

## Mandatory dichotic integration of second-formant information: Contralateral sine bleats have predictable effects on consonant place judgments

Brian Roberts, Robert J. Summers and Peter J. Bailey

Citation: [The Journal of the Acoustical Society of America](#) **150**, 3693 (2021); doi: 10.1121/10.0007132

View online: <https://doi.org/10.1121/10.0007132>

View Table of Contents: <https://asa.scitation.org/toc/jas/150/5>

Published by the [Acoustical Society of America](#)

---

---



**Advance your science and career  
as a member of the**

**ACOUSTICAL SOCIETY OF AMERICA**

LEARN MORE



## Mandatory dichotic integration of second-formant information: Contralateral sine bleats have predictable effects on consonant place judgments

Brian Roberts,<sup>1,a)</sup> Robert J. Summers,<sup>1,b)</sup> and Peter J. Bailey<sup>2</sup>

<sup>1</sup>*School of Psychology, Aston University, Birmingham B4 7ET, United Kingdom*

<sup>2</sup>*Department of Psychology, University of York, Heslington, York YO10 5DD, United Kingdom*

### ABSTRACT:

Speech-on-speech informational masking arises because the interferer disrupts target processing (e.g., capacity limitations) or corrupts it (e.g., intrusions into the target percept); the latter should produce predictable errors. Listeners identified the consonant in monaural buzz-excited three-formant analogues of approximant-vowel syllables, forming a place of articulation series (/w/-l/-j/). There were two 11-member series; the vowel was either high-front or low-back. Series members shared formant-amplitude contours, fundamental frequency, and F1+F3 frequency contours; they were distinguished solely by the F2 frequency contour before the steady portion. Targets were always presented in the left ear. For each series, F2 frequency and amplitude contours were also used to generate interferers with altered source properties—sine-wave analogues of F2 (sine bleats) matched to their buzz-excited counterparts. Accompanying each series member with a fixed mismatched sine bleat in the contralateral ear produced systematic and predictable effects on category judgments; these effects were usually largest for bleats involving the fastest rate or greatest extent of frequency change. Judgments of isolated sine bleats using the three place labels were often unsystematic or arbitrary. These results indicate that informational masking by interferers involved corruption of target processing as a result of mandatory dichotic integration of F2 information, despite the grouping cues disfavoring this integration.

© 2021 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1121/10.0007132>

(Received 17 June 2021; revised 28 September 2021; accepted 22 October 2021; published online 17 November 2021)

[Editor: Karen S. Helfer]

Pages: 3693–3710

### I. INTRODUCTION

Listeners often experience circumstances in which speech perception takes place in complex auditory scenes, involving segregating and attending to target speech in the presence of other sounds, including interfering speech (see, e.g., Bregman, 1990; Darwin, 2008; Mattys *et al.*, 2012). Such adverse listening conditions tend to lower speech intelligibility for normal-hearing listeners and the impact is even greater for listeners with hearing impairments. The masking produced by other sounds can arise through energetic masking (EM), in which encoding of the target speech in the auditory-nerve response is degraded, or through informational masking (IM) in the central auditory system, even when the properties of the target speech are faithfully represented in the peripheral response. The speech signal is sparse on a frequency  $\times$  time representation, and so IM is often the main cause of speech-on-speech interference, especially when there is only one interfering voice and that voice is similar in level or lower than the target voice (Brungart *et al.*, 2006; Darwin, 2008). IM can result from failures of object formation and selection—for example, the

interferer may intrude into the target percept—or from capacity limitations on the resources available for information processing (e.g., Shinn-Cunningham, 2008). These aspects of IM can be regarded as *corrupting* and *disrupting* effects, respectively (e.g., Roberts *et al.*, 2014; Roberts and Summers, 2018), and the experiments reported here were designed to explore the contribution of corrupting effects to IM when listening to speech in the presence of other sounds in the contralateral ear.

A convenient method for investigating speech-on-speech IM with a high degree of stimulus control is the second-formant competitor (F2C) paradigm (Remez *et al.*, 1994; Roberts *et al.*, 2010). This paradigm involves accompanying a three-formant analogue of a target sentence with an extraneous formant intended to act as a competitor by providing an alternative candidate for F2. F2C is always presented in the opposite ear to F2 and so causes relatively little EM for dichotic targets (i.e., F1 and F3 in the same ear as F2C) and none for monaural targets. Hence, the impact of the extraneous formant on target intelligibility arises mainly or entirely from IM. F2C is usually derived from the properties of F2—e.g., by time reversing or inverting the F2 frequency contour—and must be rejected by the listener to optimize recognition of the target speech. Research using this paradigm or variants thereof has shown that the impact

<sup>a)</sup>Electronic mail: b.roberts@aston.ac.uk, ORCID: 0000-0002-4232-9459.

<sup>b)</sup>ORCID: 0000-0003-4857-7354.

of extraneous formants on intelligibility is governed mainly by the time-varying properties of their formant-frequency contours, particularly the depth (extent) and the rate of that variation (Roberts *et al.*, 2010, 2014; Roberts and Summers, 2015, 2018, 2020; Summers *et al.*, 2012).

The F2C paradigm was originally conceptualized as involving competition between the target F2 and the extraneous formant to form a coherent perceptual group with F1 and F3, such that any fall in intelligibility occurred because F2 was displaced (or diluted) from the perceptual organization of the target speech (Remez *et al.*, 1994; Roberts *et al.*, 2010). Roberts *et al.* (2014) noted that this interpretation attributed the IM produced to a specific corrupting influence of the interferer—a failure to exclude acoustic variation in the extraneous formant from the perceptual evaluation of the acoustic-phonetic features of the target sentence—but that the impact of the interferer may instead have resulted from a relatively non-specific disrupting influence. For example, the interferer may have acted in a similar way to a cognitive load, limiting the resources available to process the target sentence (see, e.g., Mattys *et al.*, 2012). Roberts and Summers (2018) noted two results consistent with a role for disruption in the IM produced by the extraneous formants. First, transposing an extraneous formant of a given depth and pattern of formant-frequency variation across a wide range of frequencies had little effect on the amount of IM produced. Second, the importance of formant-frequency variation in an interferer for the IM it causes has clear parallels with a form of cross-modal interference known as the irrelevant sound effect (ISE), in which an acoustic distractor that participants must ignore nonetheless impairs serial recall of visually presented digits or words (Jones and Macken, 1993; for a review, see Ellermeier and Zimmer, 2014). Most notably, the distractor must involve frequency change for the ISE to occur—amplitude change alone is not sufficient (Tremblay and Jones, 1999)—and the ISE is usually greatest for distractors involving complex spectro-temporal change, such as speech or instrumental music (e.g., Viswanathan *et al.*, 2014; Dorsi *et al.*, 2018).

The relative contributions of corruption and disruption to speech-on-speech IM remain to be established. A property of the former that can be used to distinguish the two effects is that intrusions from the interferer into the target percept should lead to specific and predictable errors. However, using sentence-length materials limits the kind of intrusions that can easily be identified. While it is straightforward to identify intrusions of whole words from an intelligible interferer into the percept of the target speech (e.g., Summers and Roberts, 2020), attributing phonemic errors to the intrusion of specific features of a frequency-varying extraneous formant would be challenging using such long materials. However, evidence consistent with the idea that speech-on-speech IM involves specific corrupting effects at the level of the extraction and integration of the acoustic-phonetic information carried by different formants has come from a largely overlooked study by Porter and Whittaker (1980) that used short materials and an approach similar to the F2C paradigm.

Previous research has shown that a dichotic “challenge” from another syllable can interfere with consonant identification in the target syllable and that errors are usually related to the place of articulation cue carried by the interferer (e.g., Studdert-Kennedy and Shankweiler, 1970; for a review, see Berlin and McNeil, 1976). Berlin *et al.* (1976) reported that dichotic interference can also be caused by an isolated formant, typically involving an initial transition followed by a steady portion (known as a “bleat”), derived from another syllable. This observation was investigated by Porter and Whittaker (1980) using a series of synthetic two-formant consonant-vowel (CV) syllables spanning the percepts [bae], [dae], and [gae]. They observed changes in initial stop perception that were related to the properties of the initial transition of an isolated F2 bleat presented in the contralateral ear. This suggests that, despite the strong spatial cue disfavoring it and even though a “clean” peripheral representation of the target syllable was available, some form of across-ear spectro-temporal averaging occurred. In accord with this idea, Porter and Whittaker (1980) found that errors in consonant identification were not only predictable from the position of the F2 bleat on the series but that their likelihood depended on bleat intensity relative to the CV syllable. Higher bleat levels led to more errors over the 24 dB range tested, presumably because the place cues carried by the target F2 became progressively more displaced or diluted.

What distinguishes Porter and Whittaker’s (1980) study from earlier demonstrations of across-ear integration of formant information in the perception of CV syllables (Rand, 1974; Cutting, 1976) is that the target formants were all in one ear and their contours were fully specified. Hence, the contralateral bleat was an interferer rather than a necessary part of the syllable, and so one might have expected it to be rejected from the target percept. Nonetheless, this across-ear integration may have been facilitated by the target CV and isolated bleat sharing the same source properties, as both were buzz-excited on the same fundamental frequency (F0). Although using a different F0 for the contralateral bleat might be expected to decrease the likelihood of integration, there is evidence that introducing a difference in F0 between formants has only a limited effect on the identification of synthetic CV syllables (Darwin, 1981; Gardner *et al.*, 1989) and a 4-semitone  $\Delta F_0$  does not prevent extraneous formants causing substantial IM when presented in the opposite ear to a target sentence (Summers and Roberts, 2020; see also Summers *et al.*, 2010, 2017). Therefore, the study reported here used a more radical difference in acoustic source properties—buzz-excited analogues of CV syllables were accompanied in the contralateral ear by a sine-wave analogue of the F2 bleat.

It has long been known that sine-wave analogues of speech can support intelligibility (Bailey *et al.*, 1977; Remez *et al.*, 1981) and that acoustic-phonetic information carried by tonal analogues of formants can be combined with that carried by buzz-excited analogues when that integration completes an intelligible percept. For example, the

identity of a synthetic stop-vowel syllable made ambiguous by removing its initial F3 transition can be resolved by replacing the missing transition with a tonal analogue (Whalen and Liberman, 1987; Bailey and Herrmann, 1993). Indeed, the integration of formants rendered as tonal with those rendered as buzz-excited can sustain a good level of intelligibility for sentence-length materials, whether the target formants are presented monaurally or dichotically (Roberts *et al.*, 2015; Summers *et al.*, 2016). The composite nature of these hybrid stimuli is apparent to listeners despite the integration of the acoustic-phonetic information carried by the different analogues (see Cutting, 1976). However, when there is competition between alternative candidates for F2—one buzz-excited and the other tonal, as used in the study reported here—there is evidence from experiments using sentence-length materials that the buzz-excited analogue is more influential perceptually than the tonal analogue (Roberts *et al.*, 2015; Summers *et al.*, 2016).

Experiments 1 and 2 reported here used an adapted version of Porter and Whittaker's (1980) method. The most critical difference was the use of a tonal analogue of the contralateral F2 bleat to minimize grouping cues favoring its integration with the buzz-excited CV syllable. If the initial transition of the bleat nonetheless influences the perception of consonant place in predictable ways, this would indicate that dichotic integration of F2 information is mandatory and would further suggest that intrusions into the target percept at the phonemic level are an important component of speech-on-speech IM. The experiments also extended Porter and Whittaker's method, which used only an /ae/ vowel, by examining the effects of F2 bleats in contrasting vowel contexts (low back versus high front—i.e., /a/ versus /i/); the slope and direction of an initial F2 transition is heavily context dependent and so using different vowels is useful to establish the generality of the findings. Finally, we used three-formant approximant-vowel syllables rather than two-formant stop-vowel syllables. With the simple synthesis model and constrained schematic formant patterns demanded by the experimental design, it proved easier to create more natural-sounding tokens of approximant consonants than of stops. Furthermore, approximants involve longer formant transitions than stops. Experiment 3 examined the extent to which place judgments could be supported by isolated sine bleats. Overall, these experiments found predictable effects that could be accounted for only if acoustic-phonetic information carried by the sine bleat was integrated with that carried by the target syllable. The mechanism for this mandatory dichotic integration of second-formant information—and the IM arising from it—remains to be established, but possible contributing factors are considered in Sec. V.

## II. EXPERIMENT 1

This experiment involved identification of the initial consonant of three-formant buzz-excited analogues of CV syllables, lying along a place of articulation series (/w-/l-/j/).

It was designed to determine whether place perception for a CV syllable presented to one ear was affected by the contralateral presentation of an F2 sine bleat derived from a different syllable, despite the grouping cues disfavoring the fusion of the syllable and the bleat—differences in ear of presentation and in source properties—and the use of CV tokens that included the initial F2 transition. To facilitate attention to the target syllables, listeners always received the CV token in the same ear—the left—and the sine bleat contralaterally in the right ear (see Porter and Whittaker, 1980). Previous research using dichotic formant ensembles did not find any discernible effects of which ear received which formants, either for CV syllables (Rand, 1974) or for sentence-length materials (Roberts *et al.*, 2010; Summers *et al.*, 2010). The critical comparison made was between responses to CV tokens accompanied by a matched sine bleat—i.e., one crafted using the same contours as the target F2—and responses to CV tokens accompanied by a different sine bleat, representing an alternative version of F2.

## A. Method

### 1. Listeners

All listeners, most of whom were students or members of staff at Aston University, received either course credits or payment for taking part. They were first tested using a screening audiometer (Interacoustics AS208, Assens, Denmark) to ensure that their audiometric thresholds at 0.5, 1, 2, and 4 kHz did not exceed 20 dB hearing level in either ear. All listeners who passed the audiometric screening took part in a training session designed to familiarize them with the synthetic CV syllables used and to ensure that their judgments met a pre-defined criterion when categorizing the place of articulation for the initial consonant (see Sec. II A 3). The same 16 listeners (two males) completed this experiment and experiment 2 (mean age = 25.6 years; range = 19.0–48.4). All listeners were native speakers of English (mostly British) and gave informed consent. The research was approved by the Aston University Ethics Committee.

### 2. Stimuli and conditions

Two series of CV syllables, each spanning three approximant categories (/w/, /l/, and /j/), were created using simplified (piecewise linear) formant-frequency contours; the vowel was either low back (/a/) or high front (/i/). The corresponding places of articulation are bilabial and velar for /w/, dental for /l/, and palatal for /j/. The stimulus parameters used were informed by extracting frequency and amplitude contours for the first three formants from tokens of six syllables (/wa/, /la/, /ja/, /wi/, /li/, and /ji/) spoken by a British male talker with “Received Pronunciation”. These contours were estimated from the waveform automatically every 1 ms from a 25 ms-long Gaussian window, using custom scripts in PRAAT (Boersma and Weenink, 2016). Stimuli were crafted such that all members of a series had common formant-amplitude contours and shared the same F1 and F3 frequency contours; tokens were distinguished

solely by the F2 frequency contour prior to the steady vowel portion. In real speech, more than one acoustic property distinguishes place of articulation in approximant-vowel syllables, but the properties of F2 prior to the steady portion usually carry the most salient cues and it was important for our purposes to constrain the information distinguishing approximant category. Hence, for our stimuli, the F2 frequency contour before the steady vowel portion acted as the sole disambiguating cue to consonant place in the CV stimulus.

To create the /wa/-/la/-/ja/ series, first a synthetic /la/ token was made using the smoothed natural /la/ amplitude contours and stylized F1, F2, and F3 frequency contours, based on those extracted from the /la/ stimulus. The F1–F3 frequencies used for the steady vowel portion of the stimulus were 650, 1090, and 2700 Hz, respectively, and the onset frequencies for the F1 and F3 transitions were 220 Hz and 3100 Hz, respectively. Satisfactory synthetic /wa/ and /ja/ tokens were then made by modifying only the shape of the F2 frequency contour of the /la/ stimulus prior to the final steady portion; the F1 and F3 frequency contours and all three amplitude contours were left unchanged. Finally, an 11-member series spanning from /wa/ to /ja/ via /la/ was crafted using a linear increase in the F2 onset frequency (800–2100 Hz in 130 Hz steps); the time after onset at which

the F2 frequency began to change and the duration of the F2 transition were adjusted as necessary to improve syllable perception based on informal listening. Tokens 1, 6, and 11 were crafted to be perceived as clear examples of /wa/, /la/, and /ja/, respectively. A similar process was used to create the /wi/-/li/-/ji/ series, again using a linear increase in the F2 onset frequency (500–2600 Hz in 210 Hz steps). The F1–F3 frequencies used for the steady vowel were 270, 2300, and 3000 Hz, respectively, and the onset frequencies for the F1 and F3 transitions were 340 and 3100 Hz, respectively. In describing the stimuli to listeners and in presenting the data, we judged it less confusing for those unfamiliar with phonetic notation to use the label [y] for the palatal approximant /j/. Accordingly, henceforth, the stimulus and response categories are given as [w], [l], and [y].

Figure 1 illustrates the sets of formant frequency and amplitude contours used to synthesize the two CV series. All tokens were synthesized using the appropriate frequency and amplitude contours to control three parallel second-order resonators corresponding to F1, F2, and F3 whose outputs were summed; the 3 dB bandwidths of these resonators were set to constant values of 90, 110, and 130 Hz, respectively. Following Klatt (1980), the outputs of the resonators were summed using alternating signs (+, −, +) to minimize spectral notches between adjacent formants in the same ear.

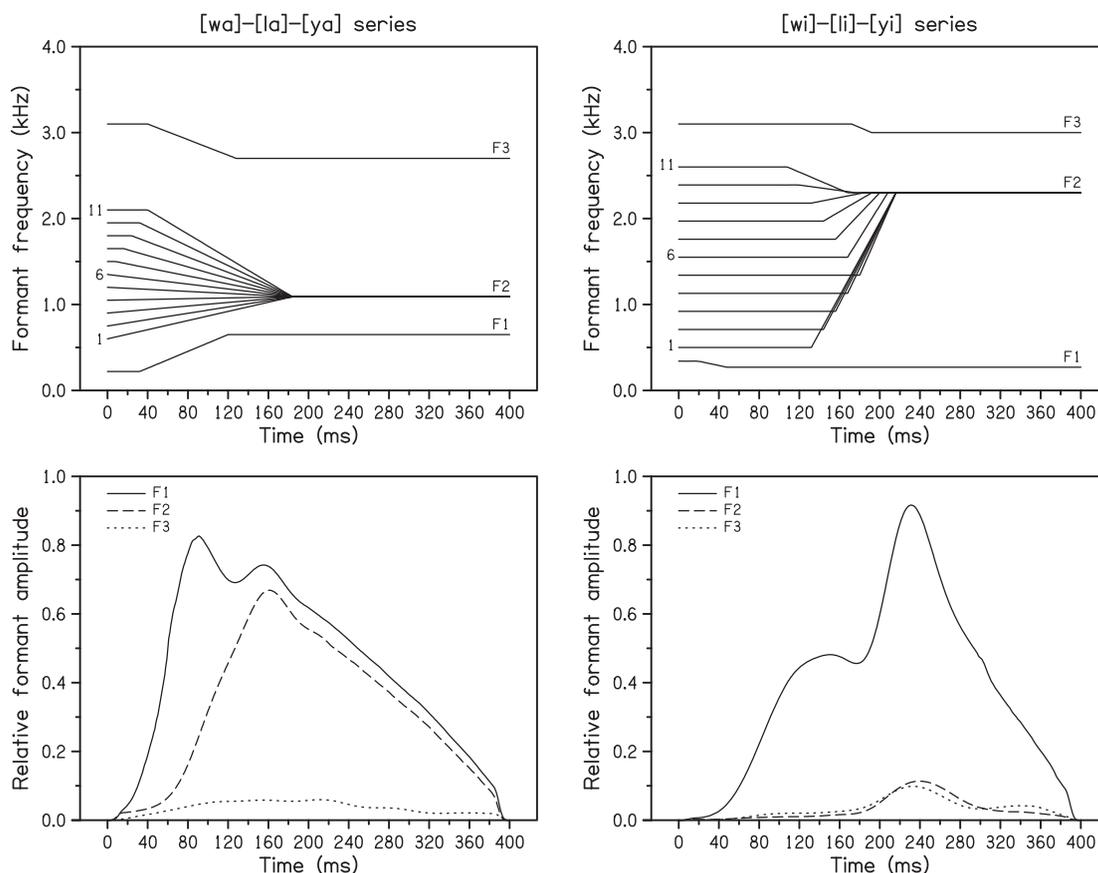


FIG. 1. Stimuli for experiment 1—schematics showing the formant-frequency contours (top row) and amplitude contours (bottom row) for F1-F3 used to create the two consonant-vowel series and corresponding sine bleats. The left- and right-hand panels illustrate the low-back and high-front vowel contexts, respectively. Note that tokens within each series were distinguished from one another only by their initial F2 transitions.

A monotonous source with an F0 of 130 Hz was used in the synthesis of all CV tokens; the excitation source was a periodic train of simple excitation pulses modeled on the glottal waveform (Rosenberg, 1971). For each series, the frequency and amplitude contours of F2 were also used to control the properties of a time-varying sinusoid, generating a set of tonal analogues of F2 that were matched to the root mean square (RMS) power of their buzz-excited counterparts. Hence, each CV token in the two series had a corresponding sine beat. All synthetic CV tokens and sine beats were 400 ms long.

There were four conditions in the main experiment. In all cases, the target stimulus was a CV syllable presented in the left ear, corresponding to one of the members of the chosen series, and it was accompanied by a sine beat in the right ear. Hence, there were no obvious cues to distinguish the experimental conditions from the reference case. In condition 1, the reference case, the frequency contour of the sine beat accompanying each CV token matched exactly that of the target F2 for the corresponding position in the 11-member series (matched beats). In conditions 2–4, the frequency contour of the sine beat accompanying each CV token always corresponded to the target F2 for token 1 ([w] sine beat), 6 ([l] sine beat), and 11 ([y] sine beat), respectively (mismatched beats). For these conditions, only 10 members of the series were tested; the one case in each condition for which the fixed sine beat matched its CV counterpart was not tested again. Example spectrograms illustrating

the matched- and mismatched-beat conditions are shown in Fig. 2. See the supplementary material for all stimuli and the parameters used to create them.<sup>1</sup>

### 3. Procedure

During testing, listeners were seated in front of a computer screen and a button pad in a double-walled sound-attenuating chamber (Industrial Acoustics 1201A, Winchester, UK). Participants were asked to listen carefully to the syllable and to judge the identity of the initial consonant, responding by pressing W, L, or Y on the button pad, as appropriate. No feedback on their responses was given. The experiment consisted of an initial session, in which each CV token was presented diotically and without any accompanying sine beat, and a four-condition main session run on a different day in which each CV token was presented in the left ear and accompanied by a contralateral sine beat. Trials were blocked by series (vowel = low back or high front) in both sessions; half the listeners began with the low-back CV series. Each block comprised two short familiarization tasks followed by the main task; listeners were free to take short breaks between the component parts of each block. Whenever responses were required (see below), there was a 0.5 s pause after the response before the next stimulus was presented.

The initial session used diotic presentation throughout and each block began with the familiarization tasks.

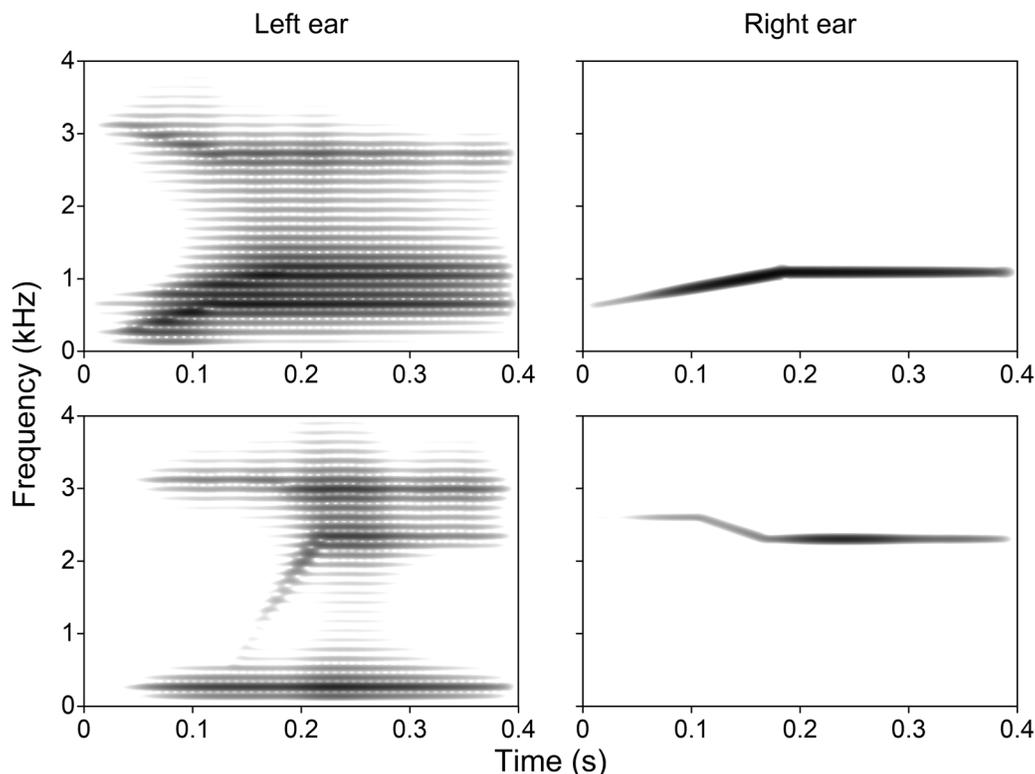


FIG. 2. Stimuli for experiment 1—narrowband spectrograms illustrating conditions in which the sine beat was either matched or mismatched with F2 of the CV token. The upper pair of panels show a matched example, in which a buzz-excited [wa] token (position 1 in low-back vowel context) in the left ear was accompanied by the corresponding sine beat in the right ear. The lower pair of panels show a mismatched example, in which a buzz-excited [wi] token (position 1 in high-front context) in the left ear was accompanied by a different sine beat (position 11, corresponding to [yi]) in the right ear.

Listeners first heard the sequence of CV series members in the correct order (1–11); they were asked to listen to the sequence (repeating if desired) but were not asked to make any responses. Listeners then heard 10 repetitions of three clear exemplars of the approximant place categories (tokens 1, 6, and 11) for the selected series (i.e., 30 trials), presented in random order, and were asked to identify the initial consonant. The main task comprised 11 repetitions of all 11 members of the series, for which listeners were again asked to identify the initial consonant. A new randomization of trial order was used for each repetition and listener. Given that these trials included stimuli with intermediate and ambiguous place cues, the first 11 trials were treated as practice and the results were discarded. Hence, the identification functions obtained for each listener were based on 110 trials per series. Listeners were invited back for the main session only if they met criterion performance for identifying the diotic CVs, which was defined as responding [w], [l], and [y] at least eight times out of 10 for tokens 1, 6, and 11, respectively, for both series.

The main session began with the same familiarization tasks, except that monaural presentation was used; sine bleats were not present during those tasks. After familiarization, the same procedure was followed during the main session as for the initial one, except that there were four conditions—the stimuli from which were intermingled randomly within each block (either low-back or high-front contexts)—and dichotic presentation was used. Listeners were told that the syllable would be presented in the left ear and that it would be accompanied in the right ear by a tonal sound; they were asked to try to focus their attention on the syllable and to ignore the tonal sound when making their judgment of the initial consonant. By always presenting the target syllable in the same ear and instructing listeners to focus on that ear (i.e., no spatial uncertainty), and by rendering the contralateral bleat with radically different source properties (tonal analogue), the aim was to provide listeners with the best opportunity to exclude the sine bleat from their perception of the target syllable. The identification functions obtained for each listener were based on 110 trials for condition C1 and 100 trials for each of conditions C2–C4; the data for positions 1, 6, and 11 in C1 were included to complete the identification functions for C2, C3, and C4, respectively. The main session typically took ~70 min to complete.

All speech analogues were synthesized using MITSYN (Henke, 2005) at a sample rate of 20 kHz and with 10 ms raised-cosine onset and offset ramps. They were played at 16 bit resolution over Sennheiser HD 480–13II earphones (Hannover, Germany) via a Sound Blaster X-Fi HD sound card (Creative Technology Ltd., Singapore), a pair of programmable attenuators (Tucker-Davis Technologies TDT PA5, Alachua, FL), and a headphone buffer (TDT HB7, Alachua, FL). Output levels were calibrated using a sound-level meter (Brüel and Kjaer, type 2209, Nærum, Denmark) coupled to the earphones by an artificial ear (type 4153). All CV tokens were presented at 72 dB sound pressure level

(SPL); the level in the ear receiving the sine bleat was lower but varied across stimuli, depending on the RMS power of the target F2 from which the sine bleat was derived.

#### 4. Data analysis and availability

For judgments of the initial diotic CVs, analysis was limited mainly to computing the identification functions for the three approximant categories for each listener, ensuring that they met criterion performance, and averaging the results across those listeners. For the main experiment, identification functions were computed for each condition, and changes across conditions in the centroid and area of the three category responses were used to assess changes in place judgments across conditions and, hence, to evaluate the impact of the different sets of sine bleats used. These measures have the virtue of being derived from responses to all stimuli in the series. The centroid for a given response category function was computed from the number of responses in that category,  $r$ , at each position (step #, 1–11) along the series,  $i$ :

$$centroid = \frac{\sum_{i=1}^{11} i \times r_i}{\sum_{i=1}^{11} r_i}$$

Changes in centroid ( $\Delta$ centroid) are expressed as changes in position along the series in step units. The area of a given identification function corresponds to the mean proportion of responses in that category when averaged across all positions in the series. Changes in area ( $\Delta$ area) are expressed as changes in percentage points (% pts).

All statistical analyses were computed using R 3.6.3 (R Core Team, 2020) and the *ez* analysis package (Lawrence, 2016). The measure of effect size reported here for repeated-measures analysis of variance (ANOVA) is partial eta squared ( $\eta_p^2$ ). For the analysis of changes in area, note that the main effects of condition and vowel context, and their interaction, are not presented because the values for these terms always sum to 100% across the three response categories ( $p = 1.0$ ). Hence, the effects of condition and vowel context on area are manifest only through their interactions with response category. All pairwise comparisons (two tailed) were computed using the restricted least-significant-difference test (Snedecor and Cochran, 1967; Keppel and Wickens, 2004). The research data underlying this publication are available on-line from a repository hosted by Aston University.<sup>2</sup>

#### 5. Predictions

If the F2 contour of a contralateral sine bleat is at least partly integrated with the F2 contour of a CV token, through some form of spectro-temporal averaging in the central auditory system, then sine bleats should have predictable effects on judgments of place of articulation. The predicted

effects of the fixed mismatched bleats on the centroid and area of the consonant identification functions in both vowel contexts, relative to the reference case (C1), were as follows: For the area data, accompanying the CVs with a given fixed sine bleat was expected to increase the area of the identification function for the corresponding place category at the expense of the other categories. For the centroid data, accompanying the CVs with fixed sine bleat 1 (C2) or 11 (C4) was anticipated to increase the probability of [w] or [y] responses, respectively, and so to move the centroids for all response categories rightwards or leftwards along the position axis. Accompanying the CVs with fixed sine bleat 6 (C3) was anticipated to increase the probability of [l] responses and so to move the centroids for [w] and [y] responses leftwards and rightwards, respectively, displacing them away from the middle of the position axis; the centroid for the [l] responses was expected to remain unchanged.

**B. Results and Discussion**

Figure 3 shows the results for identification of the diotic CV tokens presented alone in low-back (upper panel) and high-front (lower panel) vowel contexts, respectively. For each response category, the mean proportions and inter-subject standard errors are shown for each position along the series; the mean centroid and inter-subject standard error for each identification function (indicating values along the x-axis) are also shown. Listeners produced clear and systematic patterns of place judgments in both vowel contexts, progressing smoothly from mainly [w] responses through mainly [l] responses to mainly [y] responses with increasing position number. In both vowel contexts, the response profiles obtained are in accord with the notion that the F2 contours corresponding to positions 1, 6, and 11 were highly effective, when presented in syllabic context, at carrying acoustic-phonetic information that supported the perception of approximants with bilabial and velar ([w]), dental ([d]), and palatal ([y]) places of articulation, respectively. This outcome supports the choice of fixed sine bleats modeled on these F2 contours for the mismatched-bleat conditions.

Figure 4 shows the results for the main experiment, in which monaurally presented CV tokens were accompanied in the contralateral ear by a sine bleat that matched the target F2 (condition 1, matched case) or by a sine bleat that always corresponded to the target F2 for token 1 ([w] sine bleat), 6 ([l] sine bleat), or 11 ([y] sine bleat), respectively (conditions 2–4, mismatched cases). Parts (a) and (b) display the results for low-back and high-front vowel contexts, respectively. For each part, the results for conditions 1–4 are presented in descending rows. The main panel in each row shows for each response category the mean proportions and inter-subject standard errors for each position along the series; the mean centroid and inter-subject standard error for each identification function are also shown. For conditions 2–4, the mean proportions for condition 1 are reproduced (dotted lines) to facilitate comparison. In addition, the two small panels to the right of the main panel show for each

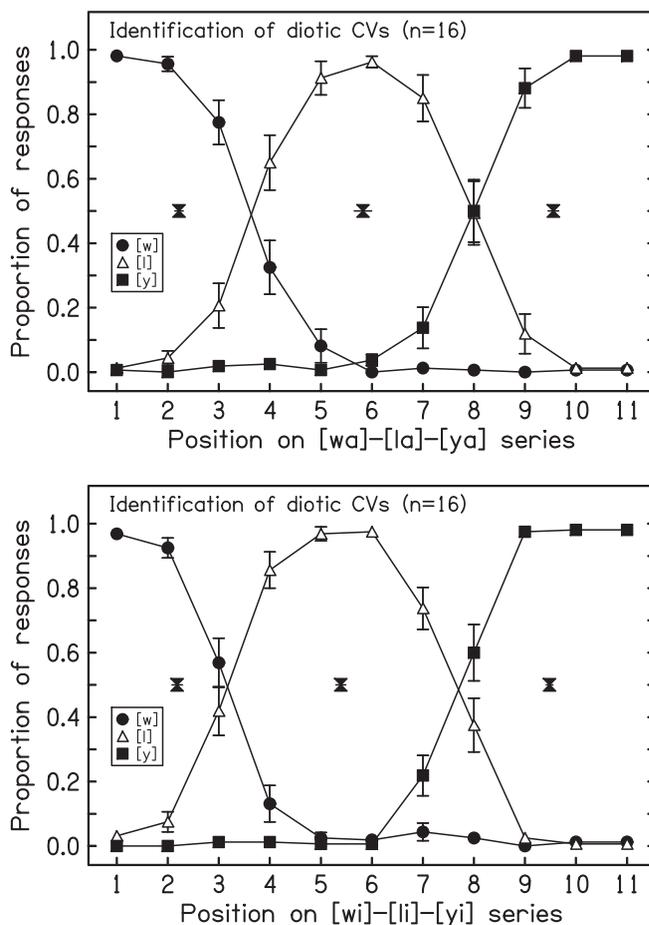


FIG. 3. Results for experiment 1—categorization of diotic CV tokens. Upper and lower panels display the results for the low-back and high-front vowel contexts, respectively. Each panel shows for each response category (see inset) the means and inter-subject standard errors (n=16) for each position along the series; the mean centroid and inter-subject standard error, indicating values along the x-axis, are also shown for each response category (hourglass symbols).

response category the changes in centroid and area relative to condition 1.

Inspection of Fig. 4 suggests that the sine bleats influenced listeners' judgments of place despite their contralateral presentation, the difference in source properties from the CV tokens, and the fact that each CV token already included a well-specified F2. To explore whether these effects were significant, we began by conducting three-factor repeated-measures ANOVAs—condition (C) × vowel context (V) × response category (R)—on the centroid and area data; the statistical outcomes are presented in Tables I and II, respectively. For the centroid data, there was a highly significant main effect of condition ( $p < 0.001$ ) and all interactions of condition with vowel context and response category were also significant (range:  $p = 0.038$  to  $p < 0.001$ ); the V × R interaction was not significant ( $p = 0.396$ ). The significant main effect of response category ( $p < 0.001$ ) is a trivial outcome reflecting large differences in centroid between the three identification functions; the marginal effect of vowel context ( $p = 0.051$ ) is also trivial. For the area data, the two interaction terms involving condition

Part (a): low-back vowel context

Part (b): high-front vowel context

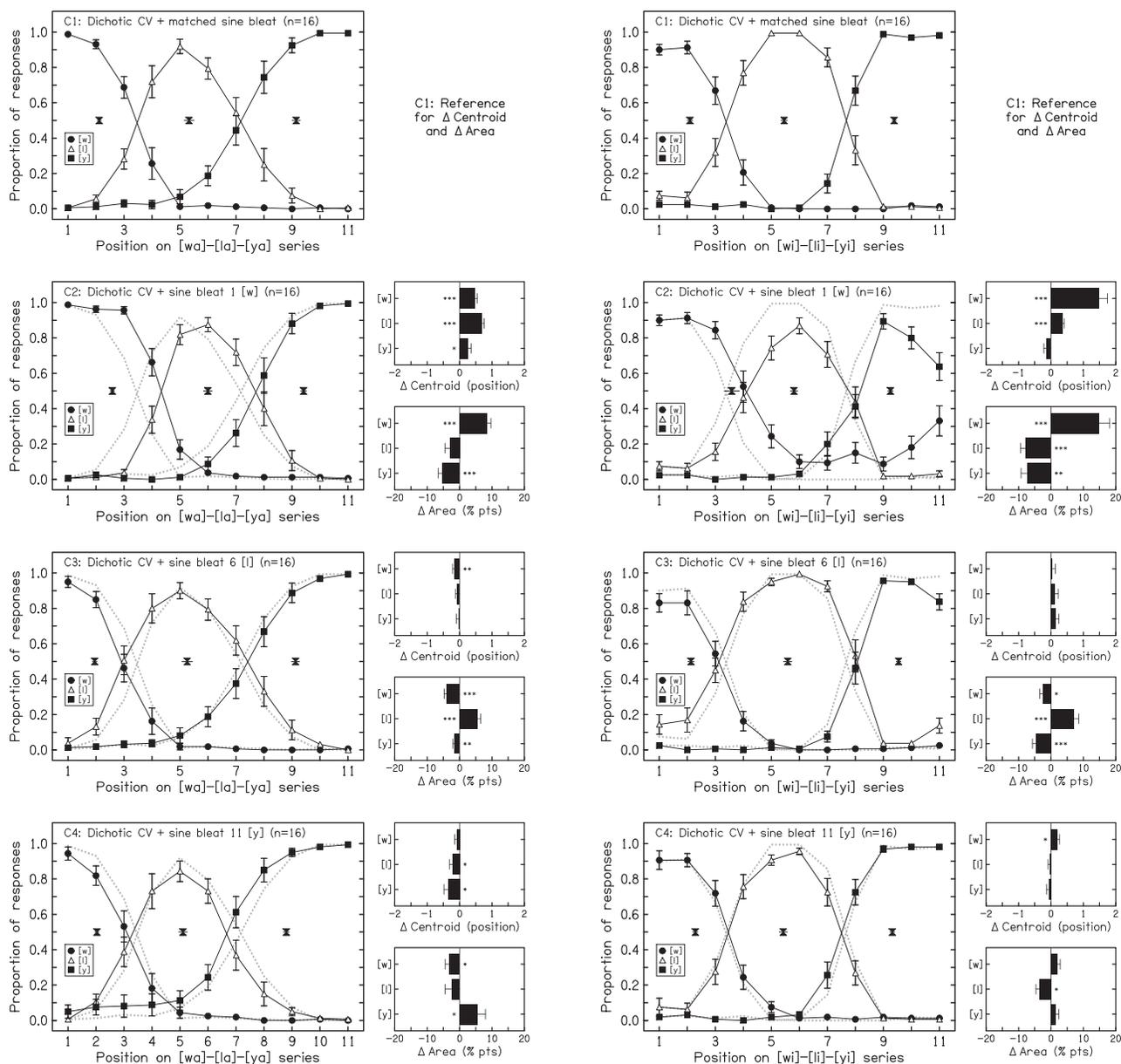


FIG. 4. Results for experiment 1—effects of mismatched contralateral sine bleats on the perception of consonant place. Parts (a) and (b) display the results for the low-back and high-front vowel contexts, respectively. For each part, the results for conditions 1-4 are presented in descending rows. The main panel in each row shows for each response category (see inset) the means and inter-subject standard errors ( $n=16$ ) for each position along the series; the mean centroid and inter-subject standard error, indicating values along the x-axis, are also shown for each response category (hourglass symbols). For conditions 2-4 (fixed mismatched sine bleats), the mean results for condition 1 are reproduced (reference case, gray dotted lines) for comparison; there are also two additional panels to the right of the main panel showing for each response category the change in centroid and area relative to the reference case (matched sine bleats, condition 1). Significance levels for these changes are indicated using asterisks (\* = <5%, \*\* = <1%, \*\*\* = <0.1%).

( $C \times R$  and  $C \times V \times R$ ) were significant (range:  $p = 0.025$  to  $p < 0.001$ ); the  $V \times R$  interaction was also significant ( $p = 0.015$ ). Once again, the significant main effect of response category ( $p = 0.001$ ) is trivial, in this case arising from differences in the overall proportion of responses made in each of the three categories. Taken together, these outcomes demonstrate that listeners' place judgments were influenced by condition in the two vowel contexts but also that the extent of these effects differed across contexts.

More detailed investigation was conducted using pairwise comparisons of the centroid and area measures for conditions 2-4, relative to condition 1, in the context of the *a priori* predictions (see Sec. II A 5) made about the likely impacts of the different mismatched sine bleats on consonant place judgments if their effect was to corrupt the perceptual estimation of  $F_2$  for the CV tokens. Tables III and IV present centroid and area measures for each response category and assess the significance of changes in those

TABLE I. Results for experiment 1—centroids. Effects of condition (CV syllable accompanied by a matched or mismatched sine bleat) and vowel context on the centroids of the responses to the three place categories. Summary of the three-way repeated-measures ANOVA; all significant terms are shown in bold.

Factor	<i>df</i>	<i>F</i>	<i>p</i>	$\eta_p^2$
<b>Condition (C)</b>	<b>(3, 45)</b>	<b>45.572</b>	<b>&lt;0.001</b>	<b>0.752</b>
Vowel context (V)	(1, 15)	4.506	0.051	0.231
<b>Response category (R)</b>	<b>(2, 30)</b>	<b>2475.927</b>	<b>&lt;0.001</b>	<b>0.994</b>
<b>C × V</b>	<b>(3, 45)</b>	<b>3.042</b>	<b>0.038</b>	<b>0.169</b>
<b>C × R</b>	<b>(6, 90)</b>	<b>17.896</b>	<b>&lt;0.001</b>	<b>0.544</b>
V × R	(2, 30)	0.957	0.396	0.060
<b>C × V × R</b>	<b>(6, 90)</b>	<b>14.980</b>	<b>&lt;0.001</b>	<b>0.500</b>

measures relative to the reference case (C1). Inspection of Fig. 4 and Tables III and IV shows that the results obtained were mainly in accord with the predictions made; see Table V for a summary of the predictions and results.

Changes in centroid relative to C1 were predicted for eight of the nine pairwise comparisons. For the low-back vowel context, significant changes occurred in six of these eight cases, all of them in the predicted direction. As predicted, the remaining case showed no significant change. The greatest overall effect was obtained when the CV tokens were accompanied by the [w] sine bleat (position 1, C2); for the [l] responses, the rise in mean centroid approached 0.7 steps. The outcome was less clear for the high-front vowel context. Only three of the eight cases for which there were predicted shifts in centroid were associated with significant changes; as predicted, the remaining case showed no significant change. The larger two of the three significant changes were in the direction predicted and once again occurred in C2; the rise of almost 1.5 steps for the [w] responses was the largest change in mean centroid found in the experiment. The other significant change occurred for the [w] response category in C4 but was not in the predicted direction. The change was, however, relatively small (<0.2 steps). Changes in area for each response category were predicted for all nine pairwise comparisons. All changes in area were in the predicted direction and, in each vowel context, these changes were significant in seven of the nine cases; the greatest area changes in the low-back (8.4% pts) and high-front contexts (14.9% pts) both occurred in C2 (fixed sine bleat 1) and represent increases in [w] responses.

TABLE II. Results for experiment 1—area under the identification functions. Effects of condition (CV syllable accompanied by a matched or mismatched sine bleat) and vowel context on the area under the identification functions for the three place categories. Summary of the three-way repeated-measures ANOVA; all significant terms are shown in bold. The main effects of condition and vowel context, and their interaction, are not included because the values for these terms always sum to 100% ( $p = 1.0$ ).

Factor	<i>df</i>	<i>F</i>	<i>p</i>	$\eta_p^2$
<b>Response category (R)</b>	<b>(2, 30)</b>	<b>8.625</b>	<b>0.001</b>	<b>0.365</b>
<b>C × R</b>	<b>(6, 90)</b>	<b>25.809</b>	<b>&lt;0.001</b>	<b>0.632</b>
<b>V × R</b>	<b>(2, 30)</b>	<b>4.859</b>	<b>0.015</b>	<b>0.245</b>
<b>C × V × R</b>	<b>(6, 90)</b>	<b>2.544</b>	<b>0.025</b>	<b>0.145</b>

TABLE III. Results for experiment 1—centroids. Effects of condition (CV syllable accompanied by a matched or mismatched sine bleat) and vowel context on the centroids of the responses to the three place categories, relative to the reference case (C1). Summary of pairwise comparisons; all significant cases are shown in bold.

A: Results for low-back vowel context					
Condition	Response category	Centroid (step #)	$\Delta$ centroid (steps)	<i>t</i> (15)	<i>p</i>
C1	[wa]	2.115	–	–	–
C1	[la]	5.325	–	–	–
C1	[ya]	9.146	–	–	–
<b>C2</b>	<b>[wa]</b>	<b>2.587</b>	<b>0.472</b>	<b>6.008</b>	<b>&lt;0.001</b>
<b>C2</b>	<b>[la]</b>	<b>5.996</b>	<b>0.671</b>	<b>7.239</b>	<b>&lt;0.001</b>
<b>C2</b>	<b>[ya]</b>	<b>9.411</b>	<b>0.265</b>	<b>2.938</b>	<b>0.010</b>
<b>C3</b>	<b>[wa]</b>	<b>1.954</b>	<b>–0.161</b>	<b>2.961</b>	<b>0.010</b>
C3	[la]	5.257	–0.067	1.261	0.227
C3	[ya]	9.129	–0.017	0.197	0.847
C4	[wa]	2.034	–0.081	1.158	0.265
<b>C4</b>	<b>[la]</b>	<b>5.107</b>	<b>–0.217</b>	<b>2.495</b>	<b>0.025</b>
<b>C4</b>	<b>[ya]</b>	<b>8.798</b>	<b>–0.348</b>	<b>2.609</b>	<b>0.020</b>

B: Results for high-front vowel context					
Condition	Response category	Centroid (step #)	$\Delta$ centroid (steps)	<i>t</i> (15)	<i>p</i>
C1	[wi]	2.098	–	–	–
C1	[li]	5.458	–	–	–
C1	[yi]	9.387	–	–	–
<b>C2</b>	<b>[wi]</b>	<b>3.589</b>	<b>1.491</b>	<b>5.841</b>	<b>&lt;0.001</b>
<b>C2</b>	<b>[li]</b>	<b>5.811</b>	<b>0.353</b>	<b>5.751</b>	<b>&lt;0.001</b>
C2	[yi]	9.253	–0.134	1.509	0.152
C3	[wi]	2.131	0.033	0.329	0.747
C3	[li]	5.584	0.126	1.272	0.223
C3	[yi]	9.544	0.156	1.718	0.106
<b>C4</b>	<b>[wi]</b>	<b>2.286</b>	<b>0.188</b>	<b>2.656</b>	<b>0.018</b>
C4	[li]	5.422	–0.036	0.683	0.505
C4	[yi]	9.322	–0.066	0.870	0.398

Changes in the identification of place of articulation arising from adding mismatched contralateral sine bleats were usually greatest for the most ambiguous CV tokens on each series, corresponding to positions 3–4 and 7–8. Nonetheless, changes in identification were sometimes considerable even for apparently unambiguous tokens. In particular, note the effect of adding fixed sine bleat 1 on responses to positions 5, 10, and 11 of the [wi]-[li]-[yi] series (C2). In the reference case (C1), responses to these positions were almost exclusively (>99%) either [l] (position 5) or [y] (positions 10 and 11), but on average, responses to these categories fell by 25.3% pts for these positions in C2 and these losses were almost exclusively to category [w]. Furthermore, responses to these positions were also unambiguous in the initial diotic session and so we can be confident that the initial F2 transition of the syllable fully specified consonant identity in the absence of a matched sine bleat. Evidently, there are at least some circumstances in which a mismatched contralateral sine bleat can influence the perception of place for a CV syllable that would

TABLE IV. Results for experiment 1—area under the identification functions. Effects of condition (CV syllable accompanied by a matched or mismatched sine bleat) and vowel context on the area under the identification functions for the three place categories, relative to the reference case (C1). Summary of pairwise comparisons; all significant cases are shown in bold.

A: Results for low-back vowel context					
Condition	Response category	Area (%)	$\Delta$ area (% pts)	$t(15)$	$p$
C1	[wa]	26.53	–	–	–
C1	[la]	33.18	–	–	–
C1	[ya]	40.28	–	–	–
<b>C2</b>	<b>[wa]</b>	<b>34.89</b>	<b>8.35</b>	<b>6.576</b>	<b>&lt;0.001</b>
C2	[la]	30.17	–3.01	2.014	0.062
<b>C2</b>	<b>[ya]</b>	<b>34.94</b>	<b>–5.34</b>	<b>4.474</b>	<b>&lt;0.001</b>
<b>C3</b>	<b>[wa]</b>	<b>22.50</b>	<b>–4.03</b>	<b>5.614</b>	<b>&lt;0.001</b>
<b>C3</b>	<b>[la]</b>	<b>38.75</b>	<b>5.57</b>	<b>6.078</b>	<b>&lt;0.001</b>
<b>C3</b>	<b>[ya]</b>	<b>38.75</b>	<b>–1.53</b>	<b>3.010</b>	<b>0.009</b>
<b>C4</b>	<b>[wa]</b>	<b>23.35</b>	<b>–3.18</b>	<b>2.567</b>	<b>0.021</b>
C4	[la]	30.85	–2.33	1.099	0.289
<b>C4</b>	<b>[ya]</b>	<b>45.80</b>	<b>5.51</b>	<b>2.245</b>	<b>0.040</b>

B: Results for high-front vowel context					
Condition	Response category	Area (%)	$\Delta$ area (% pts)	$t(15)$	$p$
C1	[wi]	24.77	–	–	–
C1	[li]	40.28	–	–	–
C1	[yi]	34.94	–	–	–
<b>C2</b>	<b>[wi]</b>	<b>39.72</b>	<b>14.94</b>	<b>4.737</b>	<b>&lt;0.001</b>
<b>C2</b>	<b>[li]</b>	<b>32.56</b>	<b>–7.73</b>	<b>4.549</b>	<b>&lt;0.001</b>
<b>C2</b>	<b>[yi]</b>	<b>27.73</b>	<b>–7.22</b>	<b>3.618</b>	<b>0.003</b>
<b>C3</b>	<b>[wi]</b>	<b>22.33</b>	<b>–2.42</b>	<b>2.447</b>	<b>0.027</b>
<b>C3</b>	<b>[li]</b>	<b>47.39</b>	<b>7.10</b>	<b>4.760</b>	<b>&lt;0.001</b>
<b>C3</b>	<b>[yi]</b>	<b>30.28</b>	<b>–4.66</b>	<b>4.333</b>	<b>0.001</b>
C4	[wi]	26.65	1.88	1.934	0.072
<b>C4</b>	<b>[li]</b>	<b>36.82</b>	<b>–3.47</b>	<b>2.915</b>	<b>0.011</b>
C4	[yi]	36.53	1.59	1.962	0.069

otherwise be judged as unambiguous. This outcome is consistent with the finding by Porter and Whittaker (1980) that mismatched F2 bleats influenced judgments of ambiguous and unambiguous targets in the same way, which they

interpreted as reflecting a primarily pre-phonetic locus for the dichotic interaction.

Overall, the pattern of results clearly supports the hypothesis that perceptual estimation of F2 for the CV tokens was influenced by the frequency contours of the contralateral sine bleats, but there are some interesting differences in the size of these effects on centroid and area, particularly between vowel contexts, that merit consideration. Overall, the largest effects were obtained from the [wi] sine bleat (C2), which involved the fastest rate and largest extent of frequency change during the initial F2 transition, and the smallest effects were obtained from the [yi] sine bleat (C4), which involved the slowest rate and smallest extent of change during the transition. Indeed, the difference in the velocity and extent of the initial F2 transitions for [wi] and [yi] is even greater on a semitone scale than is suggested by the linear frequency scale typical of speech spectrograms and used for the schematic presented in Fig. 1. This outcome suggests that the velocity and extent of frequency variation in the initial part of the sine bleat was important for their impact on consonant identity; see Sec. V for further discussion.

The final aspect of the results that merits comment concerns the extent to which the presence of a *matched* F2 sine bleat changed judgments of consonant place. Figure 5 replots the results for the matched sine-bleat condition (C1) and the initial identification of diotic CVs (no sine bleats present) to facilitate their comparison. The two sets of stimuli were tested in different sessions and stimulus contexts, which is likely to have had some impact on listeners' responses, but the observed differences in patterns of responses to the diotic and C1 stimuli were relatively large in low-back vowel context. This was especially apparent for category [l], for which the pattern of responses showed a decrease in the probability of [l] responses when matched sine bleats were present, notably for series positions 6–8, with [l] responses losing out mainly to [y] responses. Specifically, the area of the identification function for [l] was reduced by 5.7% pts, and the location of the [l]/[y] category boundary (as defined by the mean point of crossover for responses to those categories) shifted leftwards by 0.82 steps. These differences are similar in

TABLE V. Results for experiment 1—summary of predictions and outcomes for the centroid and area measures for each response category in conditions 2–4, relative to condition 1 (reference case). The extent to which the outcomes were in accord with the predictions is indicated for the two vowel contexts—low back (LB) and high front (HF)—by symbols ✓ (significant change in predicted direction or no change if none predicted), – (no significant change, unless none predicted), and x (significant change in the opposite direction).

Condition and bleat #	Response category	Predicted direction of change in centroid	Outcomes		Predicted change in area	Outcomes		
			LB	HF		LB	HF	
C2	1	[w]	Rightwards	✓	✓	Increase	✓	✓
C2	1	[l]	Rightwards	✓	✓	Decrease	–	✓
C2	1	[y]	Rightwards	✓	–	Decrease	✓	✓
C3	6	[w]	Leftwards	✓	–	Decrease	✓	✓
C3	6	[l]	No change	✓	✓	Increase	✓	✓
C3	6	[y]	Rightwards	–	–	Decrease	✓	✓
C4	11	[w]	Leftwards	–	x	Decrease	✓	–
C4	11	[l]	Leftwards	✓	–	Decrease	–	✓
C4	11	[y]	Leftwards	✓	–	Increase	✓	–

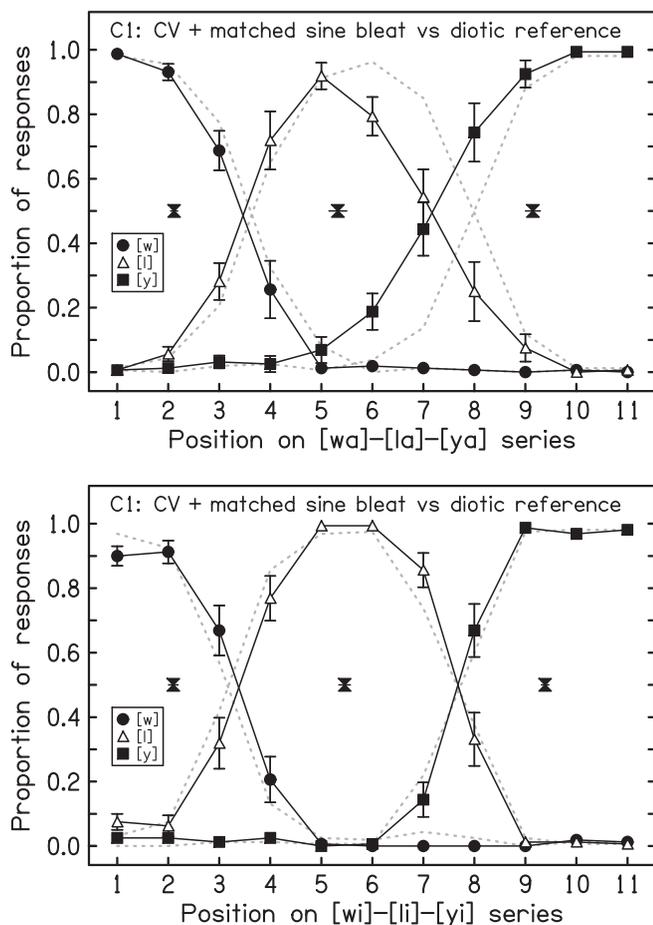


FIG. 5. Results for experiment 1—effects of matched contralateral sine bleats (condition 1) on the perception of consonant place relative to the diotic reference case. Upper and lower panels display the results for the low-back and high-front vowel contexts, respectively. Each panel reproduces the means and inter-subject standard errors for each position along the series in condition 1 (solid lines) and the mean results for the diotic reference case (gray dotted lines). The mean centroids and inter-subject standard errors for condition 1, indicating values along the x-axis, are also shown.

size to some of those observed between conditions in the main experiment and so merit further investigation.

### III. EXPERIMENT 2

This experiment examined whether the presence of an F2 sine bleat can affect the perception of place of articulation for a CV syllable when that bleat has the same frequency and amplitude contours as the corresponding target F2. This was done by ensuring that the stimuli being identified differed only in whether or not a contralateral sine bleat was present (monaural CV alone versus monaural CV plus matched sine bleat) and testing them by intermingling stimuli of the two types within a given vowel context. The results for experiment 2 confirmed that there are circumstances in which supposedly matched sine bleats nonetheless changed place judgments.

#### A. Method

Except where described, the same method was used as for experiment 1. There was no initial session involving

identification of diotic CVs; this experiment was run as a single session which typically took ~40 min to complete, including the familiarization tasks. There were only two conditions; once again, trials were blocked by series and conditions were intermingled within blocks. In condition 1, the target stimulus was a monaural CV syllable presented in the left ear, corresponding to one of the 11 members of the chosen series; there was no accompanying sound in the right ear. In condition 2, there was an accompanying sine bleat in the right ear and the frequency contour of the bleat matched exactly that of the target F2 for the corresponding position (matched bleats; same as condition 1 of experiment 1). As before, the level of the sine bleat was RMS-matched to its buzz-excited counterpart (target F2). Listeners were told that, regardless of whether or not the CV syllable was accompanied by a tonal sound, they should try to attend only to the left ear when making their judgments. Based on the results for experiment 1, it was anticipated that the presence of matched sine bleats would have a considerable effect on place judgments in low-back vowel context, particularly in the region of the [l]/[y] category boundary, but relatively little effect in high-front vowel context.

#### B. Results and Discussion

Figure 6 shows the results for the two conditions, one in which monaural CV tokens were presented alone (C1, reference case) and the other in which they were accompanied in the contralateral ear by the corresponding sine bleat (C2). The layout of Fig. 6 is the same as that used to present the main results for experiment 1 (see Fig. 4). Inspection of Fig. 6 suggests that accompanying CV tokens with matched sine bleats in the contralateral ear had relatively little impact on judgments of consonant place in the high-front vowel context, but that its impact was considerable in the low-back vowel context. Once again, we began by conducting three-factor repeated-measures ANOVAs—condition (C) × vowel context (V) × response category (R)—on the centroid and area data; the statistical outcomes are presented in Tables VI and VII, respectively. As for experiment 1, the significant main effects of response category on the two measures ( $p < 0.001$ ) were trivial outcomes. For the centroid data, there was a significant main effect of condition ( $p = 0.003$ ) and all interactions of condition with vowel context and response category were also significant (range:  $p = 0.016$  to  $p < 0.001$ ). For the area data, both interaction terms involving condition (C × R and C × V × R) were highly significant ( $p < 0.001$ ); the V × R interaction was not significant ( $p = 0.106$ ). Taken together, these outcomes demonstrate that the effect of condition on listeners' place judgments was influenced by vowel context.

Tables VIII and IX present the centroid and area measures for each response category in the two conditions; changes in those measures between C1 and C2 were explored using pairwise comparisons. For the low-back vowel context, inspection of Fig. 6 shows that the presence of the matched sine bleats considerably reduced the

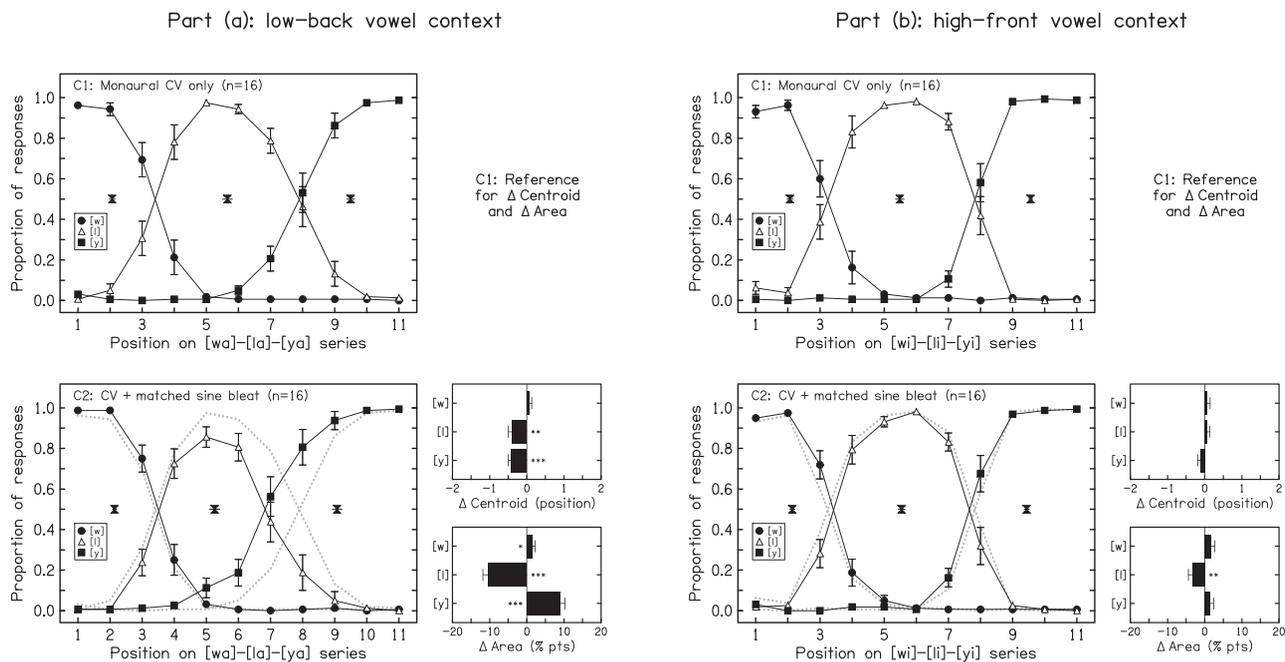


FIG. 6. Results for experiment 2—effects of matched contralateral sine bleats on the perception of consonant place. Parts (a) and (b) display the results for the low-back and high-front vowel contexts, respectively. For each part, the results for conditions 1 and 2 are presented in the top and bottom row. The main panel in each row shows for each response category (see inset) the means and inter-subject standard errors ( $n=16$ ) for each position along the series; the mean centroid and inter-subject standard error, indicating values along the x-axis, are also shown for each response category (hourglass symbols). For condition 2 (sine bleats present), the mean results for condition 1 are reproduced (reference case, gray dotted lines) for comparison; there are also two additional panels to the right of the main panel showing for each response category the change in centroid and area relative to the reference case (sine bleats absent, condition 1). Significance levels for these changes are indicated using asterisks (\* = <5%, \*\* = <1%, \*\*\* = <0.1%).

likelihood of [l] responses and that this reduction was asymmetrical, occurring primarily for series positions 5–9, and accompanied by increases in the probability of [y] responses. This change in place judgments was manifest as significant leftward shifts in the centroids for the [l] and [y] categories ( $\sim 0.4$  steps), and a significant loss of area for the [l] category (10.5% pts) with significant area gains for the other categories, particularly [y] (8.9% pts). The location of the [l]/[y] category boundary shifted leftwards by 1.06 steps, and the width of the [l] category was reduced (overall reduction = 1.18 steps). For the high-front vowel context, no significant changes in centroid were found for any response category. All changes in area were small and only the decrease in [l] responses (3.3% pts) was significant. This

loss was fairly symmetrical, leading to a slight narrowing (0.35 steps) of the identification function for the [l] category.

The comparison of place judgments for monaural CVs with those when the CVs were accompanied by matched sine bleats in the contralateral ear revealed a similar overall pattern to that observed in experiment 1 (comparison of initial judgments of diotic CV tokens with those in condition 1—monaural CV tokens accompanied by matched sine bleats). In both comparisons, the presence of matched sine bleats had a substantial effect in low-back vowel context but a fairly modest effect in high-front context, albeit also involving some loss of [l] responses. In low-back context, there was clear evidence of a considerable decrease in the probability of [l] responses, with corresponding increases in

TABLE VI. Results for experiment 2—centroids. Effects of condition (CV syllable alone or accompanied by a matched sine bleat) and vowel context on the centroids of the responses to the three place categories. Summary of the three-way repeated-measures ANOVA; all significant terms are shown in bold.

Factor	<i>df</i>	<i>F</i>	<i>p</i>	$\eta_p^2$
<b>Condition (C)</b>	<b>(1, 15)</b>	<b>12.704</b>	<b>0.003</b>	<b>0.459</b>
Vowel context (V)	(1, 15)	0.895	0.359	0.056
<b>Response category (R)</b>	<b>(2, 30)</b>	<b>4099.961</b>	<b>&lt;0.001</b>	<b>0.996</b>
<b>C × V</b>	<b>(1, 15)</b>	<b>7.355</b>	<b>0.016</b>	<b>0.329</b>
<b>C × R</b>	<b>(2, 30)</b>	<b>19.277</b>	<b>&lt;0.001</b>	<b>0.562</b>
V × R	(2, 30)	1.312	0.284	0.080
<b>C × V × R</b>	<b>(2, 30)</b>	<b>7.128</b>	<b>0.003</b>	<b>0.322</b>

TABLE VII. Results for experiment 2—area under the identification functions. Effects of condition (CV syllable alone or accompanied by a matched sine bleat) and vowel context on the area under the identification functions for the three place categories. Summary of the three-way repeated-measures ANOVA; all significant terms are shown in bold. The main effects of condition and vowel context, and their interaction, are not included because the values for these terms always sum to 100% ( $p = 1.0$ ).

Factor	<i>df</i>	<i>F</i>	<i>p</i>	$\eta_p^2$
<b>Response category (R)</b>	<b>(2, 30)</b>	<b>14.583</b>	<b>&lt;0.001</b>	<b>0.493</b>
<b>C × R</b>	<b>(2, 30)</b>	<b>38.714</b>	<b>&lt;0.001</b>	<b>0.721</b>
V × R	(2, 30)	2.426	0.106	0.139
<b>C × V × R</b>	<b>(2, 30)</b>	<b>16.042</b>	<b>&lt;0.001</b>	<b>0.517</b>

TABLE VIII. Results for experiment 2—centroids. Effects of condition (CV syllable alone or accompanied by a matched sine bleat) and vowel context on the centroids of the responses to the three place categories, relative to the reference case (C1). Summary of pairwise comparisons; all significant cases are shown in bold.

A: Results for low-back vowel context					
Condition	Response category	Centroid (step #)	$\Delta$ centroid (steps)	$t(15)$	$p$
C1	[wa]	2.069	–	–	–
C1	[la]	5.657	–	–	–
C1	[ya]	9.496	–	–	–
C2	[wa]	2.142	0.074	1.274	0.222
C2	[la]	5.261	<b>–0.396</b>	<b>4.048</b>	<b>0.001</b>
C2	[ya]	9.071	<b>–0.425</b>	<b>6.180</b>	<b>&lt;0.001</b>
B: Results for high-front vowel context					
Condition	Response category	Centroid (step #)	$\Delta$ centroid (steps)	$t(15)$	$p$
C1	[wi]	2.066	–	–	–
C1	[li]	5.486	–	–	–
C1	[yi]	9.562	–	–	–
C2	[wi]	2.139	0.073	0.957	0.354
C2	[li]	5.543	0.057	0.733	0.475
C2	[yi]	9.438	–0.124	1.812	0.090

the probability of other responses, primarily in category [y]. Indeed, the reduction in area for category [l] observed in this experiment, for which the two conditions were tested in the same context, was almost twice as large as that seen in experiment 1. From this outcome, it can be concluded that there are at least some circumstances in which a contralateral sine bleat can have a substantial influence on the

TABLE IX. Results for experiment 2—area under the identification functions. Effects of condition (CV syllable alone or accompanied by a matched sine bleat) and vowel context on the area under the identification functions for the three place categories, relative to the reference case (C1). Summary of pairwise comparisons; all significant cases are shown in bold.

A: Results for low-back vowel context					
Condition	Response category	Area (%)	$\Delta$ area (% pts)	$t(15)$	$p$
C1	[wa]	26.02	–	–	–
C1	[la]	40.68	–	–	–
C1	[ya]	33.30	–	–	–
C2	[wa]	27.61	<b>1.59</b>	<b>2.671</b>	<b>0.017</b>
C2	[la]	30.23	<b>–10.45</b>	<b>7.610</b>	<b>&lt;0.001</b>
C2	[ya]	42.16	<b>8.86</b>	<b>6.266</b>	<b>&lt;0.001</b>
B: Results for high-front vowel context					
Condition	Response category	Area (%)	$\Delta$ area (% pts)	$t(15)$	$p$
C1	[wi]	24.89	–	–	–
C1	[li]	41.59	–	–	–
C1	[yi]	33.52	–	–	–
C2	[wi]	26.59	1.70	1.695	0.111
C2	[li]	38.30	<b>–3.30</b>	<b>3.011</b>	<b>0.009</b>
C2	[yi]	35.11	1.59	1.792	0.093

perception of place, even when that bleat notionally specifies the same F2 properties as the target F2 in the CV token. Although Porter and Whittaker (1980) presented only a selection of their results in any detail, they made an observation consistent with this result. When unambiguous [dae] (position 6 on their 11-point series) was paired with bleats presented at the highest level tested, the highest proportion of [d] responses occurred when the target was paired with bleat 7 or 8 rather than with its own bleat.

Two aspects of these findings merit discussion. First, on what basis might the presence of a sine bleat notionally matched to the target F2 in the contralateral ear influence the perceptual estimation of F2? Second, what differences between vowel contexts might account for the larger and asymmetrical effects found for the [wa]-[la]-[ya] series? One reason why the two sources of F2 information may have combined in a way that changed the consonantal percept is because the contralateral sine bleat was protected from the partial masking to which the target F2 was exposed (primarily by F1). This was one of the factors shaping the acoustic-phonetic information carried by the two counterparts. Another is that presenting the sine bleat in isolation may have increased the salience of the F2 information it carries, which may be similar to the effect of increasing bleat intensity (Porter and Whittaker, 1980). Finally, there may not be full equivalence perceptually between a formant transition represented as a continuous glide (sine bleat) and one in which an acoustic resonance moved across a set of steady harmonics (CV token); this might affect, for example, the perception of the rate or even the direction of frequency change. However, this final possibility can provide at best only a partial account of the effects of adding matched sine bleats, given that even matched buzz-excited bleats can have similar effects (Porter and Whittaker, 1980).

There are two obvious differences between the two CV series. First, F2 amplitudes were considerably higher in low-back vowel context, such that the accompanying sine bleat in the contralateral ear was much more prominent than for the high-front context. Second, falling F2 transitions were dominant in low-back context (seven of 11 cases, positions 5–11) whereas rising F2 transitions were dominant in high-front context (nine of 11 cases, positions 1–9). These differences seem likely to have contributed to the greater impact of matched sine bleats in low-back vowel context, but precisely why is not immediately clear. Unfortunately, the different patterns of F1/F2 proximity and relative formant amplitude between the two vowel series make the partial masking hypothesis hard to assess as an account of the differences observed between vowel contexts. However, a speculative account of our results can be offered based on the typical acoustic characteristics of natural approximant-vowel syllables and on asymmetries in the perception of rising and falling formant transitions.

Spectrographic analysis of natural tokens of [la] (see, e.g., Potter *et al.*, 1947) shows that they typically begin with relatively low levels of mid-frequency energy followed by a low-rising second formant transition into the vowel. When

target syllables from the mid-range of the [wa]-[la]-[ya] series—in isolation identified as [la]—are accompanied by contralateral sine bleats, the additional early mid-frequency energy and flat or falling second formant implied by integration of information from the bleat may predispose listeners away from [la] towards a percept of [ya], as both of these properties are characteristic of natural tokens of /ja/. The substantially smaller effect of adding a contralateral bleat to target syllables from the [wi]-[li]-[yi] series is consistent with the lower levels of sine bleats derived from syllables with high-front vowels, but it is not obvious on the basis of acoustic–phonetic considerations why the effect is manifest as a small but broad reduction in the probability of [li] responses. It is noteworthy that the effect of matched sine bleats on [la] responses emerged for series positions  $\geq 5$ , corresponding to the point at which the direction of the F2 transition changed from rising to falling, raising the possibility that falling sine glides were more salient or were perceived as steeper than their target counterparts, biasing listeners towards [y] responses. Perhaps relevant here is the finding by Porter *et al.* (1991) that isolated F2 bleats with a falling initial transition were better discriminated than those with a rising transition, and that the just-noticeable-difference for increases in the rate of change were smaller for falling transitions. Whatever the reason for the differences between the two vowel contexts, the outcome of this experiment indicates that the matched-bleat condition in experiment 1 (C1) was a more appropriate reference case for the mismatched-bleat conditions (C2–C4) than would have been provided had monaural CVs been used instead.

#### IV. EXPERIMENT 3

This experiment explored the extent to which isolated sine bleats might support place judgments on their own; an ability to do so might arise because their trajectories inherently evoke percepts of particular consonants or because they acquire a link to those percepts through associative learning. The question of whether isolated sine bleats evoke consistent consonantal place percepts is important to investigate because a close mapping between listeners' judgments of isolated sine bleats and their judgments of CV tokens accompanied by those bleats would complicate the interpretation of experiments 1 and 2. In particular, it would make it hard to sustain the argument that listeners integrated the information carried by the two versions of F2 rather than merely responding directly to acoustic–phonetic information implied by the trajectories of the sine bleats.

The design of experiment 1 should have mitigated against learning a strong association between a particular sine bleat and a particular category, given that three of the four conditions involved sine bleats with different properties to the corresponding target F2. Therefore, for the majority of those trials, the accompanying sine bleat did not match the F2 contour of the target CV. That was not true, however, for experiment 2. Moreover, in neither case can it be ruled out *a priori* that the isolated sine bleats inherently evoked

particular consonantal place percepts. Three design features of experiment 3 were intended to minimize any further opportunities for associative learning. First, there was trial-by-trial unpredictability in stimulus type (CV alone or sine bleat alone) and the sine bleats represented only a small proportion of the trials. Second, as in experiments 1 and 2, no feedback on responses was provided, including for the isolated sine bleats. Third, one block of trials intermingled low-back syllables with high-front sine bleats and the other block was vice versa.

#### A. Method

Except where described, the same method was used as for experiments 1 and 2. Six listeners (1 male; mean age = 30.7 years, range = 21.8–48.4) took part, drawn as available from those who had completed experiments 1 and 2. This experiment involved only one condition and was run as a single session which typically took  $\sim 30$  min to complete. The form of blocking used for the trials here was adapted for the specific needs of this experiment. For each block, the set of stimuli used was the 11 CV tokens comprising the series for one vowel context plus the three archetypal sine-bleat stimuli (tokens 1, 6, and 11) drawn from the *other* series. The order of these blocks was alternated across successive listeners. Diotic presentation was used, rather than presenting CV syllables in the left ear and sine bleats in the right, to avoid the need for sudden changes in spatial attention that might disadvantage listeners on the occasional trials involving the isolated sine bleats. To increase the prominence of the isolated sine bleats, they were RMS matched with the overall level of the whole CV syllable rather than that of the corresponding target F2. This approach also removed the disparity in level that would otherwise have occurred for the sine bleats across vowel contexts (the relative amplitude of F2 was considerably lower in the high-front context). All CV and sine-bleat stimuli were presented at 72 dB SPL.

Once again, 11 blocks of trials were run and the first block was used as practice and discarded, giving 140 trials per series (110 trials for the CV identification functions and 30 trials for sine-bleat classification). Listeners were told that most of the stimuli would be like those heard during the familiarization tasks, but that they would sometimes hear tonal sounds which they should try to judge as best they could using the same category labels. Our decision to include relatively few but more distinctive sine-bleat tokens (archetypes only) was in response to the anticipated difficulty that listeners may have had in classifying them. The use of crossed contexts (i.e., CV tokens from the low-back context intermingled with sine bleats from the high-front context, and vice versa) was to reduce opportunities for associative learning of place cues between the CV tokens and the sine bleats during the experiment. This tactic was used because the large difference in F2 frequency for the steady portions in the two contexts affected the slope, and in some cases direction, of the initial F2 transitions, ensuring

that there was no simple correspondence between series steps for the syllables and sine bleats.

### B. Results and Discussion

No statistical analysis was performed on the data collected; instead, the results for the two types of stimuli (CV tokens and sine bleats) were compared descriptively across individual listeners. Figure 7 shows the results for each listener in a separate row; the left- and right-hand columns present the results for the low-back and high-front vowel contexts, respectively. Each panel shows the total number of responses out of 10 obtained in each response category at each position on the series for the CV tokens (open and black-filled symbols), and at positions 1, 6, and 11 for the archetypal sine bleats (gray symbols).

Despite some individual variation across listeners, responses to the CV tokens always progressed smoothly from mainly [w] responses through mainly [ɪ] responses to mainly [y] responses as position number increased, in both vowel contexts. In contrast, responses to the archetypal sine bleats were far less consistent across listeners, ranging from cases where the sine bleats were generally classified in a way consistent with the place judgments for the corresponding CV tokens (e.g., most listeners usually classified [a]-bleat 1 and [i]-bleat 1 as [w]) to cases where responses to the different sine bleats were near-random (e.g., L1 for the low-back context) to examples where the bleat classifications were inconsistent with responses to the corresponding CV (e.g., L1, L2, and L6 rarely responded [y] when presented with [i]-bleat 11). That judgments overall were not entirely random is consistent with previous research showing that identification of isolated rising or falling glides corresponding to the initial F3 transitions (chirps) for analogues of [da] or [ga] can show a greater than chance association with those places of articulation (Bailey and Herrmann, 1993).

A simple summary of the results was obtained by counting for each sine bleat the number of listeners who identified it as the consonant in the corresponding CV more often than either of the other two categories; this gave counts of 6 out of 6 listeners (6/6), 3/6, and 4/6 for [a]-bleats 1, 6, and 11, respectively, and 5/6, 0/6, and 2/6 for [i]-bleats 1, 6, and 11, respectively. Although this greater likelihood of reporting bleat 1 as [w] was sometimes associated with more systematic and consistent responses to the other bleats (see, e.g., the responses of L2 to the three [a]-bleats), this was not always the case. For example, [w] was the dominant response of L3 to all three [i]-bleats, accounting for ~73% of all responses. Note that the F2 transitions signaling [w] for CVs rose consistently in both vowel contexts, whereas [ɪ] and [y] could be signaled by rising or falling transitions and the balance of the two directions changed across vowel contexts. It seems likely that this is why listeners were better able to attach the correct place label to the [a]- and [i]-bleats for position 1 than for positions 6 and 11. Note also that [i]-bleat 6 was usually classified as [w], not [ɪ], presumably

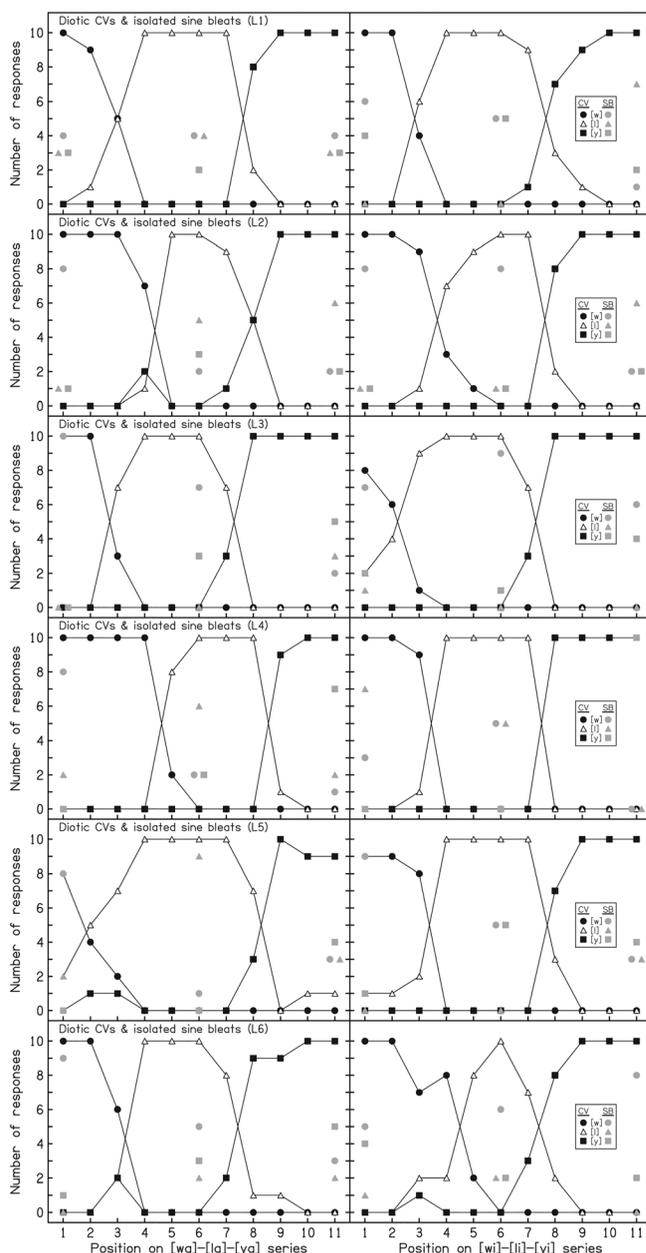


FIG. 7. Results for experiment 3—individual differences in judgments of place for isolated CV syllables and isolated sine bleats. Each pair of panels (left = low-back vowel context, right = high-front) shows the number of responses made in each category for each CV token (see inset, black symbols) and for the three archetypal sine bleats (gray symbols) from the same context; these bleats were in fact presented intermingled with CVs from the other context (see main text). For clarity, cases where responses to the sine bleats were the same in two categories are shown slightly displaced from one another. Note the relatively large variation between listeners in responses to the sine bleats compared with the corresponding CV tokens (positions 1, 6, and 11).

because the F2 transition cue for [ɪ] rose in this vowel context.

While it should be acknowledged that associative learning during experiment 2 cannot be ruled out, the variability and inconsistent pattern of responses to the archetypal sine bleats observed across listeners in experiment 3, relative to their counterparts in the CV series, suggest that any

tendency for isolated sine bleats inherently to evoke particular consonantal place percepts did not determine the outcomes of experiments 1 and 2. Overall, the results suggest that—although the sine bleats carried potentially useful acoustic–phonetic information—listeners did not routinely make place judgments based on the sine bleats alone. Rather, listeners combined information from the sine bleat with that carried by the buzz-excited formants comprising the CV token.

## V. GENERAL DISCUSSION

There are known to be limitations on the ability to listen with independent ears, even when it is advantageous to do so (e.g., Gallun *et al.*, 2007). The experiments reported here have provided evidence that judgments of place of articulation can be affected by intrusions of properties from an interfering F2 bleat in the contralateral ear, despite the presence of a complete CV syllable in the target ear (Whalen and Liberman, 1987; Bailey and Herrmann, 1993) and the radically altered source properties of the sine bleat (cf. Roberts *et al.*, 2015; Summers *et al.*, 2016). The latter outcome is striking given that several studies using non-speech sounds have shown that reducing target-masker similarity can be an effective way of reducing IM (e.g., Neff, 1995; Durlach *et al.*, 2003; see Kidd *et al.*, 2008). This finding, along with our use of approximant-vowel syllables and contrasting vowel contexts, extends the generality of Porter and Whittaker’s (1980) findings regarding the mandatory dichotic integration of second-formant information. Although there were differences between the outcomes for the two vowel contexts, our results confirm that the principal findings were not dependent on the particular pattern and frequency region of the formant-transition cue.

Experiment 1 showed that accompanying CV syllables with a fixed mismatched sine bleat in the opposite ear typically produced systematic and predictable changes in consonant judgments relative to the matched-bleat condition; these effects were reflected in changes in the centroid and area of each of the three response category functions. Experiment 2 compared judgments of place of articulation when monaural CV tokens were presented alone with when they were accompanied by matched sine bleats in the contralateral ear; it was shown that there are at least some circumstances in which even the presence of a matched sine bleat affects place judgments. Whatever the basis for this effect (e.g., because the sine bleat was more salient than the target F2 or because it was free from partial masking by F1), the result indicates that aspects of information from the contralateral bleat, even when notionally matched to the target F2, are inevitably included in the target categorization process. Experiment 3 showed that judgments of place of articulation based on listening to isolated sine bleats were often unsystematic or arbitrary, indicating that identifying the initial consonant in dichotic stimuli must involve integration of F2 information across the target CV and sine bleat. Taken together, these findings suggest that perceptual estimation of

F2 for the CV syllable involves some form of obligatory spectro-temporal averaging across ears (see also Roberts and Summers, 2019), such that the perceived place moves towards that implied by the sine bleat.

The locus of this dichotic integration is assumed to be pre-phonetic, because the presence of contralateral F2 bleats has similar effects on judgments of unambiguous and ambiguous CV syllables, indicating that the impact of the bleat depends more on the salience and “potency” of the interferer than on the ambiguity of the target (Porter and Whittaker, 1980). One factor that may facilitate or contribute to the dichotic integration of formant information is suggested by the so-called drone illusion (Deutsch, 1979). In this illusion, a familiar melody rendered unrecognizable by randomly allocating each note in the sequence to one or other ear becomes easier to identify if each target note is accompanied by a simultaneous “drone” note in the contralateral ear. The drone note does not change, and so provides no information about the melody, but its synchronous presentation with each target note appears to weaken the lateralization cues, reducing the tendency for target notes sent to the left and right ears to segregate from one another. In the context of the experiments reported here, note that the target F2 and sine bleat had the same duration and were presented simultaneously. Spectral overlap between the stimuli in the two ears may also have been a relevant factor (Darwin and Hukin, 2004). While it is true that not all IM scenarios necessarily involve limitations on the ability to listen independently with the two ears, the results presented here indicate that, when this is the case, the additional segregation cue provided by the difference in source properties between target and masker was insufficient to prevent their integration.

Bleat salience will be affected by various factors, such as presentation level, source properties, and protection from partial masking, but what determines the inherent potency of an interfering bleat? Our previous research using the F2C paradigm (or variants thereof) and sentence-length materials has shown that, within broad limits, increasing the depth or rate of formant-frequency variation in an extraneous formant increases the amount of IM it causes (Roberts *et al.*, 2014; Roberts and Summers, 2015, 2018, 2020; Summers *et al.*, 2012). Both manipulations increase the velocity of formant transitions in the interferer; increasing depth also extends the frequency range over which this variation takes place. Indeed, the effect of formant-frequency variation in the interferer appears to have two partially separable elements, one corresponding to the range over which the variation occurs and the other to the velocity of the transitions. Roberts and Summers (2020) explored this distinction by measuring the IM caused by various interferers constructed from segments of a frequency-varying extraneous formant. They found that interferers composed of segments that retained their frequency variation had a greater impact on intelligibility than those in which each segment was set to be locally constant.

The finding in the experiments reported here that the effect of mismatched contralateral sine bleats on judgments

of place was greatest for bleats involving the fastest and largest extent of frequency change, and least for bleats involving the slowest and smallest extent of frequency change, is in accord with the results of previous studies using the F2C paradigm to measure the impact of different interferers on mean overall intelligibility (e.g., Roberts and Summers, 2015; Summers *et al.*, 2012). This outcome is also consistent with the findings of the original dichotic challenge studies showing that another stop-vowel syllable (i.e., involving formant-frequency change) was an effective interferer but an isolated steady vowel was not (Berlin and McNeil, 1976). Formants with frequency-varying contours are known to be important for conveying acoustic–phonetic information (e.g., Stevens, 1998) and increases in formant velocity imply more rapid movement of the articulators. Indeed, peak velocity may be regarded as a measure of the degree of articulatory effort (e.g., Cheng and Xu, 2013). Presumably, mismatched sine bleats with more rapid transitions cause more IM because they are especially likely to interfere with the perceptual processing of the acoustic–phonetic information carried by the formants of the target speech. The results of the current study indicate that intrusion of specific properties of the interferer into the target percept constitutes an important component of this IM.

Another notable feature of the results reported here is the extent to which the dichotic integration of CV syllable and bleat appears resistant to the grouping cues favoring segregation. Furthermore, as noted earlier, Roberts *et al.* (2015) and Summers *et al.* (2016) found using composite sentence-length materials that buzz-excited analogues of F2 tended to be more influential perceptually when placed in competition with tonal ones, regardless of the source properties of F1 and F3. On the more specific point about competition, the two most obvious differences are that the current study used short materials and that the steady portions of the target F2 and the sine bleat were aligned because of their common (center) frequency. The relevance of the latter for dichotic integration might be explored by comparing the effects on place judgments of sine bleats with those of tonal analogues restricted to the initial part containing the critical transition (i.e., sine chirps). On the more general point about resistance to grouping cues, this aspect might be explored by attempting sequential capture of the sine chirp to reduce its impact (see Ciocca and Bregman, 1989) or attempting to segregate the sine chirp by attaching a seamless steady precursor to create an onset-time lead ( $\geq 300$  ms; see Darwin, 1981). Frequency transposition and asynchrony might also be used to explore the role of spectral and temporal alignment between the sine transition and the F2 transition in the target syllable (see Cutting, 1976). The importance in this context of formant velocity, as opposed to frequency range, might be explored by comparing the effects of standard sine glides with implied glides cued by a dynamic change in spectral center of gravity arising from smooth but rapid change in the relative level of two constant-frequency tones (see Fox *et al.*, 2008).

In conclusion, informational masking of speech does not result only from disruption of target processing, it can

also involve corruption of the acoustic–phonetic information that supports the target percept. The rate and extent of frequency change in the sine bleat appears to determine its impact on the identification of consonant place of articulation. This result may be another demonstration of the importance for speech-on-speech IM of the amount of frequency change in the masker (e.g., Roberts and Summers, 2015, 2018, 2020; Summers *et al.*, 2012). The approach taken here provides a useful tool that can be adapted to allow a systematic and fine-grained exploration of the factors that determine the likelihood of intrusions from the masker into the target syllable. These intrusions arise from mandatory dichotic integration—an inherent component of IM—whereby, despite grouping cues disfavoring this integration, the interferer alters the acoustic–phonetic information carried by the target formants in ways that are predictable from the properties of the interferer.

## ACKNOWLEDGMENTS

This research was supported by Grant No. ES/N014383/1 from the Economic and Social Research Council (UK), awarded to B.R. We are grateful to Adam Shephard for his assistance with data collection. Poster presentations on the first experiment were given at the 177th Meeting of the Acoustical Society of America (Louisville, KY, May 2019), and at the Basic Auditory Science Meeting of the British Society of Audiology (University College London, UK, September 2019).

<sup>1</sup>See supplementary material at <https://www.scitation.org/doi/suppl/10.1121/10.0007132> for all stimuli and the parameters used to create them.

<sup>2</sup>See <https://doi.org/10.17036/researchdata.aston.ac.uk.00000525> (Last viewed November 5, 2021).

- Bailey, P. J., and Herrmann, P. (1993). “A reexamination of duplex perception evoked by intensity differences,” *Percept. Psychophys.* **54**, 20–32.
- Bailey, P. J., Summerfield, Q., and Dorman, M. (1977). “On the identification of sine-wave analogues of certain speech sounds,” *Haskins Lab. Status Rep. Speech Res.* **SR-51/52**, 1–25.
- Berlin, C. I., Berlin, H. L., Hughes, L. F., and Dermody, P. (1976). “Dichotic vs monotic masking functions may reveal central organization for speech identification,” *J. Acoust. Soc. Am.* **59**, S5.
- Berlin, C. I., and McNeil, M. R. (1976). “Dichotic Listening,” in *Contemporary Issues in Experimental Phonetics*, edited by N. J. Lass (Academic Press, New York), pp. 327–387.
- Boersma, P., and Weenink, D. (2016). “PRAAT, a system for doing phonetics by computer (version 6.0.20) [software package],” Institute of Phonetic Sciences, University of Amsterdam, The Netherlands. <http://www.praat.org/> (Last viewed September 16, 2016).
- Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA).
- Brungart, D. S., Chang, P. S., Simpson, B. D., and Wang, D. L. (2006). “Isolating the energetic component of speech-on-speech masking with ideal time–frequency segregation,” *J. Acoust. Soc. Am.* **120**, 4007–4018.
- Cheng, C., and Xu, Y. (2013). “Articulatory limit and extreme segmental reduction in Taiwan Mandarin,” *J. Acoust. Soc. Am.* **134**, 4481–4495.
- Ciocca, V., and Bregman, A. S. (1989). “The effects of auditory streaming on duplex perception,” *Percept. Psychophys.* **46**, 39–48.
- Cutting, J. E. (1976). “Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening,” *Psychol. Rev.* **83**, 114–140.

- Darwin, C. J. (1981). "Perceptual grouping of speech components differing in fundamental frequency and onset-time," *Q. J. Exp. Psychol.* **33A**, 185–207.
- Darwin, C. J. (2008). "Listening to speech in the presence of other sounds," *Phil. Trans. R. Soc. B* **363**, 1011–1021.
- Darwin, C. J., and Hukin, R. W. (2004). "Limits to the role of a common fundamental frequency in the fusion of two sounds with different spatial cues," *J. Acoust. Soc. Am.* **116**, 502–506.
- Deutsch, D. (1979). "Binaural integration of melodic patterns," *Percept. Psychophys.* **25**, 399–405.
- Dorsi, J., Viswanathan, N., Rosenblum, L. D., and Dias, J. W. (2018). "The role of speech fidelity in the irrelevant sound effect: Insights from noise-coded speech backgrounds," *Q. J. Exp. Psychol.* **71**, 2152–2161.
- Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., and Kidd, G., Jr. (2003). "Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity," *J. Acoust. Soc. Am.* **114**, 368–379.
- Ellermeier, W., and Zimmer, K. (2014). "The psychoacoustics of the irrelevant sound effect," *Acoust. Sci. Tech.* **35**, 10–16.
- Fox, R. A., Jacewicz, E., and Feth, L. L. (2008). "Spectral integration of dynamic cues in the perception of syllable-initial stops," *Phonetica* **65**, 19–44.
- Gallun, F. J., Mason, C. R., and Kidd, G., Jr. (2007). "The ability to listen with independent ears," *J. Acoust. Soc. Am.* **122**, 2814–2825.
- Gardner, R. B., Gaskill, S. A., and Darwin, C. J. (1989). "Perceptual grouping of formants with static and dynamic differences in fundamental frequency," *J. Acoust. Soc. Am.* **85**, 1329–1337.
- Henke, W. L. (2005). "MITSYN: A coherent family of high-level languages for time signal processing [software package]," (W. L. Henke, Belmont, MA).
- Jones, D. M., and Macken, W. J. (1993). "Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory," *J. Exp. Psychol. Learn.* **19**, 369–381.
- Keppel, G., and Wickens, T. D. (2004). *Design and Analysis: A Researcher's Handbook*, 4th ed. (Pearson Prentice Hall, Englewood Cliffs, NJ).
- Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (2008). "Informational masking," in *Auditory Perception of Sound Sources*, Springer Handbook of Auditory Research, Vol. 29, edited by W. A. Yost and R. R. Fay (Springer, Boston, MA), pp. 143–189.
- Klatt, D. H. (1980). "Software for a cascade/parallel formant synthesizer," *J. Acoust. Soc. Am.* **67**, 971–995.
- Lawrence, M. A. (2016). "ez: Easy analysis and visualization of factorial experiments (R package version 4.4-0) [software]," <https://cran.r-project.org/package=ez> (Last viewed July 30, 2018).
- Mattys, S. L., Davis, M. H., Bradlow, A. R., and Scott, S. K. (2012). "Speech recognition in adverse conditions: A review," *Lang. Cogn. Process.* **27**, 953–978.
- Neff, D. L. (1995). "Signal properties that reduce masking by simultaneous, random-frequency maskers," *J. Acoust. Soc. Am.* **98**, 1909–1920.
- Porter, R. J., Cullen, J. K., Collins, M. J., and Jackson, D. F. (1991). "Discrimination of formant transition onset frequency: Psychoacoustic cues at short, moderate, and long durations," *J. Acoust. Soc. Am.* **90**, 1298–1308.
- Porter, R. J., Jr., and Whittaker, R. G. (1980). "Dichotic and monotic masking of CV's by CV second formants with different transition starting values," *J. Acoust. Soc. Am.* **67**, 1772–1780.
- Potter, R. K., Kopp, G. A., and Green, H. C. (1947). *Visible Speech* (Van Nostrand, New York).
- R Core Team. (2020). "R: A language and environment for statistical computing [software package]," The R Foundation, Vienna, Austria, <http://www.R-project.org/> (Last viewed July 31, 2020).
- Rand, T. C. (1974). "Dichotic release from masking for speech," *J. Acoust. Soc. Am.* **55**, 678–680.
- Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., and Lang, J. M. (1994). "On the perceptual organization of speech," *Psychol. Rev.* **101**, 129–156.
- Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (1981). "Speech perception without traditional speech cues," *Science* **212**, 947–950.
- Roberts, B., and Summers, R. J. (2015). "Informational masking of monaural target speech by a single contralateral formant," *J. Acoust. Soc. Am.* **137**, 2726–2736.
- Roberts, B., and Summers, R. J. (2018). "Informational masking of speech by time-varying competitors: Effects of frequency region and number of interfering formants," *J. Acoust. Soc. Am.* **143**, 891–900.
- Roberts, B., and Summers, R. J. (2019). "Dichotic integration of acoustic-phonetic information: Competition from extraneous formants increases the effect of second-formant attenuation on intelligibility," *J. Acoust. Soc. Am.* **145**, 1230–1240.
- Roberts, B., and Summers, R. J. (2020). "Informational masking of speech depends on masker spectro-temporal variation but not on its coherence," *J. Acoust. Soc. Am.* **148**, 2416–2428.
- Roberts, B., Summers, R. J., and Bailey, P. J. (2010). "The perceptual organization of sine-wave speech under competitive conditions," *J. Acoust. Soc. Am.* **128**, 804–817.
- Roberts, B., Summers, R. J., and Bailey, P. J. (2014). "Formant-frequency variation and informational masking of speech by extraneous formants: Evidence against dynamic and speech-specific acoustical constraints," *J. Exp. Psychol. Hum. Percept. Perform.* **40**, 1507–1525.
- Roberts, B., Summers, R. J., and Bailey, P. J. (2015). "Acoustic source characteristics, across-formant integration, and speech intelligibility under competitive conditions," *J. Exp. Psychol. Hum. Percept. Perform.* **41**, 680–691.
- Rosenberg, A. E. (1971). "Effect of glottal pulse shape on the quality of natural vowels," *J. Acoust. Soc. Am.* **49**, 583–590.
- Shinn-Cunningham, B. G. (2008). "Object-based auditory and visual attention," *Trends Cogn. Sci.* **12**, 182–186.
- Snedecor, G. W., and Cochran, W. G. (1967). *Statistical Methods*, 6th ed. (Iowa University Press, Ames, Iowa).
- Stevens, K. N. (1998). *Acoustic Phonetics* (MIT Press, Cambridge, MA).
- Studdert-Kennedy, M., and Shankweiler, D. (1970). "Hemispheric specialization for speech perception," *J. Acoust. Soc. Am.* **48**, 579–594.
- Summers, R. J., Bailey, P. J., and Roberts, B. (2010). "Effects of differences in fundamental frequency on across-formant grouping in speech perception," *J. Acoust. Soc. Am.* **128**, 3667–3677.
- Summers, R. J., Bailey, P. J., and Roberts, B. (2012). "Effects of the rate of formant-frequency variation on the grouping of formants in speech perception," *J. Assoc. Res. Otolaryngol.* **13**, 269–280.
- Summers, R. J., Bailey, P. J., and Roberts, B. (2016). "Across-formant integration and speech intelligibility: Effects of acoustic source properties in the presence and absence of a contralateral interferer," *J. Acoust. Soc. Am.* **140**, 1227–1238.
- Summers, R. J., Bailey, P. J., and Roberts, B. (2017). "Informational masking and the effects of differences in fundamental frequency and fundamental-frequency contour on phonetic integration in a formant ensemble," *Hear. Res.* **344**, 295–303.
- Summers, R. J., and Roberts, B. (2020). "Informational masking of speech by acoustically similar intelligible and unintelligible interferers," *J. Acoust. Soc. Am.* **147**, 1113–1125.
- Tremblay, S., and Jones, D. M. (1999). "Change of intensity fails to produce an irrelevant sound effect: Implications for the representation of unattended sound," *J. Exp. Psychol. Hum. Percept. Perform.* **25**, 1005–1015.
- Viswanathan, N., Dorsi, J., and George, S. (2014). "The role of speech-specific properties of the background in the irrelevant sound effect," *Q. J. Exp. Psychol.* **67**, 581–589.
- Whalen, D. H., and Liberman, A. M. (1987). "Speech perception takes precedence over nonspeech perception," *Science* **237**, 169–171.