningham, 2008; Mattys et al., 2012).

The voices of two talkers speaking at the same time usually differ in fundamental frequency (F0) and in F0 contour; these differences provide acoustic cues for voice segregation that may assist listeners trying to understand what is being said. In the context of the integration of acoustic-phonetic information across formants, it is known that a difference in fundamental frequency (ΔF0) between formants influences their grouping and segregation (Darwin, 1981; Gardner et al., 1989; Bird and Darwin, 1998; Summers et al., 2010). The focus of these studies differs from the many that have explored the effect of ΔF0 on the ability to separate a mixture of two voices within the same ear (e.g., Brokx and Nooteboom, 1982; Binns and Culling, 2007; Deroche et al., 2014) in that performance is limited mainly by the ability to group acoustic elements correctly across frequency regions rather than to separate overlapping harmonics (for a review, see Summers et al., 2010). Studies of the perceptual organization of a formant ensemble indicate that imposing a ΔF0

Abbreviations: CV, consonant-vowel; F0, fundamental frequency; F1, first formant; F2, second formant; F2C, second formant competitor; F3, third formant; % pts, percentage points

* Corresponding author.
E-mail address: b.roberts@aston.ac.uk (B. Roberts).

http://dx.doi.org/10.1016/j.heares.2016.10.026
0378-5955/

on one formant in the ensemble can reduce its phonetic contribution to the speech percept, but suggest that this reduction occurs solely or mainly in circumstances where there is competition between alternative candidates for one or more of the lower formants (Darwin, 1981; Gardner et al., 1989).

Summers et al. (2010) explored the effect of differences in F0 on across-formant grouping and segregation using sentence-length speech analogues and the second-formant competitor (F2C) paradigm (e.g., Remez et al., 1994; Roberts et al., 2010). This paradigm involves the dichotic presentation of two versions of F2, for which intelligibility is enhanced by the phonetic integration of one version (target F2) with the other formants (F1+F3) but impaired by the integration of the other, a single extraneous formant intended to act as a competitor to F2 (F2C). Hence, the listener must reject the competitor to optimize recognition of the utterance. The version of the F2C paradigm used by Summers et al. (2010) involved presenting the target formants on a monotonic F0 of 150 Hz to separate ears (left ear = F1± F2C+F3; right ear = F2). The inclusion of the competitor lowered intelligibility and applying a ΔF0 to F2C relative to the target formants led to a significant but relatively modest fall in interference, which was attributed to grouping by common F0. The dichotic configuration allowed competition between the two versions of F2 in a context where any interference must have arisen primarily through informational masking. Note that any contribution of energetic masking to competitor impact arising from adding F2C in the same ear as F1+F3 must have been small or negligible for two reasons. First, F1 was lower in frequency and more intense than F2C. Second, competitor impact remained the same when the possibility of upward spread of masking from F2C to F3 was eliminated by moving F3 to the opposite ear (Summers et al., 2010; cf. Rand, 1974).

There are two limitations of the study by Summers et al. (2010) that merit further investigation. First, it did not explore the effect of applying a ΔF0 to the target F2 (rather than to the competitor); second, it did not explore the role of natural F0 contours in the integration of acoustic-phonetic information across formants. The first limitation is important because the target F2 is spatially isolated from the others in the stimulus configuration used and so may be particularly susceptible to perceptual exclusion on the basis of primitive grouping cues (Bregman, 1990; Darwin, 2008). The second limitation is important because the Gestalt principle of good continuation suggests that the smooth and continuous change characteristic of a natural F0 contour might assist in binding together all acoustic elements following that contour, and yet almost no attention has been paid to whether across-formant differences in F0 contour *per se* influence the grouping and segregation of formants. Specifically, are there any direct effects of differences in F0 contour between formants, over and above those arising from the ΔF0 that inevitably results from any mismatch in F0 contour?

To our knowledge, only one experiment has examined the effect of introducing time-varying (as well as static) ΔF0s between formants in an ensemble, in this case one constituting a consonant-vowel (CV) syllable. In their second experiment, Gardner et al. (1989) manipulated a synthetic four-formant ensemble that could be perceived as /ru/ or /li/. When presented alone, formants 1, 2, and 3 elicited /ru/ percepts and formants 1, 3, and 4 elicited /li/ percepts. When all four formants were presented together on the same F0, almost all responses indicated /ru/ percepts. However, when a ΔF0 was applied to formant 2, the syllable could be heard as /ru/ or /li/ (or as both) depending on whether or not the phonetic information carried by formant 2 was integrated into the percept. In addition to static F0 differences between formant 2 and the rest, the effects of coherent and incoherent sinusoidal modulation of F0 between the two sets were compared (rate = 6 Hz or 12 Hz; depth = ±3% or ±8%; phase difference = 0° or 90°). There was no evidence that the coherence of the motion of F0 had any additional effect on the perceptual grouping of the formants over and above the effect of a static ΔF0. Nonetheless, it would be premature to generalize from this finding obtained for synthetic CV syllables and sinusoidal F0 contours and to assume that there is no additional role for F0 contour in the grouping and segregation of formants for sentence-length utterances synthesized using the natural pattern of F0 variation.

Investigations of the influence of variations in voice pitch on speech intelligibility have generally been restricted to cases where all the formants share the same F0 contour. A number of studies have shown that changing the F0 contour from the natural pattern of variation usually lowers the intelligibility of sentence-length utterances. Such effects have been found even when high-quality speech is heard in quiet, but the impact of such change tends to become more pronounced in more adverse listening conditions, such as low-pass filtering (Hillenbrand, 2003) or the presence of background noise (Miller et al., 2010) or a competing talker (Binns and Culling, 2007). The most common manipulation is to flatten the F0 contour to a monotone, removing any prosodic information carried by the natural pattern of F0 variation (Wingfield et al., 1984; Laures and Weismer, 1999; Hillenbrand, 2003; Binns and Culling, 2007; Miller et al., 2010; Deroche et al., 2014). Under otherwise similar listening conditions, the impact on intelligibility is greater when the prosodic information provided by F0 variation is not simply removed but is instead made misleading by inverting the natural pattern of variation. For example, Miller et al. (2010) found that flattening the F0 contour of speech presented in noise lowered keyword scores by ~13 percentage points (% pts) relative to the natural contour, whereas inverting the F0 contour lowered performance by ~23% pts. Their study also included a condition in which the natural F0 variation was exaggerated by × 1.75, for which the effect was similar to flattening the contour (~13% pts reduction). Presumably, exaggeration had less effect than inversion because the variations were in the same direction moment-to-moment as for the natural contour. In contrast to these studies, which were designed primarily to explore the prosodic properties of F0 contours, the current study used natural, constant, exaggerated, and inverted contours to introduce time-varying differences in F0 between one formant (F2) and the others.

The two experiments reported here addressed the limitations of Summers et al. (2010) 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. Whilst the primary goal of this study was to use speech acoustics to extend our understanding of the role of F0 as an auditory grouping cue, these experiments also cast further light on the nature of acoustic-phonetic integration in speech perception.

2. Experiment 1

In this experiment, the F0 of F2 ($F0_{F2}$) could be the same as, or different from, that of the other formants. The purpose of the experiment was to measure the extent to which the intelligibility of dichotic target speech (F1+F3; F2) was dependent on the difference in F0 between the isolated target F2 and the other formants, in the presence and absence of a competitor (F2C) that shared a common F0 and ear of presentation with the F1+F3 "frame". Note that the presence of the competitor is challenging for the listener, as maximizing intelligibility involves discarding the acoustic-phonetic information carried by a misarticulated but seemingly genuine

Please cite this article in press as: Summers, R.J., et al., Informational masking and the effects of differences in fundamental frequency and fundamental-frequency contour on phonetic integration in a formant ensemble, Hearing Research (2016), http://dx.doi.org/10.1016/j.heares.2016.10.026