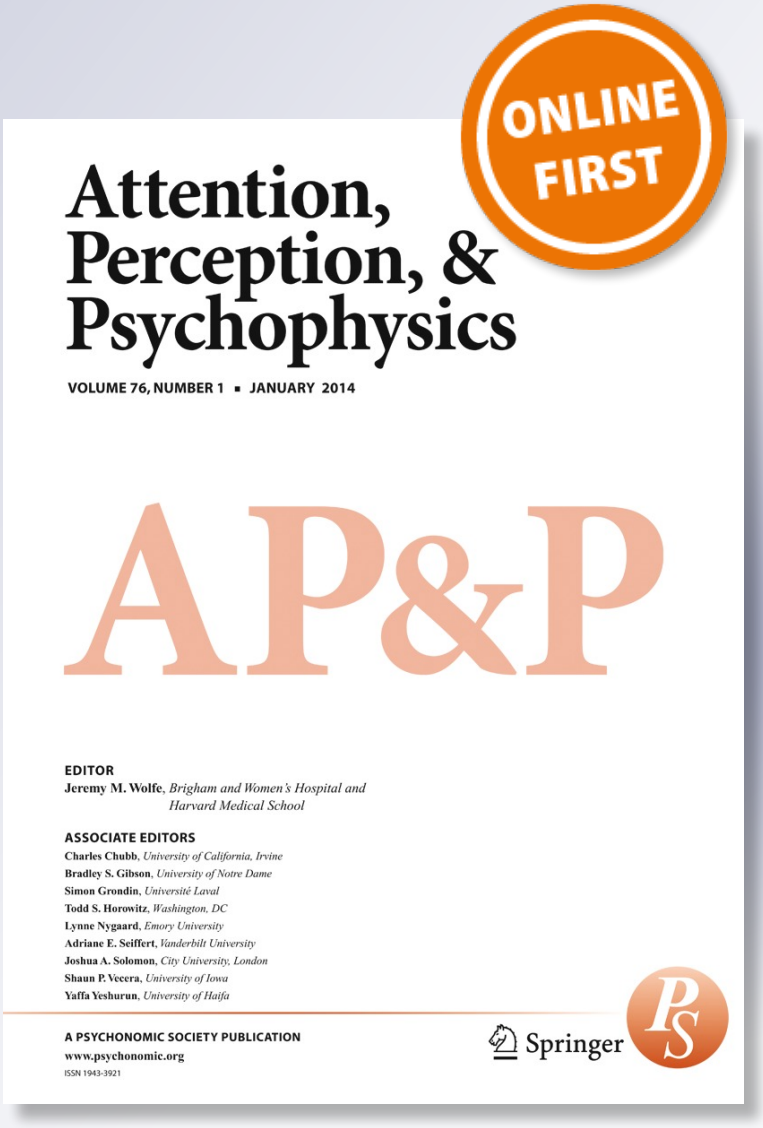
Context effects as auditory contrast

John Kingston, Shigeto Kawahara, Della Chambless, Michael Key, Daniel Mash & Sarah Watsky

Attention, Perception, & Psychophysics

ISSN 1943-3921

Atten Percept Psychophys
DOI 10.3758/s13414-013-0593-z



Springer

Atten Percept Psychophys
DOI 10.3758/s13414-013-0593-z

Context effects as auditory contrast

John Kingston · Shigeto Kawahara · Della Chambless · Michael Key · Daniel Mash · Sarah Watsky

Abstract Three experiments are reported that collectively show that listeners perceive speech sounds as contrasting auditorily with neighboring sounds. Experiment 1 replicates the well-established finding that listeners categorize more of a [d–g] continuum as [g] after [l] than after [r]. Experiments 2 and 3 show that listeners discriminate stimuli in which the energy concentrations differ in frequency between the spectra of neighboring sounds better than those in which they do not differ. In Experiment 2, [alga–arda] pairs, in which the energy concentrations in the liquid-stop sequences are H(igh) L(ow)–LH, were more discriminable than [alda–arga] pairs, in which they are HH–LL. In Experiment 3, [da] and [ga] syllables were more easily discriminated when they were preceded by lower and higher pure tones, respectively—that is, tones that differed from the stops' higher and lower *F*3 onset frequencies—than when they were preceded by H and L pure tones with similar frequencies. These discrimination results show that contrast with the target's context exaggerates its perceived value when energy concentrations differ in frequency between the target's spectrum and its context's spectrum. Because contrast with its context does more that merely shift the criterion for categorizing the target, it cannot be produced by neural adaptation. The finding that nonspeech contexts exaggerate the perceived values of speech targets also rules out compensation for coarticulation by showing that their values depend on the proximal auditory qualities evoked by the stimuli's acoustic properties, rather than the distal articulatory gestures.

Keywords Context effects · Auditory contrast · Compensation for coarticulation · Speech · Nonspeech

The percept of a speech sound is affected by its context (Repp 1982, and many others). Among the well-studied context effects is Mann's (1980) finding that listeners respond "ga" more often to the [da–ga] continuum after [al] than after [ar]. This effect has provided considerable empirical evidence in the debates between "auditorists," who argue that the objects of speech perception are auditory qualities (Diehl and Kluender 1989; Diehl et al. 2004; Lotto and Holt 2006; Lotto and Kluender 1998), and "gesturalists," who argue that they are, instead, articulatory gestures (Fowler 1986, 2006; Liberman et al. 1967; Liberman and Mattingly 1985, 1989).[1] Auditorists explain this and similar context effects as a product of auditory contrast between the target sound, the syllable from the [da–ga] continuum, and its context, the preceding [al] or [ar]. A stop whose *F*3 onset frequency is intermediate between [d]'s high

J. Kingston (✉)
Linguistics Department, University of Massachusetts, 150 Hicks Way, 226 South College, Amherst, MA 01003-9274, USA
e-mail: jkingston@linguist.umass.edu

S. Kawahara
Institute of Cultural and Linguistic Studies, Keio University, Tokyo, Japan

D. Chambless
Italian Program, Duke University, Durham, NC, USA

M. Key
Linguistics Department and Center for the Advanced Study of Language, University of Maryland, College Park, MD, USA

D. Mash · S. Watsky
Linguistics Department, University of Massachusetts, Amherst, MA, USA

[1] The gesturalist account encompasses two distinct accounts of these perceptual objects, the motor theory (Liberman and Mattingly 1985, 1989; Liberman et al. 1967) and direct realism (Fowler 1986, 2006), that differ in their theoretical rationales for choosing gestures as the objects of speech perception. For our purposes, all that matters is that both theories propose that the objects of speech perception are articulatory gestures and that many context effects, including the effect of [al] versus [ar] on percepts of a following [da–ga] continuum examined here, are produced by listeners' compensating for coarticulation. Since the direct-realist account treats the perception of speech sounds like the perception of other events in the world, it is a more general alternative than the motor-theory account to the auditorist account and, therefore, is the focus of our discussion of the gesturalist alternative.

Published online: 25 March 2014

value and [g]'s low one is heard more often as having a lower, more [g]-like *F*3 after [l]'s high *F*3 than after [r]'s low *F*3. According to gesturalists, listeners instead compensate for coarticulation with the preceding liquid. The stop with an ambiguous place (= intermediate *F*3 onset frequency) is perceived as having been pulled forward by [l]'s more anterior articulation, and the listener compensates for this fronting and hears it more often as the posterior alternative [g]. We report three experiments that jointly support the auditorist account over the gesturalist account—that is, contrast over compensation. The results also show that auditory contrast between target and context exaggerates the perceived value of the affected acoustic property in the target, rather than shifting the criterion for deciding what category the target sound belongs to. In the next section, we develop a detection-theoretic model of exaggeration and criterion shifts to show how we test these claims empirically.

Modeling context effects on response bias and sensitivity

The panels in Fig. 1 display response-likelihood distributions to adjacent stimuli along a [d–g] continuum with respect to the corresponding perceptual dimension (Macmillan and Creelman 2005). The sensitivity measure d' is the distance between the means of these two distributions, in units of their standard deviation. The criterion used to decide whether the stop is "d" or "g" is represented by the vertical line labeled "?". Its value, *c*, represents the listener's response bias. Percepts that fall to the left of it lead to "d" responses, those to the right to "g" responses. The criterion is placed midway between the distributions in Fig. 1(i) to show that responding is unbiased when no context precedes (c=0). The value of d' is 2; it equals the perceptual distance between the means of the response-likelihood distributions.

Figure 1(ii) shows how a preceding [l] context might shift the criterion, in the figure from 0 to −0.5, for deciding whether the following stop is [d] or [g] without changing the listener's sensitivity to the difference between those two stops; compare the horizontal position of the dotted decision criterion in Fig. 1(ii) with the solid one in Fig. 1(i). This criterion shift decreases the proportions of the areas below the criterion under both the "d" and "g" distributions, from 0.841 and 0.159 to 0.691 and 0.067. Responses are biased toward "g": *c* changes from 0 to 0.5, but sensitivity to the difference between the two stimuli does not change, because the means of the response-likelihood distributions do not shift and d' remains 2.[2]

Figure 1(iii) shows how the preceding [l] could, instead, change sensitivity to the difference between adjacent steps along the [d–g] continuum rather than shifting the criterion, which remains fixed at 0; compare Figure 1(i). By shifting the entire "g" response-likelihood distribution away from the "d" distribution, by as much (0.5) as the criterion shift in Figure 1(ii), the proportion of the area under the "g" response-likelihood distribution below the decision criterion again decreases from 0.159 to 0.067, but that under the "d" response-likelihood distribution remains unchanged at 0.841, and the d' value jumps from 2 to 2.5. The bias toward "g" responses again increases, although only to c=0.25.

This modeling exercise shows that both a criterion shift and exaggeration alter response biases (*c* values) in categorization tasks but that only exaggeration also changes sensitivity to differences between adjacent stimuli in discrimination tasks. This discussion of context effects' psychophysics does not, however, reveal the psychological processes responsible for these effects.

Psychophysics and psychological processes

How might the target "contrast" with its context? In discussing Lotto and Kluender's (1998) account, Fowler (2006) ascribed the context's contrastive effects to reduced perceptual sensitivity[3] to acoustic values in the target that have been shifted by coarticulation with the context—that is, to the "assimilative effects of /l/ and /r/ during /d/ and /g/" (p. 163). The high versus low energy concentrations in [l]'s versus [r]'s spectrum reduce listeners' sensitivity to the raised or lowered energy distributions in the stop's spectrum caused by coarticulation with a preceding [l] or [r], which in turn lowers or raises the perceived energy concentration in an intermediate stop's spectrum and causes it to be preceived more like [g] or [d], respectively. If contrast between context and target works by changing sensitivity in this way, it should shift the listener's decision criterion, as depicted in Fig. 1(ii).

Alternatively, contrast could exaggerate the target's perceived value for whichever acoustic properties differ from those in its context and increase the listener's sensitivity to differences between adjacent stimuli along the continuum, as depicted in Fig. 1(iii). Stephens and Holt (2003) showed that listeners were significantly better at discriminating both adjacent syllables from a [da-ga] continuum and adjacent nonspeech stimuli consisting of the first 80 ms of the syllables' *F*2 and *F*3 transitions when the more [da]-like syllable or transitions were preceded by [ar] and the more [ga]-like ones by [al] than vice versa. Contrast with the preceding context increased sensitivity for both sorts of targets even though listeners could categorize the syllables far more consistently than the transitions, and they discriminated the syllables, but not the transitions,

[2] In discussing Experiment 2 below, we show how two criterion shifts, in opposite directions, could in principle change sensitivity in a discrimination task.

[3] Fowler (2006) used "sensitivity" in a quite different sense than we do. Her sense refers to how well listeners perceive that a sound has a particular value for an acoustic property, and not to how well they perceive the *difference* between the values of neighboring sounds along the continuum. We use it in this alternative sense in the next paragraph, where we discuss contrast as exaggeration.

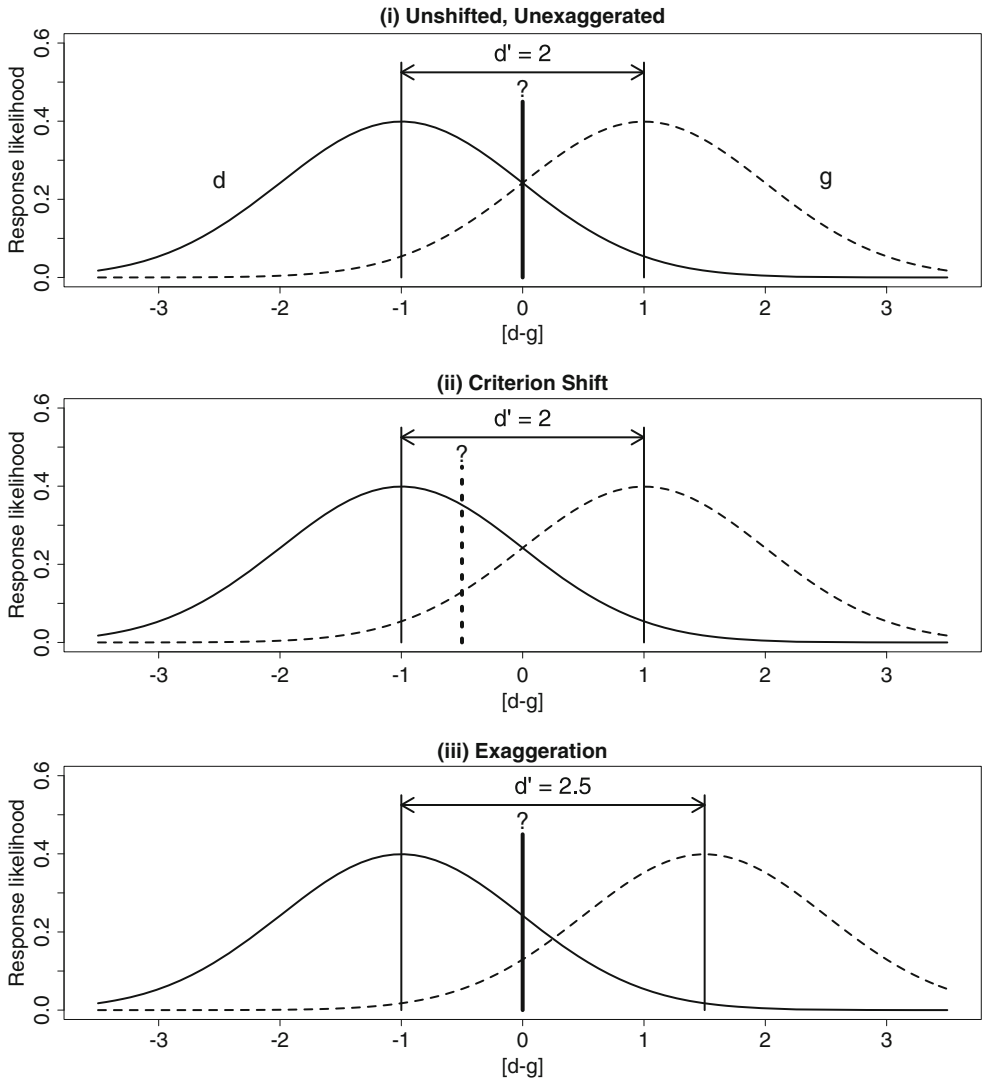


Fig. 1 "d" (solid line) and "g" (dashed line) *response-likelihood distributions*, their means (thinner, unlabeled solid lines), and *decision criteria* (thicker vertical lines topped with "?") for when there is no preceding context and responding is unbiased (i), as compared with when [l] precedes and shifts the decision criterion (dotted line) 0.5 units to the left (ii) or exaggerates the perceived lowness of the [g]-like stimulus and shifts the "g" response-likelihood distribution 0.5 units to the right (iii)

better near the category boundary than near the continuum endpoints. These findings show that contrast between a target and its context can exaggerate the target's perceived value independently of any shift in the criterion for categorizing it.

Both a criterion shift and exaggeration are compatible with the auditorist account of context effects, because contrast between the context's and target's spectra either can shift the criterion for deciding whether energy is concentrated low enough in the stop's spectrum for it to qualify as an instance of the category [g] or can exaggerate the perceived lowness of that energy concentration. However, only a criterion shift is compatible with direct realism's compensation account, because an exaggerated percept does not represent the signal's acoustics veridically and such a distortion of the information provided by the acoustics would prevent the listener from reliably recovering the articulations that produced them. An exaggerated percept is "heteromorphic" to the extent that its values differ from those of the proximal stimulus and, thus, deviates from the "homomorphy" between the proximal stimulus's values and the articulatory gestures that produced it (Fowler 1990, pp. 1246–1247). Shifting a criterion to compensate for coarticulation would not, however, introduce heteromorphy—quite the contrary, because it is produced by an accurate parsing of the target sound's and its context's articulatory influences on the signal's acoustics (Fowler 1994, 1996, 2005, 2006; Fowler and Brown 1997; Fowler and Smith 1986; Pardo and Fowler 1997). Therefore, finding evidence that the target sound's context not only shifts the criterion for categorizing it, but also exaggerates its perceived value would rule out compensation for coarticulation as it is conceived in the direct-realist account as the mechanism responsible for the criterion shift itself. Stephens and Holt

(2003) present such evidence for nonspeech targets; Experiments 2 and 3 present similar evidence of exaggeration for speech and nonspeech contexts, respectively.

Is contrast the only means by which the target's perceived value could be exaggerated? Couldn't the context's effect on the perception of the target—that is, the extent to which the listener compensates—be a direct function of how much the listener perceives the target as coarticulated with that context, and couldn't greater compensation for coarticulation with more extreme contexts increase the perceived difference between the targets? If a coarticulated segment overlaps more with an articulatorily different context in one token than in another, its acoustic properties should be altered more in that token, and therefore, the listener would have to compensate more for those effects in recognizing the affected sound. They could succeed in doing so by shifting the criterion farther for deciding whether that sound belongs to the category corresponding to the direction in which the context has shifted its acoustic properties. For example, if [l] overlaps more with a following [g] in a particular token and, as a result, raises that [g]'s *F*3 more, the listener could compensate for that greater raising by also raising the criterion of what counts as a [g]-like *F*3 to higher values than would be necessary for a [g] token that is overlapped less by a preceding [l]. Similarly, a token of [d] that is overlapped more by a preceding [r] would evoke more extensive compensation for the lowering of its *F*3, which could also be accomplished by shifting the criterion for what counts as a [d]—in this case, toward lower *F*3 values. In both cases, the extent of compensation varies directly with the extent of perceived coarticulation, and in both cases, compensation is achieved by shifting the decision criterion. If compensation is achieved entirely by shifting the decision criteria in a direction and to an extent that corresponds to the extent of coarticulation, the likelihood of categorizing the stops as "g" or "d" increases, but sensitivity to the difference between the two stops does not. In short, completely successful parsing would at most return [g] and [d] perceptually to their uncoarticulated state, not exaggerate how [g]- or [d]-like they are.

The products of compensating for coarticulation are described here as the idealized gestures that would be produced if the target sound's pronunciation were not altered by coarticulation with its context. In its description of these products, the direct-realist account does not, therefore, differ from the motor-theory's account, even though they differ profoundly in the means by which these products are obtained. The acoustics of the proximal stimulus inform the listener about how gestures overlap or coarticulate and how they are blended by the task dynamics in the distal stimulus. Listeners cannot stop with recognizing that *some* gestures have been overlapped, coarticulated, and blended but must go on to recognize *which* gestures. Doing so, in compensating for coarticulation, thus undoes its effects just as much in the direct-realist as in the motor-theory account.

As the comparison of Fig. 1(ii) with Fig. 1(iii) above showed, it is impossible to tell whether the context has shifted the criterion or altered sensitivity from a finding that the context changed the listener's response bias in a categorization task. Only a discrimination task like that in Experiment 2 or like that carried out by Stephens and Holt (2003) can determine whether the context shifts the decision criterion or, instead, the mean of the response-likelihood distribution. We describe below how criterion shifts in opposite directions could produce the illusion of an increase in sensitivity to the difference between adjacent stops along the [d–g] continuum in a discrimination task, together with the other conditions that must also be met before an apparent increase in sensitivity can plausibly be attributed to such criterion shifts.[4]

Our first experiment replicates Mann's (1980) finding that listeners respond "ga" more often after [al] than after [ar]. The second tests whether a preceding liquid context exaggerates the perceived value of the stop's spectrum by comparing the discriminability of [alga] versus [arda] pairs with that of [alda] versus [arga] pairs. The energy distributions in the spectra of the contexts and targets in these stimulus pairs are high (H)–low (L) versus LH for [alga] versus [arda] and HH versus LL for [alda] versus [arga]. The liquid context could exaggerate the perceived value of the energy distribution in the stop target's spectrum in the HL versus LH pairs because the context's and target's spectra differ in their energy distributions, but not in HH versus LL pairs because their spectra do not differ. Exaggeration therefore predicts that HL versus LH pairs should be more discriminable than HH versus LL pairs, even though the acoustic differences between each interval in both the contexts and targets are equal in the two kinds of pairs. Experiment 3 tests whether any exaggeration induced by context arises in the auditory response to the stimuli by substituting nonspeech analogues as contexts (cf. Fowler et al. 2000; Lotto and Kluender 1998). This last experiment tests the hypothesis that it is specifically the proximal *auditory* qualities evoked by the acoustic properties of the context and target that contrast. If nonspeech contexts exaggerate the perceived values of speech targets in the same way that speech contexts do, speech-on-speech exaggeration can plausibly and parsimoniously be attributed to auditory contrast, too. In summary, the two accounts compared here predict the outcomes listed in Table 1.

Experiment 1: Replicating Mann (1980)

Mann (1980) presented listeners with syllables from a [da–ga] continuum following [al] versus [ar] syllables. The stop continuum was produced by incrementally

[4] We postpone until the General Discussion section responding to challenges to the contrast account presented by Fowler et al. (2000) and Viswanathan et al. (2009, 2010).

Table 1 Predictions of the contrast (auditorist) versus compensation (gesturalist) accounts of context effects, along with experiments that test them

	Criterion Shift	Exaggeration
	Figure 1(ii)	Figure 1(iii)
	p("ga" | al_) > p("ga" | ar_)	*d′* alga vs. arda > *d′* alda vs. arga: Exp. 2
	Exp. 1	*d′* H–ga vs. L–da > *d′* H–da vs. L–ga: Exp. 3
Contrast	yes	yes
Compensation	yes	no

varying *F*3's onset frequency from a high frequency of 2690 Hz to a low frequency of 2104 Hz. Multiple tokens of naturally produced [al] and [ar] syllables were used as contexts. These syllables were originally produced before [da] and [ga] syllables. Before [da], the average *F*3 offset frequencies were 2773 and 1680 Hz for [al] and [ar], respectively, while before [ga], they were slightly closer together, 2649 and 1786 Hz.[5] Following [al], listeners responded "ga" to more of the continuum than following [ar]; however, they did not respond "ga" to any less of the continuum after [ar] than when no context preceded.

The [da–ga] continuum in our stimuli was also synthesized by varying *F*3's onset frequency alone, but we synthesized the preceding liquid contexts, too. These liquid contexts resembled those in Mann's (1980) stimuli in being modeled on the acoustics of naturally produced [l] and [r], which differ in more than their *F*3 frequencies (see Stevens 1998, and the description of the stimuli's acoustics below). Unlike Mann, we also presented listeners with a continuum between [al] and [ar], rather than just categorically different endpoint contexts. Our listeners identified the liquid on each trial as well as the stop. We can, therefore, separate the effects of the liquid's acoustics on the stop percept from those of the category to which the liquid is assigned. We can also determine whether the stop's acoustics and/or category assignment affects the percept of the liquid; that is, does the stop act as a context for categorizing the liquid, and if so, how?[6] This experiment determines whether our stimuli and procedures can produce the same contextual effects as those that Mann reported and, thus, lays the foundation for determining whether those effects are due to auditory contrast or compensation for coarticulation.

[5] Mann (1980) reported that listeners were more likely to respond "ga" after either liquid when it was originally pronounced before [ga]. This effect was small for [al] contexts but large for [ar] contexts.

[6] Mann's (1980) listeners also categorized the liquid as well as the stop, but she does not report those responses or how they influenced listeners' categorization of the following stop.

Method

Stimuli

The stimuli consisted of two syllables, the first drawn from a seven-step [al-ar] continuum and the second from a seven-step [da-ga] continuum (see Fig. 2 for the corners of the 7×7 array). Both syllables were synthesized using the Sensimetrics implementation of the KLSYN88 terminal analogue synthesizer (Klatt and Klatt 1990).

The vowel portion of the first syllable lasted 160 ms; it was followed by a 60-ms transition to the liquid portion, which lasted the remaining 80 ms of the syllable (for a total duration of 300 ms). The choices of parameter values for the [l] and [r] were guided by Stevens's (1998) description of the acoustics of naturally produced liquids. In the [ar] endpoint, *F*2 was 1300 Hz for the syllable's entire duration, while it fell to 1100 Hz during the liquid portion of the [al] endpoint. In both endpoints, *F*3 was at 2500 Hz during the vowel portion, then rose to 2800 Hz in the [al] endpoint or fell to 2000 Hz in the [ar] endpoint. During the liquid portion of the [al] endpoint, a lower pole and zero were set at 1500 Hz, and a higher pole and zero were set at 3900 and 3300 Hz, while in the [ar] endpoint, the lower pole and zero were set at 1500 and 2000 Hz, and the higher pole and zero were both set at 3900 Hz. (When pole and zero have the same frequencies, the zero cancels the pole.) These settings introduce an additional high-frequency spectral prominence above *F*3 in the [al] endpoint and an additional low-frequency spectral prominence above *F*2 in the [ar] endpoint; the zero at 2000 Hz also cancels *F*3 during the [r] interval. The result is a strong high-frequency concentration of energy in the endpoint [l]'s spectrum and a strong low-frequency concentration in the endpoint [r]'s spectrum (compare Figs. 3a and 3c).

The second syllable followed a 95-ms gap containing only low-frequency periodic energy that simulated voicing during a stop closure. The syllable began with a 60-ms transition, followed by a 240-ms steady state (for a total duration of 300 ms). In both the [da] and [ga] endpoints, *F*2 began at 1988 Hz, and its steady state value was 1300 Hz; in the [da] endpoint, *F*3 began at 2690 Hz, while in the [ga] endpoint, it began at 2104 Hz, and its steady state value for all stimuli was 2400 Hz.[7] This manipulation diffused energy by separating the peaks corresponding to *F*2 and *F*3 at the beginning of the syllable at the [da] endpoint (Fig. 3a, c) or concentrated it

[7] Figure 2 also shows that *F*3 rose briefly at the end of the vowel. This was an inadvertent carryover from an earlier version of the stimuli in which a [t] followed the vowel. The change was barely audible and present in all the stimuli.

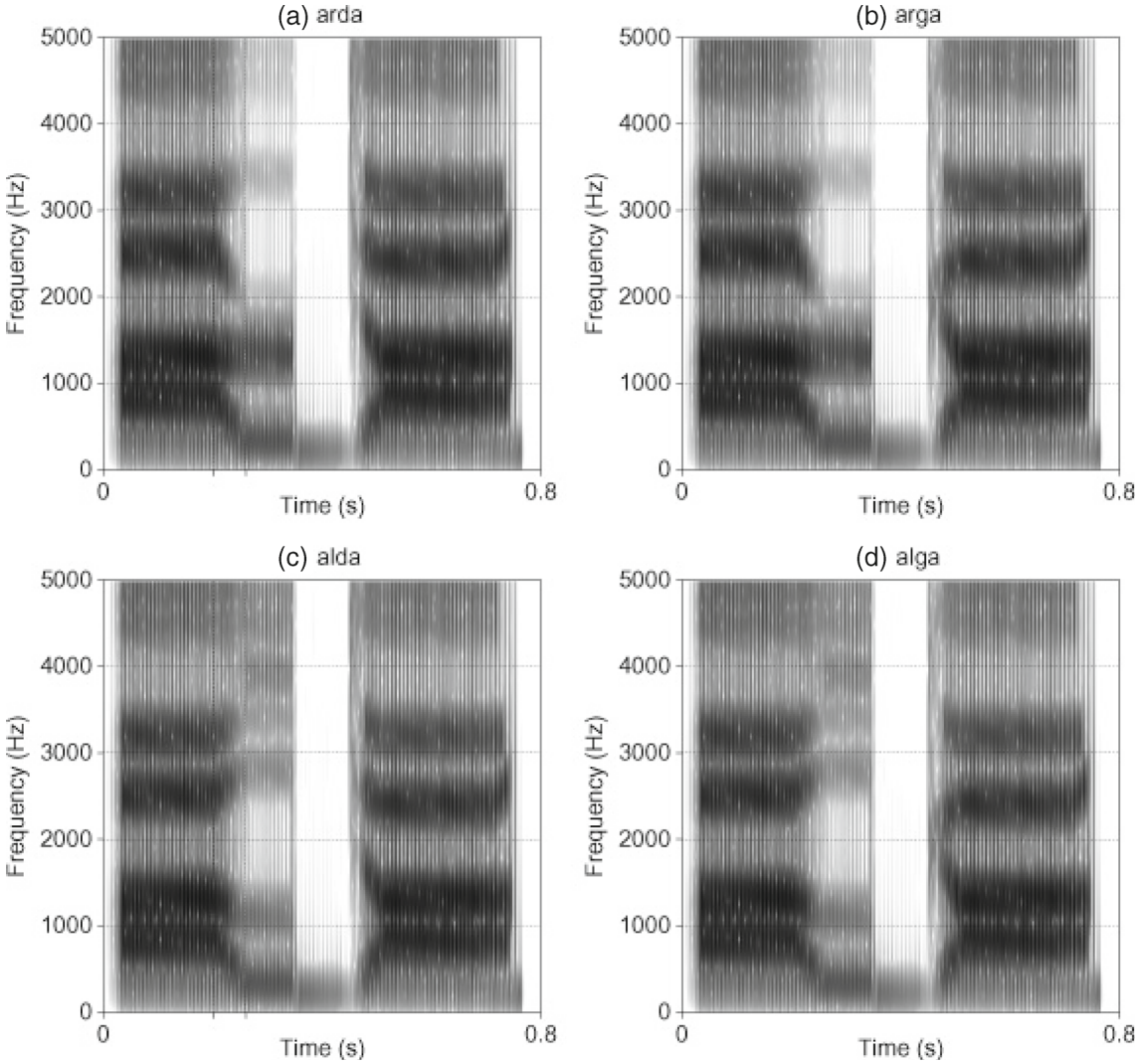


Fig. 2 Spectrograms of the corner (endpoint) stimuli from the 7× 7 array: [arda] (**a**), [arga] (**b**), [alda] (**c**), and [alga] (**d**). The vertical lines and ticks on the bottom edges in panels a and c mark the beginning and end of the transitions from the vowel steady-state to the liquid steady-state in the first syllable

by putting them next to one another at the [ga] endpoint (Fig. 3b, d).

Parameter values for intermediate stimuli along the [l–r] and [d–g] dimensions of the stimulus space were obtained by linear interpolation between the endpoint values. All other parameter values were the same in all stimuli. They are listed in the Appendix.

Consent and instructions

In this and the two other experiments, listeners gave informed consent before participating. Instructions were given first in writing and then repeated verbally by the experimenter, who also answered any questions. The instructions described what would happen on each trial, how the trials were organized into training and test blocks, and when breaks would occur. At the end of the experiment, listeners were debriefed before the purpose of the experiment was explained to them. These procedures were followed in all the experiments reported.

Participants

All participants were adult native speakers of English, who reported that they had not been exposed to any language other than English before the age of 6 years and who reported no hearing or speaking disorders. They were recruited from the University of Massachusetts, Amherst community and were either paid for their time or granted course credit. Participation in all the experiments came from the same population. Twenty-three listeners participated in this experiment.

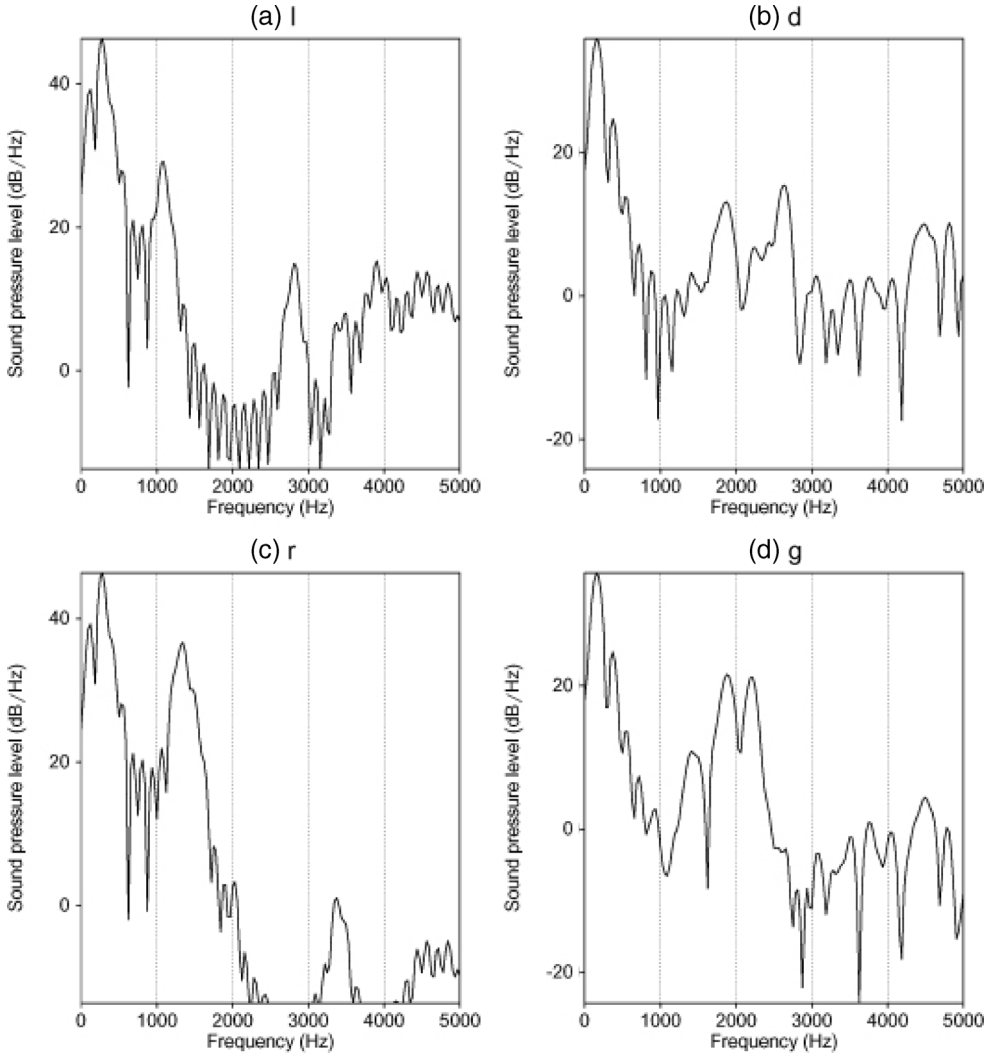


Fig. 3 Spectra from 25-ms-wide Gaussian windows in the [l] (**a**) and [r] (**c**) steady-states and immediately following the release of the [d] (**b**) and [g] (**d**)

Procedure

Listeners were trained with feedback to identify the four corners of the 7×7 stimulus array as "LD," "RD," "LG," and "RD" (Fig. 4). Each corner stimulus was presented 4 times in random order for a total of 16 training trials. In each of the ensuing test blocks, the 49 stimuli in the array were presented once in random order. Whenever a corner stimulus was presented in a test block, listeners received feedback, and at the beginning of every subsequent pair of test blocks (every 98 trials), they were retrained with feedback with four randomized repetitions of the four corner stimuli. A total of 24 test blocks were presented to each listener. Listeners took self-timed short breaks between every 2 blocks of trials and longer ones between every 6 blocks. They were assigned at random to one of four conditions, which differed in how the four responses were assigned to buttons on the response box.

All aspects of stimulus presentation and response collection in the experiment were controlled by SuperLab version 2.04. Each trial began with the presentation of the stimulus. Once the stimulus ended, the four alternative responses were then immediately displayed on the screen in the same arrangement and colors as the corresponding buttons on the response box. The listener had 1,500 ms in which to respond. The response prompts disappeared once this time had elapsed or the listener responded, whichever came earlier. If feedback was presented on that trial, it appeared in the form of the correct answer (in the appropriate color) in the middle of the screen for 750 ms. All trials ended with the screen going blank for 750 ms before the next stimulus was presented.

Listeners sat in front of an LCD display in a sound-treated room, where they listened to the stimuli on Sennheiser HD 280 (64Ω) headphones and used a Cedrus RB-834 response box to respond. The four outer buttons on the response box were used to collect responses. Listeners rested the index and middle fingers or the thumbs and index fingers of their two hands on these buttons, so that they did not need to move their

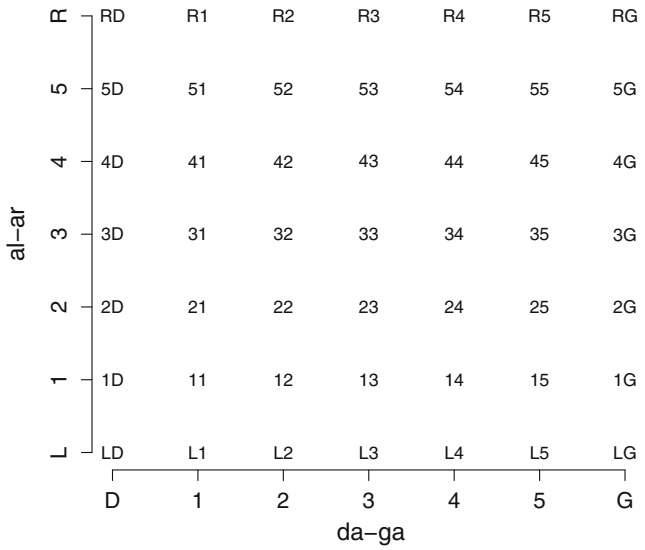


Fig. 4 The 7×7 [al–ar] by [da–ga] stimulus array. The high–high corner of the array is the lower left

hands to respond. They were told that they could not respond until the stimulus had finished playing and response prompts appeared on the display but that they should respond as soon as the stimulus ended and the prompts appeared.

Analysis

Due to experimenter error, on trials where stimulus 31 (see Fig. 4) was to be presented, stimulus 41 was presented instead; all responses on these trials were omitted from the analysis.

The first analysis assessed the influences of the liquids' or stops' acoustics on the number of "g" relative to "d" responses or the number of "l" relative to "r" responses. The second analysis assessed how categorizing the liquid as "l" versus "r" influenced the likelihood of the listener categorizing the following stop as "g" versus "d" and vice versa. The first analysis thus ignores the influence of the contexts' categorization, while the second ignores the influence of the contexts' acoustics. The first analysis thus shows how the psychoacoustic correlates of the context influence categorization of the target, while ignoring their psychophysical influence—namely, any bias to categorize the context as "l" versus "r"—while the second analysis shows how such a psychophysical bias affects categorization of the target independently of the context's psychoacoustic values.

In both analyses, the relative proportions of "g" versus "d" or "l" versus "r" responses served as binomially distributed dependent variables in mixed-effects logistic-regression models in which the fixed effects were the targets' acoustics and the contexts' acoustics or categorization and the random effects were differences between listeners in the intercept and the slopes of the fixed effects (Baayen 2011; Bates et al. 2011; R Development Core Team 2011).[8] Including random effects of subjects on the intercepts captures differences between listeners in overall response biases, while including random effects of subjects on the slopes of the fixed effects captures differences in their sensitivity to the manipulations represented by the fixed effects. Barr et al. (2013) argued that not using such maximal random-effect structures inflates type 1 error rates. In these and subsequent models, the fixed effects were centered to reduce correlations between their slopes and intercepts(as recommended by Baayen 2008, pp. 254–255): (1) Steps along the stop acoustics continuum [d, 1, 2, 3, 4, 5, g] became [−3, −2, −1, 0, 1, 2, 3], (2) those along the liquid acoustics continuum [l, 1, 2, 3, 4, 5, r] became [−3, −2, −1, 0, 1, 2, 3], (3) categorization of the stops as "d" versus "g" became −1 versus 1, and (4) categorization of the liquids as "l" versus "r" became 1 versus −1.[9]

Results

Context acoustics

Figure 5a shows that "g" responses increase from the [da] to the [ga] end of the stop continuum and that they increase from the [ar] to [al] end of the liquid continuum.

A two-step hierarchy of mixed-effects logistic-regression models of the stop responses was constructed, the first in which the fixed effects of the stops' and liquids' acoustics were independent and the second in which the interaction between them was included (see Jaeger 2008, for the advantages of such models over ANOVAs with logit-transformed response proportions). A comparison of the models' log likelihood ratios showed that including the interaction only marginally improved the fit to the data, $\chi^2(5) = 9.4139$, $p = .094$, and the estimate for the interaction was not significant ($p > .10$), so the simpler model without the interaction is interpreted here. The estimates in Table 2 show that the proportion of "g" responses relative to "d" responses increased significantly as the stop became [g]-like and decreased significantly as the liquid became more [r]-like.

Since this is a logistic regression, the estimates serve as arguments of the exponential function:

$$exp(0.95106 + 1.50397*\text{d–g} + -0.08334*\text{l–r}), \quad (1)$$

where "d–g" and "l–r" represent the centered values, −3, −2, −1, 0, 1, 2, 3, corresponding to the seven steps along the [d–g] and [l–r] continua. The value of this expression represents the predicted odds of a "g" response relative to "a" "d" response.

[8] Prior analyses of response assignments to buttons showed no significant effects of this variable, so it was omitted from further analysis.

[9] Neither the stop nor the liquid categorization was centered when they served as dependent variables.

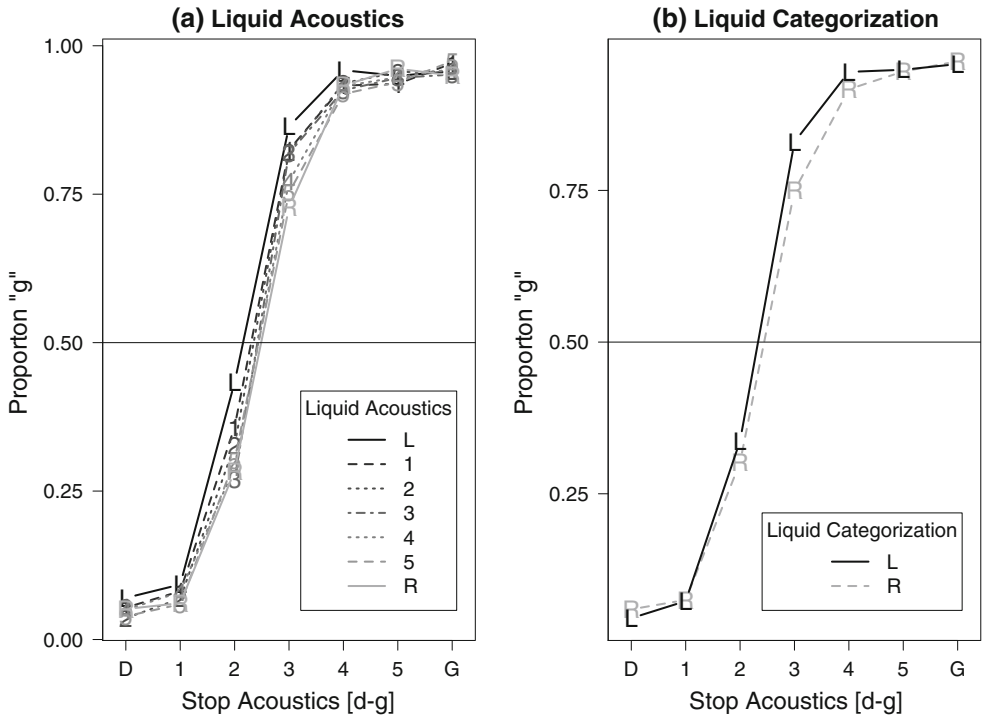


Fig. 5 Proportions of "g" responses as a function of the stops' acoustics and the liquids' acoustics (**a**) or the liquids' categorization (**b**)

For example, at the first steps along the two continua, where "d–g" and "l–r" both equal −3, the predicted odds ratio equals:

$$exp(0.95106 + 1.50397*-3 + -0.08334*-3) = 0.03649 \quad (2)$$

—that is, very low odds of a "g" response. An increase by one step along the [d–g] continuum to −2 increases that ratio to 0.16417 and an increase by one step to −2 along the [l–r] continuum decreases it to 0.03357. Holding the centered value of step along the [l–r] continuum constant at its midpoint, 0, the predicted odds ratios shift from 0.02879 to 235.80 across the range of [d–g] values; similarly, holding the centered value of step along the [d–g] continuum constant at its midpoint, the ratios shift from 3.3237 to 2.0158 across the range of [l–r] values.

Odds ratios can, in their turn, be transformed into the more familiar (predicted) probabilities, as in

$$\frac{exp(0.95106 + 1.50397*\text{d–g} + -0.08334*\text{l–r})}{1 + exp(0.95106 + 1.50397*\text{d–g} + -0.08334*\text{l–r})}. \quad (3)$$

Using this formula, the odds ratios calculated above, 0.03649, 0.16417, 0.03357, 0.02879, 235.8, 3.3237, and 2.0158, correspond to probabilities of 0.03521, 0.14102, 0.03248, 0.02798, 0.99578, 0.76872, and 0.668413, respectively

Figure 6a shows that "l" responses became less frequent as the liquid became more [r]-like and that they were affected little, if at all, by the acoustics of the stop.

A two-step hierarchy of mixed-effects models of the liquid responses like that used in the analysis of the stop responses was constructed. Once again, including the interaction did not significantly improve the fit, $\chi^2(5) = 5.4842$, $p > .10$. The estimates obtained from the simpler model in Table 3 show that only the liquids' acoustics significantly influenced the relative proportion of "l" versus "r" responses. The nonsignificant estimate for the

Table 2 Independent acoustic effects on "g" versus "d" response proportions

	Estimate	*SE*	*z*	*p*
Intercept	0.95106	0.11899	7.993	1.32e-15
Stop acoustics=[d–g]	1.50397	0.07671	19.606	< 2e-16
Liquid acoustics=[l–r]	−0.08334	0.01984	−4.201	2.65e-05

Table 3 Interaction of acoustic effects on "g" versus "d" response proportions

	Estimate	*SE*	*z*	*p*
Intercept	0.800659	0.121644	6.582	4.64e-11
Stop acoustics=[d–g]	−0.006576	0.010485	−0.627	0.531
Liquid acoustics=[l–r]	−0.928745	0.092496	−10.041	< 2e-16

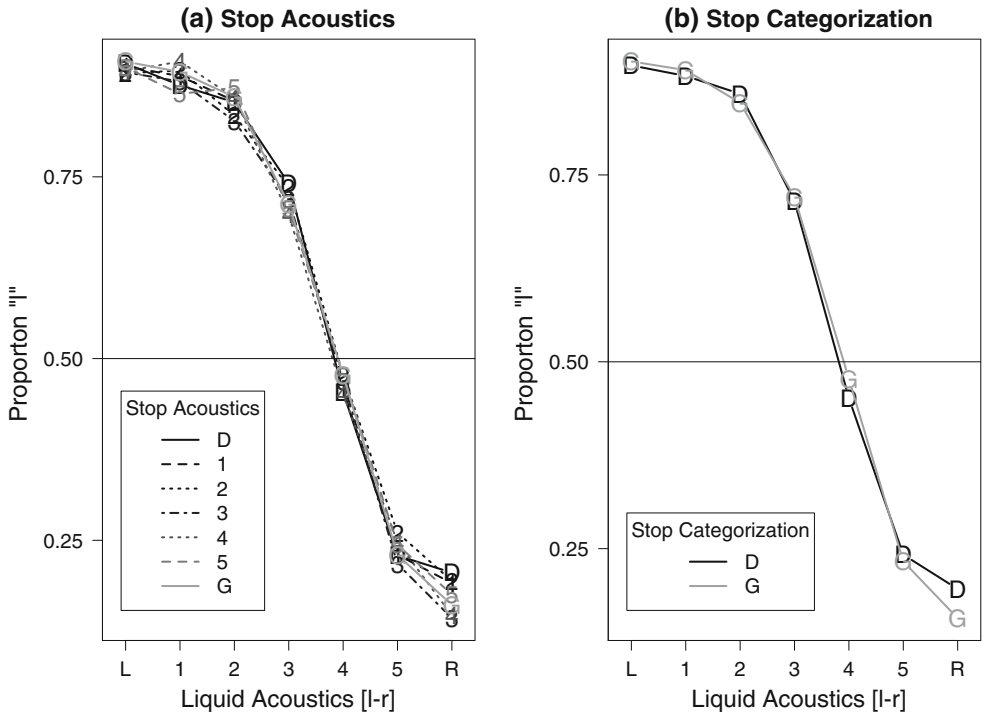


Fig. 6 Proportion of "l" responses as a function of the liquids' acoustics and the stops' acoustics (**a**) or the stops' identity (**b**)

effect of [d–g] acoustics shows that listeners' categorization of the liquid was not affected by the stops' acoustics.

Context categorization

In the second analysis, the categorization of the contexts as "l" versus "r" or "g" versus "d" were used as fixed effects in place of the context's acoustics.

Figure 5b shows that listeners responded "g" more often when they also responded "l" than when they responded "r," while Fig. 6b shows that the "l" responses did not differ depending on whether listeners responded "d" or "g."

Two-step hierarchies of mixed-effects models were again constructed for each set of responses. Including the interaction significantly improved fit in the model of stop responses, $\chi^2(5) = 19.193$, $p = .002$, even though the estimate for the interaction is not significant. The model with the interaction is interpreted here. The estimates in Table 4 show that listeners responded "g" significantly more often when they categorized the liquid as "l", and there is a nonsignificant trend for this effect of the context's categorization to become stronger as the stop becomes more [g]-like.

Comparison of the two models in the analyses of the liquid responses showed that including the interaction did not significantly improve the fit, $\chi^2(5) = 2.1498$, $p > .10$. The estimates in Table 5 also show that categorization of the following stop as "d" versus "g" did not significantly influence how often the listener categorized the liquid as "l" versus "r" any more than the stops' acoustics did (Table 3).[10]

Table 4 Interaction of acoustic and categorical effects on "g" versus "d" response proportions

	Estimate	*SE*	*z*	*p*
Intercept	0.93411	0.11676	8.000	1.24e-15
Stop acoustics=[d-g]	1.49580	0.07648	19.557	< 2e-16
Liquid categorization="l" versus "r"	0.10681	0.04419	2.417	0.0156
Stop acoustics by liquid categorization	0.04191	0.02705	1.549	0.1213

[10] We examined models like those described by Smits (2001a) that included the categorization of the context, as well as its acoustic value. For "g" versus "d" responses, adding the categorization of the liquid as "l" versus "r" significantly improved the fit of the model, $\chi^2(5)=12.711$, $p<.05$; however, the estimate for this fixed effect, −0.01055, was not significant, $p=.755$. The sign of the estimate for the liquid categorization is also opposite that in the model in Table 4, in which the liquid's categorization, but not its acoustics, was a fixed effect. Adding the categorization of the stop as "g" or "d" as a fixed effect to the model of the "l" versus "r" responses that included the stop's acoustics did not significantly improve the model's fit, $\chi^2(5)=5.6463$, p=.3422, nor was the parameter estimate significant (−0.02104, $p=.497$). Statistically, either the categorization of the liquid or its acoustics influences stop judgments, but not the two independently, while neither the stop's categorization nor its acoustics influences liquid judgments.

Table 5 Interaction of acoustic and categorical effects on "l" versus "r" response proportions

	Estimate	*SE*	*z*	*p*
Intercept	0.80356	0.12224	6.574	4.91e-11
Liquid acoustics=[l–r]	−0.92905	0.09254	−10.039	< 2e-16
Stop categorization="d" versus "g"	−0.02155	0.02315	−0.931	0.352

Discussion

Experiment 1 replicated Mann (1980), in that listeners responded "g" more often when the preceding liquid was more [l]-like. Listeners also responded "g" more often when they categorized the liquid as "l" rather than "r." However, neither the acoustics nor the categorization of the following stop influenced the categorization of the liquid, leaving only the liquids' own acoustics to affect how often listeners categorized it as "l" versus "r." The liquid serves as both an acoustic and a categorical context for categorizing the stop, but the stop does not serve as either kind of context for categorizing the liquid.

Finding that neither the acoustics nor the category of the stop influenced listeners' categorization of the preceding liquid could be interpreted as evidence in favor of Fowler's (2006) argument that a target sound can contrast only with a preceding context. However, Diehl and Walsh (1989), Mitterer et al. (2006b), Pisoni, Carrell, and Gans (1983), and Wade and Holt (2005) present results that can be interpreted as evidence of contrast between a target sound and a following context (but see Fowler 2006, p. 176, for additional arguments). In any case, it is a null effect and cannot, therefore, confirm a positive prediction of the contrast account.

This absence also fails to confirm a positive prediction of the gestural account that listeners use the acoustic perturbation of one sound produced by coarticulation with another as information about that coarticulatory source (Fowler 2005, 2006; Fowler and Smith 1986; Whalen 1984). If listeners attribute the acoustic effects of coarticulation to their source, they should treat a relatively low and [g]-like *F*3 in the stop or relatively high and [d]-like *F*3 as evidence that the preceding liquid is [r] or [l], respectively. Any evidence for such an attribution is, however, negligible (see Table 3 and Fig. 6).

The results of Experiments 2 and 3 are more decisive.

Experiment 2: Discrimination of HL-LH [alga]–[arda] versus HH-LL [alda]–[arga] pairs

The results of Experiment 1 do not reveal whether the context effects are produced by a criterion shift or exaggeration [Fig. 1(ii) or Fig. 1(iii)]. The discrimination experiment reported here tests the prediction of the exaggeration account that HL [alga] versus LH [arda] pairs should be more discriminable than HH [alda] versus LL [arga] pairs. This experiment follows up and replicates Stephens and Holt's (2003) finding that pairs of stimuli from a [da–ga] continuum and from a continuum consisting of just the first 80 ms of the *F*2 and *F*3 of the [da–ga] continuum were more discriminable when [l] preceded the more [ga]-like stimulus in the pair and [r] preceded the more [da]-like stimulus than when [r] preceded the more [ga]-like stimulus and [l] the more [da]-like stimulus. In the discussion, we argue that our findings, like those of Stephens and Holt, rule out compensation for coarticulation as the mechanism responsible for the liquids' effect on following stop-place judgments.

Method

Participants

Thirty-four new listeners from the same population as that in Experiment 1 participated in this experiment.

Stimuli

The stimuli in this experiment were drawn from the 5 × 5 subarray inside the 7 × 7 array used in Experiment 1 (Fig. 7; cf. Fig. 4). The different stimuli on a trial differed in two steps along either the [lg–rd] HL–LH or [ld–rg] HH–LL diagonals (gray and black lines in Fig. 7, respectively). Taking the values along the horizontal and vertical dimensions of the array to be [d, 1, 2, 3, 4, 5, g] and [l, 1, 2,3, 4, 5, r], where "d," "g," "l," and "r" represent the extreme values and, thus, the edges of the stimulus array, the different pairs along the [lg–rd] HL–LH diagonals were 31–13, 33–15, 42–24, 51–33, and 53–35 (circles connected by gray lines in Fig. 7), and those along the [ld–rg] HH–LL diagonals were 11–33, 13–35, 22–44, 31–53, and 33–55 (squares connected by black lines in Fig. 7).

Procedure

The four-interval same–different (4IAX) format was used so that listeners could use the auditory qualities evoked by the acoustic differences between the stimuli to discriminate them, rather than relying on categorical differences between them (Gerrits and Schouten 2004; Pisoni 1973). Listeners heard two pairs of stimuli on each trial, in which the stimuli in either the first or the second pair were different and those in the other pair were the same. The stimuli in each pair were separated by 250 ms, and the pairs by 500 ms. Listeners identified which pair was different, the first or the second. After the four stimuli

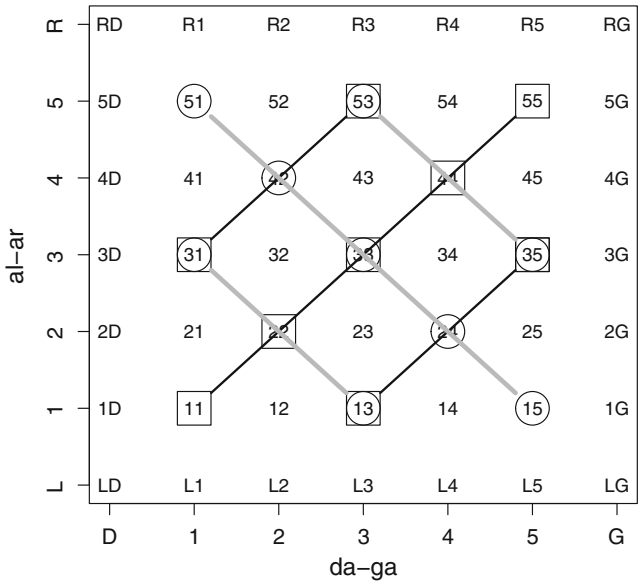


Fig. 7 The 7×7 [al–ar] by [da–ga] stimulus array from Fig. 4, in which circles and gray lines represent stimulus pairs along the HL–LH [lg–rd] diagonals, and squares and black lines represent stimulus pairs along the HH–LL [ld–rg] diagonals. The HH corner of the array is the lower left

were presented, the response prompts "1st" and "2nd" appeared immediately in red and blue on the left and right sides of the screen. Listeners then had 1,500 ms to identify the different pair by pressing either the corresponding red or blue button on the left or right of the response box. Feedback in the form of the correct answer appeared on the screen for 750 ms after the listener responded or the 1,500 ms had elapsed, whichever was earlier. The screen then went blank for 750 ms before the next sequence of four stimuli began.

Each of the 10 different stimulus pairs—for example, 31–31—was presented in a separate block of trials. Presenting just a single stimulus pair in a block reduces uncertainty and, like the 4IAX format, encourages responding to the acoustic, rather than categorical, differences between the stimuli (Macmillan et al. 1988). The order in which the 10 blocks were presented was counterbalanced using a balanced Latin square, and listeners were assigned randomly to a particular order.

Each block began with one repetition of the eight possible combinations of stimuli (for example, first pair different: 31–13—31–31, 13–31—31–31, 31–13—13–13, 13–31—13–13; second pair different: 31–31—31–13, 31–31—13–31, 13–13—31–13, 13–13—13–31) in random order. These eight trials were treated as training trials and were not included in the analysis. Test trials consisted of nine more repetitions of the eight stimulus combinations, for a total of 72 per stimulus pair. Procedures were otherwise the same as in Experiment 1.

On the one hand, if listeners categorized the sounds in the two intervals before discriminating them, the two kinds of possible different pairs should be equally discriminable, because HH–LL [alda]–[arga] pairs differ just as much in the category membership of the two consonants as HL–LH [alga]–[arda] pairs. On the other hand, if contrast between the liquid context and stop target influenced their responses, the HL–LH [alga]–[arda] pairs should be discriminated better than the HH–LL [alda]–[arga] pairs.

Analysis

The discriminability of each stimulus pair by each listener was measured by calculating d' values using Equation 4 (Micheyl and Messing 2006; Micheyl and Oxenham 2005)[11]:

$$d' = 2 \times \Phi^{-1}\left[1/2 + \sqrt{p(c)/2 - 1/4}\right], \quad (4)$$

where Φ^{-1} is the inverse of the cumulative standard normal distribution and $p(c)$ is the proportion of correct responses. These d' values served as the dependent variable in linear mixed-effects models in which the centered fixed effects were the diagonal from which the stimulus pair was drawn (lg–rd versus ld–rg, coded as 1 versus −1) and the halves of the stimulus array (d-half vs. center vs. g-half, coded as 1 vs. 0 vs. −1, and l-half vs. center vs. r-half, also coded as 1 vs. 0 vs. −1) in which the stimulus pair occurred (cf. Fig. 7). The random effects were the effects of listener on the intercepts and slopes of the fixed effects.

Results

Figure 8 shows that d' values were consistently larger for the [lg–rd] than for the [ld–rg] diagonal and in the d-half and center (LD, RD, C) than in the g-half (LG, RG).

A two-step hierarchy of models was again constructed. Including the interactions between the diagonal and the array halves did not significantly improve the fit of the data to the model for the difference in the number of parameters between them, $\chi^2(30)=24.612$, $p>.10$; moreover, none of the estimates representing the interactions among the fixed effects were significant. The simpler model without any of these

[11] Receiver operating characteristics (ROCs) in a *z*(Hits) by *z*(False Alarms) space were obtained for each of the conditions in this experiment by using each participant's hit and false alarm proportions. Straight lines fit the ROCs well enough that their responses can be considered to be normally distributed, and that d' can be used as the measure of discriminability, rather than an alternative such as A'. In any case, Macmillan and Creelman (1996) showed that, contrary to popular belief, A' is not free of distributional assumptions. Straight lines also fit the ROCs obtained in the same fashion for the results of Experiment 3.

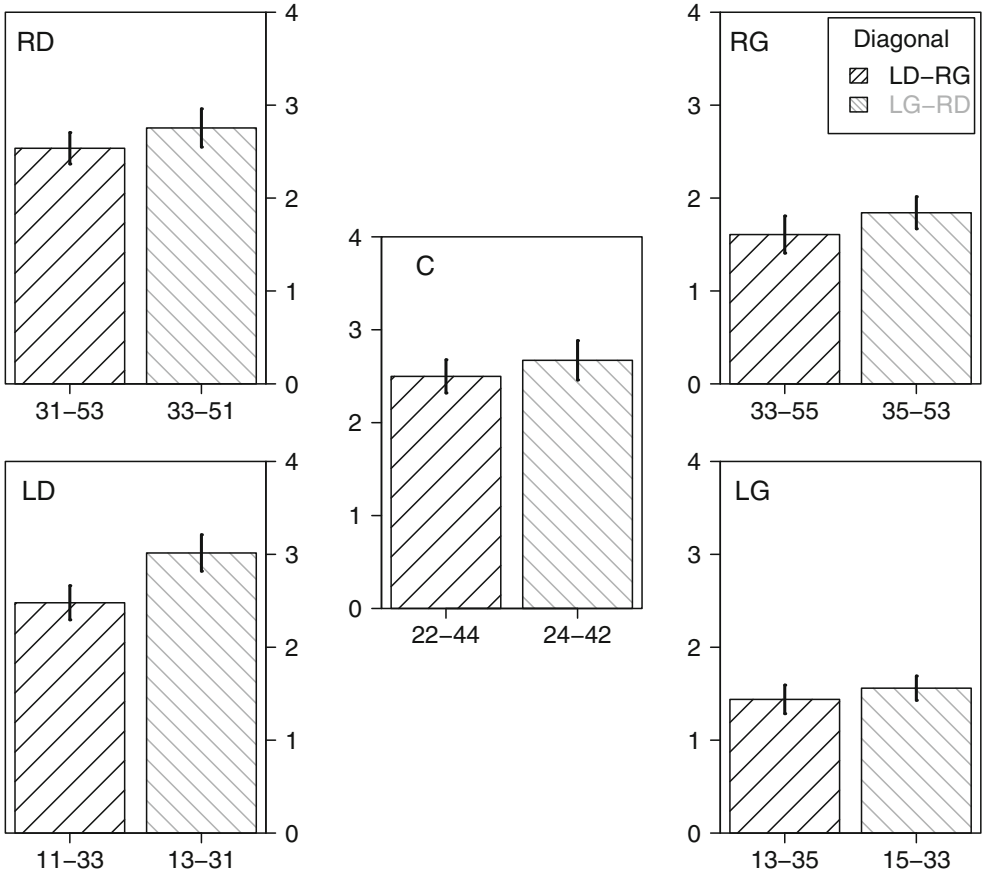


Fig. 8 Values of *d′* averaged across listeners (standard errors) for stimulus pairs along the HH–LL [ld–rg] diagonal (black, 45° hash lines) versus the HL–LH [lg–rd] diagonal (gray, 315° hash lines) by location in the stimulus space: "L" versus "C" versus "R"=l-half versus center versus r-half, and "D" versus "C" versus "G"=d-half versus center versus g-half. The panels' locations correspond to the arrangement in Fig. 7, the angle and color of the hash marks within each bar match the corresponding diagonals within that figure, and the stimulus pair represented by each bar/diagonal is listed below it

interactions is therefore interpreted here. The estimates in Table 6[12] show that *d′* values are significantly greater for stimulus pairs along the HL–LH [lg–rd] diagonal than for those along the HH–LL [ld–rg] diagonal: The significant difference of ± 0.1283 is a little more than .04 in proportion correct [estimated $p(c)$=.7915 for HL–LH [lg–rd] pairs versus .7513 for HH–LL [ld–rg] pairs]. The estimates show that *d′* values are also significantly higher in the d-half than in the center of the array [estimated $p(c)$ for d-half = .8493, as compared with .7718 in the center] and significantly lower in the g-half than in the center [estimated $p(c)$ for g-half = .6825 versus .7718) but that *d′* values do not differ significantly between the l- or r-halves and the center.[13]

Although the mean differences are in the same direction in all five regions of the stimulus space (Fig. 8), the difference between the HL–LH lg–rd and HH–LL ld–rg means is only large in the LD region. When we reran the analysis leaving out the stimuli in that region, the estimate for the HL–LH lg–rd versus HH–LL ld–rg comparison shrank from 0.12830 to 0.09338, and the corresponding *t*-value shrank from 2.586 to 1.905, which is only marginally significant, $p < .10$. One-tailed paired *t*-

Table 6 Independent effects of diagonal and location on *d′* values

	Estimate	*SE*	*t*	*p*
Intercept	2.24009	0.11114	20.156	<0.001
HL–LH lg–rd vs. HH–LL ld–rg	0.12830	0.04961	2.586	0.009
d-half vs. center vs. g-half	0.54211	0.07161	7.570	<0.001
l-half vs. center vs. r-half	−0.03086	0.05981	−0.516	0.572

[12] Because this is not a logistic-regression model, the intercept and estimates can be interpreted directly: The intercept estimates the grand mean *d′* value, and values of the fixed effect estimates are increments or decrements from this value; for example, for the [lg–rd] pair in the LD corner of the stimulus array, the predicted *d′* value = 2.24009+0.12830+(0.54211 ∗ 1)+(−0.030806 ∗ 1)=2.87969.

[13] The *p* values in this table are for a model that does not include random effects of listener on the slopes of the fixed effects, but only on the intercept, since Markov chain Monte Carlo sampling is not yet implemented in R for random effects on slopes. All *p* values reported for *t* tests below are obtained in this way. Alternatively, any $|t|$ value greater than 2 can be treated as significant (Baayen 2008, p. 270).

tests comparing the two diagonals yielded the following results for each region in the space: $t(33)$ = LD 2.1876, p = .01794; LG 0.8016, p = .2143; RD 1.1047, p = .1386; RG 1.3593, p = .09164; and C 1.0474, p = .1513. With α corrected to $\frac{0.05}{5} = 0.01$ for multiple tests, none of these differences is significant. Because there was no significant interaction between region and diagonal, we take the results of these post hoc tests to mean that the size, but not the direction, of the HL–LH lg–rd advantage over HH–LL ld–rg differs between regions in the stimulus space (Nieuwenhuis, Forstmann, and Wagenmakers 2011). A larger N should therefore provide the power needed for each pairwise comparison to reach significance.

Discussion

Summary

The results of this experiment confirm the prediction of the exaggeration account that stimulus pairs would be more discriminable when the context differs enough acoustically from the target to contrast auditorily with it. They also replicate Stephens and Holt's (2003) finding that pairs of stimuli from a [da–ga] continuum are more discriminable when the energy distribution in the preceding liquid's spectrum differs from, and can thus contrast auditorily with, that in the following stop's—that is, when [l] precedes the more [ga]-like stimulus in the pair and [r] precedes the more [da]-like stimulus, rather than vice versa. These new results extend their finding by showing that liquids that are not categorically different from one another also exaggerate the perceived difference between the following stops when the energy distributions in their spectra differ in frequency from those in the stops' spectra.

In the final part of this discussion, we examine the reasons why the greater discriminability of contrasting HL–LH [alga–arga] pairs probably cannot be attributed to [l] and [r] causing independent criterion shifts (cf. Norris 1995). Before doing so, we lay out in more detail here the reasons why these results cannot be attributed to compensation for coarticulation as formulated in the direct-realist account of context effects. The argument has two parts; the first describes the listener's behavior in direct-realist terms, and the second identifies the only mechanism compatible with direct realism that could produce that behavior. Because that mechanism could not exaggerate the perceived value of the stop's spectrum when it differs from the preceding liquid's spectrum, it cannot produce the observed improvement in discriminability.

Why compensation for coarticulation cannot improve discriminability

First, a preceding [l] would pull a [g] forward, making it sound more like [d] if it were taken out of that context, and a preceding [r] would pull a [d] backward, making it sound more like [g] out of that context. Each context would therefore make one of the stops more ambiguous. In the gestural account, listeners correct for these coarticulatory effects perceptually, undoing the fronting of [g] by [l] and the backing of [d] by [r] by attributing those shifts in the location of the stops' articulations to the preceding liquid contexts. Fowler (2006; Fowler and Smith 1986) aptly describes these perceptual effects as listeners "parsing" the acoustic properties of the signal into the contributions of the target's and context's articulations. Completely successful parsing would at most return [g] and [d] perceptually to their uncoarticulated state, not exaggerate how [g]- or [d]-like they are. Second, the only mechanism that can correct the percept in the direct-realist account is a shift in the decision criterion for deciding whether energy is concentrated low enough in the stop's spectrum to convey that the stop is [g]; this criterion would be raised after [l], increasing the portion of the continuum that is identified as "g" [Fig. 1(ii)], and lowered after [r], increasing the portion that is identified as "d." That is, the only mechanism that is compatible with the direct-realist account of context effects is one that changes the likelihood that a listener will assign a particular stimulus to one category rather than another. Any mechanism other than a criterion shift is ruled out by direct realism's assumption that a speech sound's acoustic properties inform the listener about its articulation, which is what is perceived using the information in those acoustic properties. For those properties to inform the listener about the articulation that produced them, they must be perceived undistorted, but exaggeration of their perceived values would distort them. This is, we think, the principal reason why Fowler (2006) argued against the auditorist proposal that the target sound contrasts with its context: contrast is a distortion, or a source of heteromorphy that would render the target sound's acoustic properties less veridical and, thus, less informative about the articulation that produced them.

If different contexts shift the criterion for categorizing the target in opposite directions and those shifts, in turn, influence discriminability in the ways we discuss in the next section, then direct realism would certainly predict/permit that influence, because a criterion shift does not alter the acoustic properties' perceived values but, instead, only the category that a stimulus with those values is assigned to. Thus, we do not argue that the direct-realist account makes no prediction about context effects in a discrimination task but, instead, that it does not and cannot predict that contexts could exaggerate the perceived values of neighboring ambiguous targets.

Why a criterion shift cannot improve discriminability

If the perceived values of the stops' spectra are exaggerated when their energy concentrations differ from those of a

preceding liquid, any account is ruled out in which context only shifts the criterion for categorizing the stimuli, not just the direct-realist account. Here, we discuss the mechanism proposed by Holt (2005, 2006) to account for the results of experiments in which a [da–ga] continuum was preceded by a long sequence of pure tones whose frequencies were distributed uniformly about a high mean frequency of 2800 Hz, like [l]'s *F*3, or about a low mean frequency of 1800 Hz, like [r]'s *F*3. Listeners responded "ga" to more of the continuum after the sequence of high pure tones than after the sequence of low ones.

Holt (2006) attributed this change in response likelihoods to "stimulus-specific adaptation" by neurons in the auditory cortex to the mean of the tone sequence's distribution. Like adaptation at any other stage in the auditory system or in any sensory modality, stimulus-specific adaptation is a decrease in responsiveness of neurons tuned to stimulus properties or values that have been encountered recently. In the experiments reported in Holt (2005, 2006), these are the high or low means about which the frequencies of the tones in the sequences are distributed. Adaptation's effect is *subtractive*: The neurons tuned to the mean frequency of the preceding tone sequence become less responsive. Unadapted neurons tuned to other frequencies would thereby become more responsive, relative to the adapted ones.

If the listener is presented with a stop that is ambiguous between [d] and [g]—that is, if energy is not concentrated in the stop's spectrum any more at high than at low frequencies—then a preceding series of high-frequency tones would decrease responsiveness in neurons that would otherwise be excited by whatever energy was present at higher frequencies in the stop's spectrum. This decrease in responsiveness could be modeled as an increase in the variance (1.5) of the higher, more [d]-like response likelihood distribution on the left in Fig. 9b, as compared with its variance (1) in Fig. 9a, which displays the unadapted state. An increase in variance captures the reduced responsiveness of the adapted neurons by both decreasing the likelihood of a "d" response and increasing the likelihood of a "g" response when the more [d]-like stimulus is presented, as shown by the increase in the false alarm proportion on the right-hand side of Fig. 9b.

Of greater importance for our argument that stimulus-specific adaptation would not increase sensitivity to place differences between adjacent stimuli along the [d–g] continuum, the increased variance does not also increase the hit proportions, as compared with the unadapted state. As a result, the d' value is smaller (Fig. 9b) than when no adapting context precedes the stop (Fig. 9a). Figure 9c shows that d' values only increase relative to the unadapted state if the variance of the "g" response likelihood decreases at the same time as that of the "d" response likelihood distribution increases. But if stimulus-specific adaptation is an entirely subtractive process and the responsiveness of the unadapted neurons only increases relative to the adapted neurons and not absolutely, the variance of the "g" response likelihood distribution should not change, and sensitivity to the difference between a more [d]-like and a more [g]-like stimulus can only decrease.

So far, we have not treated the effect of stimulus-specific adaptation as a criterion shift, but as the c values in Fig. 9b, c show, the changes in variance also change the response bias in favor of more "g" responses. So when modeled as an increase in variance, stimulus-specific adaptation predicts the same change in the listener's categorization of an ambiguous stop as shifting the decision criterion to the left. This shift is an increase in sensitivity to the energy concentrations at lower frequencies in the stop's spectrum, but not an increase in sensitivity to the difference between that stop and one next to it along the [d–g] continuum, unless adaptation not only subtracts from the responsiveness of the adapted neurons, but also adds to the responsiveness of unadapted neurons nearby.

Our characterization of the effects of adaptation as a criterion shift appears to be challenged by the finding that adaptation occurs early enough to be preattentive. This challenge can be met by distinguishing temporally between the cause, adaptation, and its effect, a criterion shift. We consider adaptation, wherever or whenever it occurs in the sensory encoding of a stimulus, to be the neural *mechanism* responsible for a possible criterion shift, which may only become evident later when the listener actually decides what category the stimulus belongs to. Adaptation is not the criterion shift itself, but merely one among possibly many possible contributors to that shift. Because other effects may intervene before the listener responds, the criterion may shift more, less, not at all, or even in the opposite direction, regardless of how neurons were adapted earlier. We therefore do not dispute that adaptation is early or preattentive, because the neural mechanism need not be simultaneous with its psychophysical effect. It is only necessary for the mechanism to create the conditions for that effect to emerge, whenever or wherever the listener finally pools the outputs of all the mechanisms that might bias the response and decides which category to assign the stimulus to.

If stimulus-specific adaptation can only have the effect of shifting the decision criterion and not also increase sensitivity to the difference between adjacent stimuli along the [d–g] continuum, how can the results of Experiment 2 and those reported by Stephens and Holt (2003) be explained? Our answer is that a spectrally different context must push away the mean of the response likelihood distribution corresponding to the target, as in Fig. 1(iii); that is, contrast between the target and its context must exaggerate the target's value along the relevant perceptual dimension. Still missing from this explanation is a neural mechanism that would produce this exaggeration.

Do criterion shifts in opposite directions increase sensitivity?

At the beginning of this article, we used detection-theoretic models to argue that a criterion shift induced by a neighboring

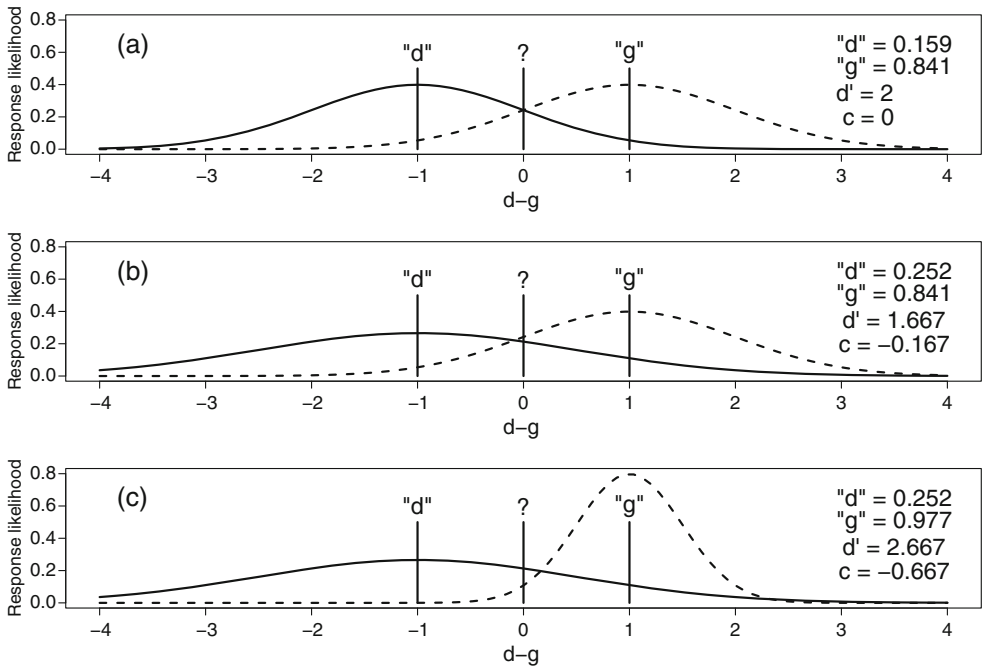


Fig. 9 "d" (solid lines) and "g" (dashed lines) response likelihood distributions when (**a**) there is no preceding context, (**b**) a preceding context has adapted neurons tuned to the higher frequencies and increased the variance of the higher "d" response likelihood distribution on the left to 1.5, or (**c**) the variance of the lower "g" response likelihood distribution on the right is also decreased to 0.5. On the right hand side in each panel are listed the proportions of the "d" and "g" response likelihood distributions to the right of the decision criterion, which equal the false alarm and hit proportions, along with the corresponding *d′* and *c* values

sound, the context, would not change sensitivity to the difference between two target stimuli and that only an increase in the perceptual distance between the response-likelihood distributions—that is, exaggeration of the stimuli's perceived values—could do so. A single criterion shift, like that introduced by one neighboring sound, could not increase sensitivity, but if the target sounds occur next to different contexts and those contexts shift the criteria for categorizing their respective targets in opposite directions, then sensitivity should appear to change, too. Since the contexts always differed in the stimuli used in the discrimination task run in Experiment 2, it is therefore possible that listeners' greater sensitivity to the differences in HL–LH [lg–rd] pairs, as compared with HH–LL [ld–rg] pairs can be attributed to criterion shifts, rather than exaggeration of the stimuli's perceived spectral differences in HL–LH [lg–rd] pairs. Figure 10 shows how this alternative could arise. (We assume that the stops are the target sounds and the preceding liquids are their contexts because stop categorization did not influence liquid categorization in Experiment 1.)

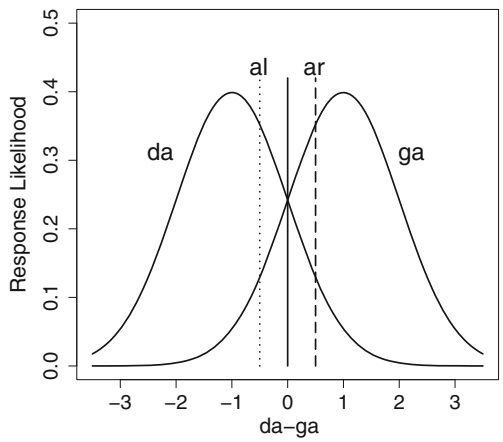


Fig. 10 "da" and "ga" response-likelihood distributions and criteria shifted relative to the neutral criterion (solid line) by a preceding "ar" (dashed line) versus a preceding "al" (dotted line)

If the listener uses the left-hand criterion after [l] and the right-hand one after [r], the probability of a hit increases and the probability of a false alarm decreases for HL–LH [alga–arda] pairs, while the probability of a hit decreases and the probability of a false alarm increases for HH–LL [alda–arga] pairs, such that predicted *d′* values increase from 2 to 3 for the HL–LH [alga–arda] pairs and decrease from 2 to 1 for HH–LL [alda–arga] pairs. The criterion shifts induced by the [l] and [r] contexts alone thus predict a *d′* value that is three times larger for the HL–LH [lg–rd] pair than for the HH–LL [ld–rg] pair. This modeling exercise does not mean that the context could not also exaggerate the perceived value of the target, but it does mean that the better discriminability of HL–LH [lg–rd] pairs, as compared with HH–LL [ld–rg] pairs, obtained in Experiment 2 may not depend on nor necessarily be diagnostic of such exaggeration.

This case is a simpler version of a scenario that Norris (1995) used to argue that multiple criterion shifts, each

induced by a different lexical item, could create the illusion that feedback from the lexicon increased perceptual sensitivity in a word recognition task.[14] It is simpler in that decisions are made with respect to the stimuli's values along a single perceptual dimension, rather than multiple dimensions, but similar in that listeners implicitly choose different decision criteria depending on the context in which each target occurs, and those choices create the illusion of an increase in sensitivity to differences between the targets. As measured by *d′* values, sensitivity does appear to change in opposite directions depending on whether the targets differ spectrally from their contexts; these changes are illusions of changes in the targets' perceived values if the means of their response likelihood distributions have not shifted, but only the likelihoods of categorizing them as "d" versus "g."

We nonetheless think that such criterion shifts are not responsible for the difference in discriminability between HL–LH [lg–rd] and HH–LL [ld–rg] pairs in our experiment, because this difference does not correspond to the likelihood of the listener categorizing the liquid as "l" versus "r." For the [al]-like context to increase the "ga" response likelihood and the [ar]-like context to increase the "da" response likelihood in the way modeled above, the listener must be able to distinguish the two contexts from one another. The size of the resulting criterion shifts should, moreover, depend directly on the extent to which they can distinguish these contexts. We have no direct measure of how well the participants in Experiment 2 distinguished the liquid contexts from one another, because the liquid always covaried with the stop. We did, however, measure categorization of the liquids in Experiment 1 by a different group of listeners. Since the stimuli were the same, the performance of that group of listeners can be used to predict roughly how well listeners could have distinguished the various pairs of liquids in Experiment 2.

Averaging across the negligible differences between the different stop contexts, listeners in Experiment 1 categorized steps 1, 3, and 5 on the [l–r] continuum as "l" on 0.887 (*SE*= 0.067), 0.719 (0.094), and 0.238 (0.089) of trials, respectively (see Fig. 6). The difference in proportions is three times larger between steps 3 and 5 (.481) than it is between steps 1 and 3 (.168); the corresponding differences expressed in *d′* values are 1.293 for steps 3 and 5 versus 0.631 for steps 1 and 3, a ratio of 2:1. This difference predicts that listeners in Experiment 2 should distinguish the more [l]-like from the more [r]-like context better for stimuli in the r-half of the stimulus array than for those in the l-half and that the criterion shifts should therefore also be greater in that half of the array. A glance at Fig. 8 shows that, contrary to this prediction, the advantage of HL–LH [lg–rd] pairs over HH–LL [ld–rg] pairs is no larger in the [r]-half of the stimulus array than in the [l]-half. The analysis reported in Table 6 above also showed that the diagonal from which the stimulus pair was drawn did not interact significantly with either the l–r or the d–g halvings of the stimulus array. That HL–LH [lg-rd] pairs were no more discriminable from HH–LL [ld–rg] pairs in the r-half of the array than in the l-half disconfirms a positive prediction of the criterion shift explanation.

This argument depends on listeners in Experiment 2 being as likely to distinguish two liquids as those who categorized the liquids in Experiment 1. The stimuli were the same, so it would not be surprising that these likelihoods would correspond. Moreover, the difference in "l" response proportions was so much larger between steps 3 and 5 than between steps 1 and 3 that some evidence should have emerged that the liquids were distinguished better and the criterion shifts were correspondingly larger in the r- than in the l-half of the stimulus array if the difference in discriminability between HL–LH [lg–rd] pairs versus HH–LL [ld–rg] pairs were produced by such criterion shifts. More direct tests of this alternative certainly are needed, in which discrimination and categorization data are collected from the same listeners, and they discriminate stimuli that differ in just the liquid or the stop.[15]

Experiment 3: Discrimination in nonspeech contexts

Experiment 2 showed that the energy distribution in the context's spectrum can exaggerate the perceived value of the energy distribution in the target's spectrum when they differ. That experiment did not show, however, that exaggeration is caused by *auditory*, as opposed to linguistic, processing of the stimuli or by the interpretation of the stimuli's acoustic properties as evidence about the coarticulation of the target with its context.[16] That is the purpose of this experiment, which replaces the original speech contexts with nonspeech analogues—specifically, pure tones whose frequencies mimic the spectral differences between [l] and [r]. This experiment has the same rationale as Lotto and Kluender's (1998) Experiment 3 in which pure tones replaced the original [al] and [ar] syllables. They found that listeners responded "ga" more often

[14] We are grateful to Lori Holt for alerting us to the relevance of Norris's (1995) paper.

[15] The format in which stimuli were presented for discrimination, 4IAX, and the presentation of just a single stimulus pair in each block of trials should also have encouraged listeners to respond in terms of the stimuli's auditory qualities, rather than the categories to which they might be assigned.

[16] It did at least show that responses could not be based on the linguistic categories to which the target and context were assigned, since that would have predicted, incorrectly, that HL–LH [lg–rd] pairs would be no more discriminable than HH–LL [ld–rg] pairs.

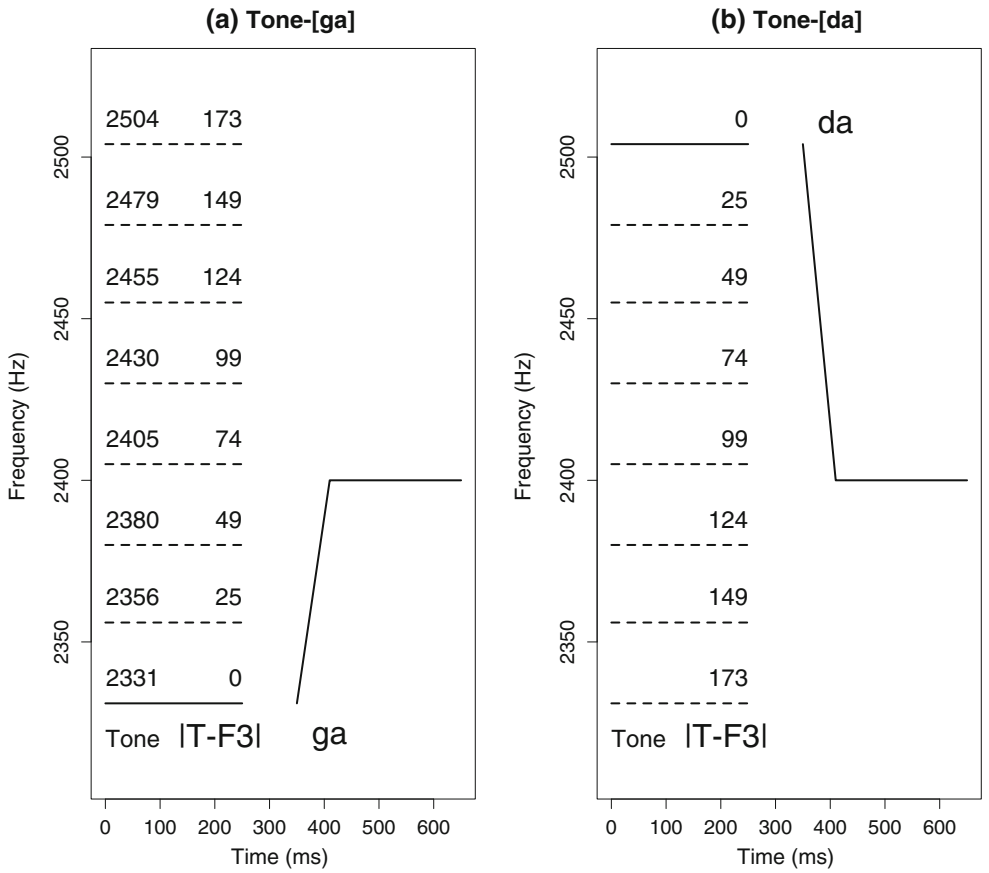


Fig. 11 (**a**) An exemplary [g]-like *F*3 trajectory preceded by pure tones differing in frequency by 0–173 Hz |*T*−*F*3| in 24- to 25 Hz steps (**b**) the corresponding [d]-like exemplar, also preceded by tones differing in frequency from 0 to 173 Hz

after the higher pure tone than after the lower one, just as they responded "ga" more often after [al]'s high *F*3 than after [ar]'s low *F*3, and they interpreted this result as evidence that auditory contrast between the target syllable and the nonspeech context shifts the response bias. This shift could not be the product of compensation for coarticulation because listeners would not perceive a syllable as coarticulating with a pure tone. Because auditory contrast between context and target could also explain the response biases they obtained with speech contexts in their Experiment 1, Lotto and Kluender argued that it provided the most parsimonious account of the response biases they observed with speech as well as nonspeech contexts. Experiment 3 is also the mirror image of Stephens and Holt's use of speech contexts to produce contrast with nonspeech targets.

We argued above that compensation for coarticulation, as it is conceived in the direct-realist account of context effects, cannot change sensitivity, even when the contexts are speech, because a change in sensitivity distorts the percept of the signal's acoustic properties and, thus, corrupts the information those properties would convey about the articulations that produced them. However, a less theory-bound alternative could explain the results of Experiment 2 as a product of compensation for coarticulation when the contexts are speech. Lotto and Kluender (1998) showed that nonspeech contexts affect categorization, or in the terms of this article, the listener's criterion for deciding whether the following syllable began with [g] or [d]. Experiment 3 tests whether such contexts are also capable of altering sensitivity to the difference between a more [g]-like and a more [d]-like stop—specifically, whether the two stops become more discriminable as the frequencies of preceding pure tones differ more from the stops' *F*3 values. It thereby intertwines this article's goals: determining whether contexts that differ spectrally from their targets can increase sensitivity to differences in neighboring targets and whether those sensitivity changes arise from the auditory response to the signal's acoustic properties, and not from their possible articulatory origins.

Method

Participants

Twenty new listeners from the same population as Experiments 1 and 2 participated in this experiment.

Stimuli

In this experiment, the stimuli consisted of a 250-ms-long pure tone followed by a syllable. The RMS amplitude of the tone equalled the average of the original [al–ar] syllables. The frequencies of the tone and of the onset frequency of the syllable's *F*3 varied in 11 equal linear steps from a minimum, relatively low [g]-like value of 2260 Hz to a maximum, relatively high [d]-like value of 2504 Hz. These values are in the middle of the [d–g] continuum used in Experiments 1 and 2, and the stops were thus neither strongly [g]- nor [d]-like. The interval between the end of the tone and the onset of the formants in the syllable was 100 ms. It consisted of 15 ms of silence, followed by 85 ms of low-frequency periodic energy simulating closure voicing. On each trial, a pair of these stimuli were played.

As is shown in Fig. 11, the syllables' *F*3 onset frequencies in a pair always differed by seven equal linear steps (173 Hz), and the tones' frequencies corresponded to each of the eight 24 to 25 Hz steps between that pair of stops' *F*3 onset frequencies. For example, if the syllables' *F*3 onset frequencies were 2331 and 2504 Hz (the highest pair), the preceding tones' frequencies varied from 2331 to 2504 Hz (for this pair of *F*3 onset frequencies, see Table 7 as well as the figure) in zero to seven steps or differed from the stop's *F*3 values from 0 to 173 Hz in 24 to 25 Hz steps. The three lower pairs of [g]-like and [d]-like *F*3 onset frequencies were 2307 and 2480, 2282 and 2455, and 2260 and 2433 Hz, and the tones' frequencies varied accordingly, like those illustrated in Table 7 and Fig. 11.

As Table 7 and Fig. 11 show, the tone's frequency was always greater than or equal to the *F*3 onset frequency of the more [g]-like stop in the pair and less than or equal to the *F*3 onset frequency of the more [d]-like stop. Thus, when the tone's frequency differed from the stop's *F*3, the *F*3 percept could contrast with the preceding tone, and if it did, the [g]-like stop could sound even more like [g] and the [d]-like stop even more like [d]. As is shown in the table, as the tone before the [g]-like stop increased in frequency and became progressively more different from the stop, the one before the [d]-like stop decreased in frequency and likewise became progressively more different from the stop. If the percept of the stops' *F*3 onset frequencies contrasts with the frequency of the preceding tone, listeners should discriminate a pair that combines the highest tone (in the example, 2504 Hz, the highest tone in Fig. 11a) with the [g]-like stop and the lowest tone (2331 Hz, the lowest tone in Fig. 11b) with the [d]-like stop (bottom row, next-to-last column) better than the pair that combines the lowest tone (2331 Hz) with the [g]-like stop and highest tone with [d]-like stop (top row), even though the differences in the tones' frequencies are equally large. Table 7 also shows that the stimuli can be grouped in pairs by the size of the difference between the tones (last column). If listeners, instead, discriminate the stimuli by the difference in tones alone, they should be better at discriminating pairs of stimuli in which frequency differences between the two tones are large (closer to the top and bottom rows) and equally good for pairs in which the absolute values of the frequency difference between the tones, $|tone_g - tone_d|$, are equal (values the same number of rows up and down from the middle).

Table 7 Exemplary stimulus characteristics: Tone and *F*3 onset frequencies in Hz, the absolute values of the differences between tone and *F*3 onset frequencies, and the frequency difference in steps between the tone preceding the [g]-like stop and that preceding the [d]-like stop, for the [g]- and [d]-like stops with the highest pair of *F*3 onset frequencies

$tone_g$	[g]-like stop	$tone_d$	[d]-like stop	$\lvert tone - syllable \rvert$	$tone_g - tone_d$
2331	2331	2504	2504	0=0	−7=−173
2356	2331	2479	2504	1=25	−5=−123
2380	2331	2455	2504	2=49	−3=−75
2405	2331	2430	2504	3=74	−1=−25
2430	2331	2405	2504	4=99	1=25
2455	2331	2380	2504	5=124	3=75
2479	2331	2356	2504	6=149	5=123
2504	2331	2331	2504	7=173	7=173

Procedure

A trial consisted of a presentation of a pair of tone–syllable sequences, separated by a 500-ms silent interval, followed by the appearance of the response prompts "1" and "2" on the left and right sides of an LCD screen. These prompts remained on the screen for 2,000 ms or until the listener responded, whichever was earlier. Listeners clicked the corresponding button on the response box to identify which stimulus contained their target syllable. On training trials, feedback in the form of the words "correct" or "incorrect" was displayed on screen for 1,000 ms immediately afterward; no feedback was provided on test trials. The two-interval forced choice (2IFC) task was used because it requires fewer trials and is generally easier for participants than the AX (same–different) task, so long as they understand what categories the stimuli are to be assigned to. In addition, listeners' attention was not drawn in any way to the preceding tones.

The experiment began with 32 training trials in which listeners heard two repetitions of all four possible pairs of [g]-like and [d]-like stops in both orders, preceded by tones whose frequencies differed from the following stops' *F*3 onset frequencies by zero or seven steps. Each block of test trials consisted of a single presentation of each of the four possible pairs of [g]-like and [d]-like stops in both orders, preceded by the eight possible tones (a total of 64 trials per block). A total of eight test blocks were presented, which yielded 16 trials for

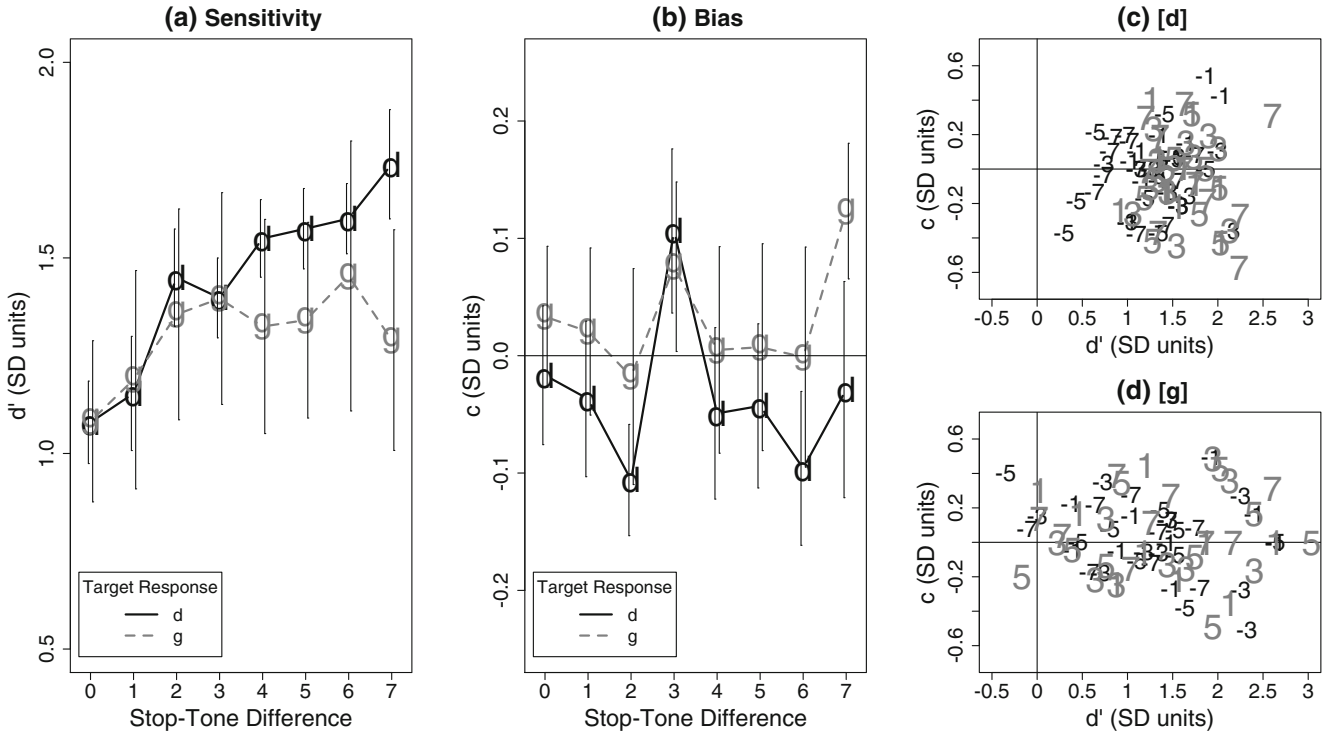


Fig. 12 (**a**) Averaged *d′* values (standard errors) and (**b**) averaged *c* values (standard errors) as a function of the size of the frequency difference between the preceding tone and the stop's *F*3 onset frequency (in steps) and whether the target response was "d" (black) or "g" (gray). *c* by *d′* values for (**c**) [d]-listeners and (**d**) [g]-listeners. The plotting symbols represent the sign and the size of the frequency difference between the tones as listed in the last column of Table 7 (smaller black=positive and larger gray=negative differences)

each stimulus pair. Listeners took self-timed breaks between blocks of trials.

Eleven "[d]" listeners identified which stop in the two syllables, the first or second, sounded more like [d], and 9 "[g]" listeners which sounded more like [g].

Procedures were otherwise the same as those in Experiments 1 and 2.

Analysis

Responses were pooled across the four possible [g]-like and [d]-like pairs to obtain enough trials (64) to calculate reliable *d′* values. Since the format is 2IFC, the formula in Equation 5 was used to calculate these values (Macmillan and Creelman 2005):

$$d' = [z(Hits) - z(False\ alarms)]/\sqrt{2} \quad (5)$$

Table 8 Interaction of stop-tone frequency difference and the target response

	Estimate	*SE*	*t*	*p*
Intercept	−0.0070	0.1333	−0.052	>0.10
g versus d	0.0696	0.1333	0.522	>0.10
Stop-tone difference	0.0299	0.0060	4.986	<0.0001
g versus d × stop-tone difference	0.0140	0.0060	2.342	0.0089

Results

Figure 12 shows that *d′* values increased with the frequency difference between the tones and the *F*3 onset frequencies of the following stops, and they increased more for the [d] than the [g] listeners.

A two-step hierarchy of linear mixed-effects models was again constructed. The dependent variable was the *d′* values, centered by subtracting the mean from each value, and the fixed effects were the centered target response ("g" = −1, "d" = 1) and the stop-tone differences (centered by subtracting 7 from twice the difference in steps between the stop and the tone). The random effects were the effects of listener on the intercept and slopes of the fixed effects. Listener is nested within target response in these random effects.

The model that included the interaction between the target response and the stop-tone difference fit significantly better, $\chi^2(8) = 16.018$, $p = .0421$, and the estimate for this term was also significant, so this model is interpreted. The estimates in Table 8 show that the stimuli became significantly more discriminable as the frequency difference between the tones and the following stops' *F*3 onsets increased and that this increase was significantly greater for the [d] than for the [g] listeners. The two groups of listeners did not differ significantly in their overall ability to discriminate the stimuli. This is the result predicted by the contrast, but not the compensation account.

Were listeners using the tones alone to discriminate the stimuli?

There is a subtle way in which listeners could have used the tones to discriminate the stimuli rather than differences in the perceived values of the syllables' *F*3 onset frequencies. Since the tone's frequency was always higher than or equal to *F*3's onset frequency in the more [g]-like stop in a pair and lower than or equal to *F*3's onset frequency in the more [d]-like stop, listeners could have used either the direction of the difference between the tone and the stop's *F*3 onset frequency or which of the tones was higher or lower to predict which syllable was more [g]- or [d]-like. The first strategy depends on listeners being able to compare the pure tone's frequency with the syllable's *F*3 onset frequency, which may be difficult to do, while the latter depends only on the much simpler task of determining which of the two tones is higher (or lower). The latter strategy would become especially effective as the tones' frequencies came to differ more, and it would thus appear to yield results very much like those obtained.

This strategy also predicts a noticeable bias toward identifying the more [g]-like stimulus as the one preceded by the higher tone and the more [d]-like stimulus as the one preceded by the lower tone. In its simplest form, where listeners use the tones alone, they would be equally likely to mistake the second stimulus for the more [g]-like one in a 2331–2331 low-[ga] and 2504–2504 high-[da] tone–syllable pair as to correctly identify that stimulus as more [g]-like in a 2331–2504 low-[da] and 2504–2331 high-[ga] pair. For [g] listeners, such a bias would show up as increasingly negative values for the criterion *c* as the tone difference between the two stimuli becomes more positive, while for [d] listeners, *c* values should instead become more positive as the tone difference becomes more positive.

Figure 12b confirms neither prediction: *c* values do not become more positive for [d] listeners, nor do they become more negative for [g] listeners, as the difference in tones became more positive. Figures 12c, d show that bias is unrelated to sensitivity for both groups of listeners [[d] listeners, $r = .0081$, $t(86) = 0.0753$, $p > .10$; [g] listeners, $r = -.1064$, $t(70) = -0.8958$, $p > .10$].

A two-step hierarchy of linear mixed effect models was again constructed in which the bias measure *c* calculated from grouping the stimuli by the frequency difference between the tones served as the dependent variable. Because this measure already varies about 0, it was not centered. The fixed effects were the centered target response and the tone difference (in steps), and the random effects were the effects of listener on the intercepts and slopes of the fixed effects. None of the estimates were significant in either model ($p > .10$), and adding the interaction term did not significantly improve the fit of the model to the data.

Discussion

These results show that the stops differing in their *F*3 onset frequencies become more discriminable as the difference between the preceding tones and the stops' *F*3 onset frequencies gets larger. They support interpreting the finding in Experiment 2 that HL–LH [lg–rd] pairs are more discriminable than HH–LL [ld–rg] pairs as evidence that the perceived value of the stops' *F*3 onset frequencies is exaggerated during prelinguistic auditory processing when the energy distribution in the spectrum of the preceding context differs. This result dovetails with that obtained in Experiment 2, where a spectrally different speech context exaggerated the perceived value of the *F*3 onset frequency of a following stop, and in doing so, confirms the hypothesis that this effect of context is auditory in origin. It also confirms the hypothesis that contrast with its context can exaggerate the target's perceived value for an acoustic property, and not only shift the criterion for categorizing the target. Further analysis tested an alternative explanation in which listeners used the difference in tones to determine which syllable was more [d]- or [g]-like. A difference in the relationship between response bias and the size and sign of the difference between the two tones in a stimulus predicted by this alternative did not materialize.

General discussion

Summary

Experiment 1 replicated Mann's (1980) original finding that listeners respond "ga" more often after [al] than after [ar]. The likelihood of a "ga" response increased continuously as the liquid became more [l]-like. When listeners categorized the liquid as [l], they were also more likely to categorize the stop as [g]. However, neither the acoustics nor the categorization of the stop influenced the categorization of the liquid.

Experiment 2 showed that the context exaggerated the perceived value of the target, rather than merely shifting the criterion for deciding which category it belonged to. Stimulus pairs in which the liquid and stop differed spectrally (HL–LH [lg]–[rd] pairs) were more discriminable across the stimulus space than those in which the liquid and stop were spectrally similar (HH–LL [ld]–[rg] pairs). By replacing the liquid contexts with pure tones that mimicked the acoustic difference between [l] and [r], Experiment 3 showed that exaggeration originates in the auditory response to the stimuli. The discriminability of [da] from [ga] syllables increased as the frequency difference between the stops' *F*3 onset frequencies and the preceding tones increased.

Taken together, these results support the auditory contrast account of context effects in speech perception, rather than the compensation for coarticulation account. They also show that auditory contrast between context and target does not just shift the criterion for categorizing the target but actually exaggerates the perceived value of the energy distribution in its spectrum.

In this article, we have argued that the perceptual effects of a sound's context on its perception are best accounted for as a product of contrast between the acoustic properties of the target sound and its context. We have also argued that the perceptual effect of auditory contrast between context and target cannot be limited to shifting the criterion for deciding what category the target sound belongs to but must also exaggerate the perceived value of the target's affected acoustic properties to account for listeners' greater sensitivity to acoustic differences between adjacent targets when they can contrast with their contexts. Finally, we have argued that contrast is an auditory effect, and not compensation for coarticulation, because nonspeech contexts also increase sensitivity to differences between speech targets. We do not mean to suggest, however, that the only kind of perceptual effects of a sound's context are those produced by online, auditory contrast between a target sound and its context or even that listeners may not, on occasion, compensate for coarticulation. In the next section, we distinguish once again between the effects of a criterion shift and exaggeration by reviewing results reported by Holt (2005, 2006). The three sections that follow review evidence that context effects can arise in other ways than contrast.

A criterion shift or exaggeration?

Holt (2005, 2006) showed that following a long series of brief pure tones whose frequencies varied randomly about a high frequency mean, listeners responded "ga" more often to a [da–ga] continuum than after an otherwise similar series whose frequencies varied randomly about a lowfrequency mean. In both high and low contexts, the last tone in the series was intermediate in frequency, so the tone that immediately preceded the syllable was neither high nor low and, thus, not the source of contrast. Holt (2005) showed that response biases did not diminish as the interval between the end of the high or low tone series and the syllable lengthened to 1.3 s. The persistence of the bias was also unaffected by whether that interval was lengthened by adding silence or repetitions of the intermediate tone. Holt (2006) attributed the response biases induced by the tone series to "stimulus-specific adaptation" at a central stage in the auditory pathway, possibly in the primary auditory cortex, where the time constants for adaptation are relatively long. Adaptation is a product of depressing neural responses to repeated characteristics of the auditory input—in this case, the mean frequency of the tone series (Ulanovsky, Las, and Nelken 2003; Ulanovsky, Las, Farkas, and Nelken 2004). Adaptation is also most readily interpreted as the neural mechanism that makes a criterion shift possible later, because it reduces the likelihood of neurons responding at one range of frequencies, as compared with others, and thus the likelihood of the response associated with that frequency range.

The results of Experiments 2 and 3 show that the context does more to the percept of the target than reduce the likelihood of neurons responding and the associated response. It apparently exaggerates the target's value along the affected perceptual dimension. A context that concentrates energy at high frequencies makes an ambiguous target sound lower and does not just reduce the likelihood of the response associated with the low endpoint of a continuum. If all that the [l] and [r] contexts did was shift the criterion for categorizing the stop as [g] or [d], the [lg–rd] HL–LH pairs would have been no more discriminable than the [ld–rg] HH–LL pairs across the stimulus space in Experiment 2. These findings replicate Stephens and Holt's (2003) findings that contrasting speech contexts increased sensitivity for both speech and nonspeech targets. They extend them by showing that liquids that differ subcategorically can exaggerate the perceived values of following stops' spectra and that the nonspeech contexts used in Experiment 3 can do so, too.

Are context effects exclusively auditory?

Although listeners' responses can be biased by nonauditory context properties—for example, knowledge of lexicality (Ganong 1980; Pitt and Samuel 1993), transitional probabilities (Magnuson et al. 2003; McQueen 2003; McQueen et al. 2009; Pitt and McQueen 1998; Samuel and Pitt 2003), phonotactics (Berent et al. 2009; Berent et al. 2007; Breen et al. 2013; Coetzee 2005; Dehaene-Lambertz et al. 2000; Dupoux et al. 1999; Dupoux et al. 2001; Hallé and Best 2007; Massaro and Cohen 1983; Moreton 1999, 2002; Moreton and Amano 1999), and phonological processes (Darcy et al. 2007; Darcy et al. 2009; Gaskell 2001; Gaskell and Marslen-Wilson 1996, 1998; Gow 2001, 2002, 2003; Gow and Im 2004; Mitterer et al. 2006a; Mitterer et al. 2006b)—such knowledge-driven context effects are tangential to this article's focus.

Of more immediate relevance is Smits's (2001a) finding that Dutch listeners are more likely to identify a sibilant fricative as "s" than as "ʃ" before the close front rounded vowel [y] than its unrounded counterpart [i] and that the lowness of the vowel's *F*3, an acoustic correlate of rounding, and the categorization of the vowel as "y" influenced the fricative judgment independently (see also Smits 2001b). The greater likelihood of an "s" response before a

vowel with a lower *F*3 could be interpreted as an effect of auditory contrast: The energy concentration in the fricative is perceived as higher and more like [s] before a vowel with a lower *F*3. However, the independent effect of the vowel's categorization cannot be a product of contrast. Smits (2001a) proposes, instead, that listeners respond "s" more often when they categorize the vowel as "y" because they have learned that the energy concentration in a fricative is lowered when it is coarticulated with a rounded vowel.

Mitterer (2006) suggests that the vowel's *F*3 may affect the fricative response only indirectly through its effect on the vowel's categorization—that is, by what he refers to as "phonological mediation." In his first experiment, Mitterer used a seven by seven, [s-ʃ] by [y–i] stimulus array and replicated Smits's (2001a) findings that both the lowness of the following vowel's *F*3 and its categorization as "y" increased the likelihood of an "s" response to the fricative. He also found evidence suggestive of phonological mediation: Listeners were more likely to respond "s" before a vowel with a lower *F*3 when they were also more likely to categorize that vowel as "y."

Mitterer's (2006) next two experiments distinguished this interpretation from one in which the vowel's *F*3 influenced the fricative response directly via auditory contrast. In the first, the vowels following the fricatives were replaced with pure tones whose frequencies matched the vowels' *F*3s. If the fricative directly contrasted with its following context, listeners should have responded "s" more often as the frequency of the tone got lower. Their responses, instead, did not differ significantly as a function of the tones' frequencies. In the second, a subset of the original fricative by vowel stimulus array consisting of all seven fricatives and intermediate steps four–six along the [y–i] continuum was combined with video recordings of a Dutch speaker pronouncing the syllables [sy, ʃy, si, ʃi]. When the visual vowel was rounded [y], "s" responses were significantly more frequent overall than when it was unrounded [i]. Also, "s" responses were significantly more frequent overall when the auditory vowel's *F*3 was lower. The results of these last two experiments suggest that the effect of the vowel's acoustics in Mitterer's first experiment and that reported by Smits (2001a) are indirect and mediated by their influence on the vowel's categorization as "y" versus "i."

Even though these results do not support an auditory contrast account, as Mitterer (2006) himself observes, they also do not distinguish between an account like Smits's (2001a), where listeners have learned that fricatives' energy concentrations are lower next to rounded vowels, and one like Fowler's (2006), where they instead parse the signal's acoustic properties into the articulations that produced them. As Mitterer also observed, they do not rule out auditory contrast accounts of other context effects, like the one that has been this article's cynosure.

Our purpose has not been to show that auditory contrast is responsible for all context effects but, instead, that it is the only mechanism that could exaggerate the perceived difference between two intervals when an acoustic property changes between them. Such exaggeration predicts the common finding that a target sound is perceived as different from its context, but the contrast account has nothing to say about context effects that arise from combining visual with auditory information, from the context's categorization, or from the listeners' knowledge. It should not be a surprise that context effects in speech perception arise from multiple sources. After all, speech sounds are realized as articulations, then as acoustic properties, and finally as auditory qualities in the speech chain; they are constituents of lexical entries; they are manipulated and regulated by phonological grammars; the statistics of their (co)-occurrence vary considerably; and finally, their realizations vary enormously with the contexts, broadly construed, in which they occur. The contrast account is not undermined by the number and variety of context effects, but only by failures of its own predictions, such as failures of nonspeech contexts to alter percepts in the same way as acoustically similar speech contexts.

Are the perceptual adjustments for context learned?

Smits (2001a, 2001b) interpreted his finding that Dutch listeners respond "s" more often before a vowel that they categorize as "y" than before one that they categorize as "i" as evidence that they had learned that the energy concentration in a sibilant fricative is lower before a rounded than before an unrounded vowel. Mitterer's (2006) description of that finding and his own as "phonological mediation" does not conflict with this interpretation. Cross-linguistic comparisons of listeners' perceptual adjustments for systematic variation of target sounds with their contexts provide further evidence that these adjustments may be learned: These adjustments differ between listeners as a direct function of how much the target sound's articulation and, thus, its acoustics are *typically* altered by its context in their native language.

For example, in an acoustic study of trans-consonantal vowel-to-vowel coarticulation, Beddor et al. (2002) showed that speakers of both Shona and English anticipated the articulation of the following vowel in the current one, but only speakers of English also carried over the articulation of the preceding vowel well into the current one. Unstressed vowels coarticulated more than stressed vowels in English, but not Shona, even though the unstressed vowels were unreduced in English. The compensatory perceptual adjustments of speakers of these languages corresponded to a large extent with the production data, but not completely. In a discrimination task, both Shona and English listeners compensated for anticipatory coarticulation, but Shona listeners did not compensate any less than English listeners for carryover articulation, apparently because the Shona listeners performed poorly overall in this condition. In a categorization task, Shona and

English listeners again compensated equally for anticipatory coarticulation, but the English listeners compensated very little for carryover coarticulation in categorizing an [a–e] continuum, as compared with the Shona listeners. Beddor et al. attributed this anomaly to the English listeners' hearing intermediate steps along this continuum as [ə] and, thus, more like [a] than [e]. When the English listeners categorized an [o–e] continuum instead, they compensated as predicted for carryover as well as anticipatory coarticulation. The extent of compensation in the categorization tasks by listeners from both languages also closely matched the sizes of the coarticulatory effects observed in the production study.

Although some uncertainties linger, these results confirm the prediction that listeners will compensate more for familiar than for unfamiliar patterns of coarticulation and, more generally, that listeners learn what perceptual adjustments to make for the contexts in which sounds occur from their experience of how much those sounds are typically altered by coarticulation with those contexts. Even so, we may ask whether what they have learned is to *compensate* more when coarticulation is more extensive. Alternatively, they could have learned that the target sound contrasts more with its context when coarticulation is more extensive. For example, if the acoustic influence of a preceding [i] extends further into the vowel in the next syllable and, thus, raises *F*2 more at the beginning of that vowel, as it does in English, then this context would be a more extreme and a nearer source of auditory contrast with the following vowel than it would in Shona, where carryover coarticulation does not extend so far. Effects of learning are, in any case, unexpected only if no further perceptual adjustments are made for the target's context after or in addition to those auditory contrast is responsible for.

Yet other context effects

Fowler (2006) argued that compensation for coarticulation is a better account than auditory contrast of the perceptual adjustments listeners make for context, because it explains a larger variety of cases. One of these is the finding reported by Silverman (1987) that listeners judge a given *F*0 peak as less prominent when it occurs on a close than when it occurs on an open vowel—for example, on the [i] in *feasting* rather than the [æ] in *fasting*. Like Silverman, Fowler (2006) attributed this effect to listeners' compensating for the intrinsically higher *F*0 values of close, as compared with open, vowels. The important difference between this example and all the others considered here is that the context, the closeness of the vowel and its intrinsic *F*0, occurs in the same interval as the target, the intonationally determined height of the *F*0 peak on the vowel, rather than being adjacent to it. Because the vowel's closeness and the intonation affect the same acoustic property within the same interval, there is no way in which they can contrast. However, their articulatory sources can be separated: Raising the tongue body to produce a close vowel by contracting the posterior genioglossus also pulls the hyoid bone forward and, thereby, tilts the thyroid cartilage forward and stretches the vocal folds, while raising *F*0 to produce an *F*0 peak is achieved by contracting the cricothyroid, which also stretches the vocal folds by tilting the thyroid cartilage forward. If listeners perceive the articulations that determine the signal's acoustic properties, they can parse the two ways in which the thyroid cartilage was tilted forward and, thus, their independent effects on *F*0 and appropriately attribute some of the *F*0 peak's height to the vowel's closeness, rather than the intonation.

We agree that this perceptual adjustment cannot be attributed to auditory contrast, but we dispute attributing the *F*0 difference between close and open vowels to a mechanical perturbation of the hyoid bone and thyroid cartilage by the posterior genioglossus contraction that raises the tongue body. This argument is developed in detail in Kingston (1991, 1992) and is not repeated here (but see also Whalen, Gick, Kumada, and Honda 1999, for objections to Kingston's arguments). Instead, we sketch an alternative in which *F*0 is deliberately raised in close vowels to bring the first harmonic close to their low *F*1 and narrow the lowest band of energy in the vowel's spectrum. This narrowing enhances the distinction between a close vowel and a more open one, where a lower *F*0 and and a higher *F*1 create a broader lowest frequency energy band (for evidence that intrinsic *F*0 differences may enhance the contrast between close and more open vowels in this way, see Diehl and Kluender 1989; Fahey et al. 1996; Hoemeke and Diehl 1994; Syrdal and Gopal 1986). If listeners treat the raising of *F*0 in a close vowel as contributing to the percept of its closeness, they might discount that contribution from their judgment of how prominent the *F*0 peak is on that vowel, producing the lower judgment reported by Silverman (1987). This parses the vowel's *F*0 value not into the contributions of different articulations but, instead, into its strictly acoustic contributions to conveying different parts of the linguistic message, the vowel's closeness and prominence.

Listeners parse *F*0 in this account, too, but quite differently than they do in the compensation account. Instead of perceiving a close vowel's intrinsically higher *F*0 as an unintended, mechanical by-product of tongue body raising that must be discounted before the listener can properly judge the height of the *F*0 peak on the vowel, listeners instead interpret some of that peak's height as positive information about the vowel's closeness and use only the remainder to judge the peak's prominence. And instead of parsing *F*0 into the articulations that produce it, listeners parse it into the different parts of the message's phonic

Table 9 Individual *F*2–*F*4 offset frequencies and average frequencies of adjacent formants

Language	Context	*F*2	*F*3	*F*4	μ(*F*2,*F*3)	μ(*F*2,*F*3)	μ(*F*2, *F*3,*F*4)
English	l	1060	2600	3600	1830	3100	2420
	ɻ	1350	1800	3050	1575	2425	2067
Tamil	r	1440	2010	3610	1725	2810	2353
	l	1600	1780	3100	1690	2400	2160

content that it provides information about. Finally, like the auditory contrast account of how listeners adjust their percepts for adjacent targets and contexts, this alternative refers only to the acoustic properties of the signal and their proximal auditory correlates, and not to the distal articulations that might have produced them.

Responding to further challenges to the contrast account

That listeners perceive a target as contrasting with its context, rather than compensating for its coarticulation with that context, appears to be supported by Lotto and Kluender's (1998) finding that frequency-modulated (FM) and pure tones that mimic the *F*2, *F*3, and *F*4 differences between [al] and [ar] shift the likelihood of a "g" response in the same direction as the spoken contexts do. Neither FM nor pure tones would be perceived as a source of coarticulatory perturbations of the stop's place that listeners would be expected to compensate for. Recently, Viswanathan, Fowler, and Magnuson (2009) showed that Lotto and Kluender's results may have depended on the FM and pure tones having the smallest possible bandwidths and the same total RMS amplitudes as the original syllables. They thus concentrated considerable energy into the narrowest possible frequency range. When Viswanathan et al. (2009) manipulated the intensity, trajectory, and bandwidth of nonspeech analogues of [al] and [ar] contexts so as to match those of *F*3 alone in the original syllables more and more closely, the differences in the likelihood of "g" response shrank and eventually disappeared. Because the *F*3 differences between [al] and [ar] are sufficient to shift "g" response likelihoods when they are heard in the context of the remainder of these syllables, despite their low intensities, time-varying trajectories, and broad bandwidths, Viswanathan et al. (2009) argued that contrast cannot be the responsible mechanism in these speech contexts. Following Fowler et al. (2000), they argued that the speech targets may be masked by the intense, minimal bandwidth nonspeech contexts used by Lotto and Kluender. Viswanathan's (2009) finding that shifts in response likelihoods shrink as the frequency difference between nonspeech pure tone contexts and the stops' *F*3 increase also appears to support a masking account of shifts in such nonspeech contexts. We challenge these attributions of Lotto and Kluender's nonspeech results to masking in experiments to be reported elsewhere, by showing the predicted effects of masking do not emerge.

Viswanathan, Fowler, and Magnuson (2010) challenged the contrast account in another way, by showing that the Tamil alveolar trill [r] increased the likelihood of "g" responses just as much as the American English alveolar [l] and that the Tamil retroflex lateral [ɭ] decreased them just as much as the American English retroflex approximant [ɻ] (which is represented as "ɹ" in Viswanathan et al. 2010). This grouping of the liquid contexts into these two pairs is expected if their perceptual effects are determined by their places of articulation, the more anterior alveolars and the more posterior retroflexes, respectively, but is unexpected if those effects are determined by their acoustics, because *F*3 in Tamil [r] is nearly as low as *F*3 in English [ɻ]. The lower likelihood of "g" responses after the Tamil retroflex lateral [ɭ] could either be attributed to its place of articulation or its low *F*3. When the spoken contexts were replaced by pure tones at their *F*3 offset frequencies and with RMS amplitudes equaling the original syllables, "g" response likelihoods were greater after the highest of these pure tones, the one mimicking American English [l], than after any of the lower pure tones, after which "g" response likelihoods did not differ. Essentially, the same results were obtained when the single pure tones were replaced with a pair of pure tones at the *F*2 and *F*3 offset frequencies of the four liquids, and when a third pure tone at the liquids' *F*4 offset frequencies was added, "g" response likelihoods were equally low after the nonspeech analogues of both Tamil liquids, noticeably higher after the analogue of American English [ɻ], and highest after the analogue of American English [l].

While these findings challenge any account that attributes the perceptual effects of a context to the locations of peaks in its spectrum, the conclusions that Viswanathan et al. (2010) drew from them overlook how Lotto and Kluender (1998) chose the frequencies of the pure tones for nonspeech stimuli. The lower tone was the average frequency of the *F*2 and *F*3 offset frequencies of [ar], and the higher tone was the average frequency of the *F*3 and *F*4 offset frequencies of [al]. The tones thus do not reflect any single formant's offset frequency or any single peak in the original syllable's spectra but, instead, a coarser concentration of energy, at lower frequencies at the end of [ar] and at higher frequencies at the end of [al]. Calculating similar averages from the *F*2–*F*4 offset frequencies reported for the four liquids by Viswanathan et al. (2010), as in Table 9, suggests an alternative, auditory explanation for their findings in the original speech contexts.

Liquids shift "g" response likelihoods similarly when the average offset frequencies of adjacent formants are similarly low or high. This explanation is more plausible if the average offset frequencies that matter are those between *F*3 and *F*4 or, even more simply, the average offset frequencies of all three formants. The average offset frequencies of *F*2 and *F*3 are in

the right low-to-high order, but they do not group into low and high pairs as tightly as these other averages do. Undoubtedly, the contribution of each formant to these averages should be weighted by its intensity, but Viswanathan et al. (2010) did not report the formants' intensities, and such adjustments cannot, therefore, be made here. Accounting for the divergent perceptual effects of the pairs and triplets of pure tones also requires that their intensities be measured (or manipulated). The perceptual effects of these four liquids therefore do not uniquely select compensation for coarticulation over contrast as the responsible mechanism.

Concluding remarks

By supporting the contrast over compensation accounts of context effects in speech perception, the results of these experiments show that the objects of speech perception are auditory qualities, rather than articulatory gestures. They thus challenge both the motor and direct-realist theories of speech perception. They also show that a speech sound's context can exaggerate its perceived value for some acoustic property, and not merely change the likelihood that it will be assigned to one category rather than another. They thus raise the possibility that the acoustic signal produced by the speaker's articulations may be transformed by auditory processing before any linguistic value is assigned to it.

Acknowledgment This research was supported by grant R01 DC006241 from NIDCD to the first author, which is gratefully acknowledged.

Appendix

Other stimulus parameters

The KLSYN88 stimulus parameters that did not vary in the stimuli are listed in Table 10 below (Klatt and Klatt 1990). In addition to these parameters, the values of A2F, A3F, A4F, A5F, and A6F, the amplitudes of the frication-excited second–sixth formants, were set at 36, 42, 48, 54, and 60 dB in the [d] endpoint and at 60, 54, 48, 42, and 36 dB in the [g] endpoint

Table 10 Fixed synthesis parameters: times (in milliseconds) and values for the synthesis parameters that did not vary between stimuli

Segment	Time	AV	AH	AF	*F*0	*F*1	B1	B2, B3	*F*4	B4	*F*5	B5	*F*6	B6	BNP, BNZ	BTP, BTZ
a	0	0	0	0	155	800	1000	1000	3200	1000	4500	1000	4900	1000	1000	1000
	40	62			150		125	100		125		350		250		
	210					800			3200						1000	1000
Liquid	270					325			3400						200	250
	350	62			130	325	125	100	3400	125		350		250	200	250
Stop	360	59					1000	1000		1000		1000		1000	1000	1000
	435	51														
	455	39	0	0	115	180	1000	1000	3500	1000		1000		1000		
Burst	460		54	45												
	470			39			200	200		750		750		750		
a	480	62	48	0	145											
	485						125	100		125		350		250		
	490		0													
	515								3200							
	525					800										
	680	62			130											
	705	61														
	725											350		250		
	735									125						
	745						125	100								
	760	58			110		1000	1000		1500		2000		2000		
	770	0														
	800	0	0	0	90	800	1000	1000	3200	1500	4500	2000	4900	2000	1000	1000

Note. AV, AH, AF=amplitudes of the voice, aspiration, and frication sources in dB; *F*0=fundamental frequency in Hz: *F*1, *F*4–F6=first, fourth–sixth formant frequencies in Hz; B1–B6=bandwidths of the first–sixth formants; BNZ, BNP, BTP, BTZ=bandwidths of the "nasal" and "tracheal" poles and zeros in Hz. These parameters were used along with their corresponding frequencies, FNP, FNZ, FTP, FTZ, to introduce additional poles and zeros in the synthesis of the liquids (see the text for further details). Each parameter's values were interpolated between the times when a value is listed for it in the table. The sections in the table correspond roughly to the acoustic intervals of the segments listed in the first column.

during the 585- to 595-ms interval. These parameters ramped up to these values from 0 dB at 580 ms and then back down to 0dB at 605 ms. As with all other parameters that varied along the stop or liquid continua, their values were interpolated linearly for each of the five steps between the endpoints.

References

Baayen, R. H. (2008). *Analyzing linguistic data.* Cambridge, UK: Cambridge University Press.

Baayen, R. H. (2011). languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics". R package version 1.2.

Barr, D., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure in mixed-effects models: Keep it maximal. *Journal of Memory and Language, 68,* 255–278.

Bates, D., Maechler, M., and Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375–39.

Beddor, P. S., Harnsberger, J. D., & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates. *Journal of Phonetics, 30,* 591–627.

Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition, 104,* 591–630.

Berent, I., Lennertz, T., Smolensky, P., & Vaknin-Nusbaum, V. (2009). Listeners' knowledge of phonological universals: Evidence from nasal clusters. *Phonology, 26,* 75–108.

Breen, M., Kingston, J., & Sanders, L. D. (2013). Perceptual representations of phonotactically illegal syllables. *Attention, Perception, & Psychophysics, 75,* 101–120.

Coetzee, A. (2005). The OCP in the perception of English. In S. Frota, M. Vigario, & M. J. Freitas (Eds.), *Prosodies. Selected Papers from the Phonetics and Phonology in Iberia Conference, 2003* (pp. 223–245). Berlin: Mouton de Gruyter.

Darcy, I., Peperkamp, S., & Dupoux, E. (2007). Bilinguals play by the rules. Perceptual compensation for assimilation in late L2-learners. In J. Cole & J. I. Hualde (Eds.), *Laboratory Phonology 9* (pp. 411–442). Berlin: Mouton De Gruyter.

Darcy, I., Ramus, F., Christophe, A., Kinzler, K., & Dupoux, E. (2009). Phonological knowledge in compensation for native and non-native assimilation. In F. Kügler, C. Féry, & R. van de Vijver (Eds.), *Variation and Gradience in Phonetics and Phonology* (pp. 265–309). Berlin: Mouton De Gruyter.

Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlates of phonological processing: A cross linguistic study. *Journal of Cognitive Neuroscience, 12,* 635–647.

Development Core Team, R. (2011). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology, 1,* 121–144.

Diehl, R. L., & Walsh, M. A. (1989). An auditory basis for the stimulus-length effect in the perception of stops and glides. *Journal of Acoustical Society of America, 85,* 2154–2164.

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology, 55,* 149–179.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1568–1578.

Dupoux, E., Pallier, C., Kakehi, K., & Mehler, J. (2001). New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes, 16,* 491–505.

Fahey, R. P., Diehl, R. L., & Traunmüller, H. (1996). Perception of back vowels: Effects of varying f1-f0 distance. *Journal of the Acoustical Society of America, 99,* 2350–2357.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct realist perspective. *Journal of Phonetics, 14,* 3–28.

Fowler, C. A. (1990). Sound-producing sources as the objects of perception: Rate normalization and nonspeech perception. *Journal of the Acoustical Society of America, 88,* 1236–1249.

Fowler, C. A. (1994). Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation. *Perception & Psychophysics, 55,* 597–610.

Fowler, C. A. (1996). Listeners do hear sounds not tongues. *Journal of Acoustical Society of America, 99,* 1730–1741.

Fowler, C. A. (2005). Parsing coarticulated speech in perception: Effects of coarticulation resistance. *Journal of Phonetics, 33,* 199–213.

Fowler, C. A. (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception and Psychophysics, 68,* 161–177.

Fowler, C. A., & Brown, J. M. (1997). Intrinsic *f*0 differences in spoken and sung vowels and their perception by listeners. *Perception and Psychophysics, 57,* 729–738.

Fowler, C. A., & Smith, M. (1986). Speech perception as vector analysis: An approach to the problems of segmentation and invariance. In J. Perkell & D. H. Klatt (Eds.), *Invariance and Variability of Speech Processes* (pp. 123–136). Hillsdale: Lawrence Erlbaum Associates.

Fowler, C. A., Brown, J., & Mann, V. (2000). Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 877–888.

Ganong, W. F., III. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 110–125.

Gaskell, G. (2001). Phonological variation and its consequences for the word recognition system. *Language and Cognitive Processes, 16,* 723–729.

Gaskell, G., & Marslen-Wilson, W. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 144–158.

Gaskell, G., & Marslen-Wilson, W. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 380–396.

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics, 66,* 363–376.

Gow, D. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language, 45,* 133–159.

Gow, D. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception & Performance, 28,* 163–179.

Gow, D. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics, 65,* 575–590.

Gow, D., & Im, A. M. (2004). A cross-language examination of assimilation context effects. *Journal of Memory and Language, 51,* 279–296.

Hallé, P. A., & Best, C. T. (2007). Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop+/l/ clusters. *Journal of the Acoustical Society of America, 121,* 2899–2914.

Hoemeke, K. A., & Diehl, R. L. (1994). Perception of vowel height: The role of f1-f0 distance. *Journal of the Acoustical Society of America, 96,* 661–674.

Holt, L. L. (2005). Temporally non-adjacent non-linguistic sounds affect speech categorization. *Psychological Science, 16,* 305–312.

Holt, L. L. (2006). The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization. *Journal of Acoustical Society of America, 120,* 2801–2817.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434–446.

Kingston, J. (1991). Integrating articulations in the perception of vowel height. *Phonetica, 48,* 149–179.

Kingston, J. (1992). The phonetics and phonology of perceptually motivated articulatory coordination. *Language and Speech, 35,* 99–113.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America, 87,* 820–857.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1–36.

Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science, 243,* 489–494.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431–461.

Lotto, A. J., & Holt, L. L. (2006). Putting phonetic context effects into context: A commentary on Fowler (2006). *Perception and Psychophysics, 68,* 178–183.

Lotto, A. J., & Kluender, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics, 60,* 602–619.

Macmillan, N. A., & Creelman, C. D. (1996). Triangles in ROC space: History and theory of "nonparametric" measures of sensitivity and response bias. *Psychnomic Bulletin & Review, 3,* 164–170.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide* (2nd ed.). Mahwah: Lawrence Erlbaum Associates Publishers.

Macmillan, N. A., Goldberg, R. F., & Braida, L. D. (1988). Resolution for speech sounds: Basic sensitivity and context memory on vowel and consonant continua. *Journal of the Acoustical Society of America, 84,* 1262–1280.

Magnuson, J. S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003). Lexical effects on compensation for coarticulation: the ghost of christmash past. *Cognitive Science, 27,* 285–298.

Mann, V. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics, 28,* 407–412.

Massaro, D. W., & Cohen, M. M. (1983). Phonological context in speech perception. *Perception & Psychophysics, 34,* 338–348.

McQueen, J. M. (2003). The ghost of Christmas future: didn't scrooge learn to be good?: Commentary on Magnuson, McMurray, Tanenhaus, and Aslin (2003). *Cognitive Science, 27,* 795–799.

McQueen, J. M., Jesse, A., & Norris, D. (2009). No lexical-prelexical feedback during speech perception or: Is it time to stop playing those Christmas tapes? *Journal of Memory and Language, 61,* 1–18.

Micheyl, C., & Messing, D. P. (2006). Likelihood ratio, optimal decision rules, and correct response probabilities in a signal detection theoretic, equal-variance gaussian model of the observer in the 4IAX paradigm. *Perception & Psychophysics, 68,* 725–735.

Micheyl, C., & Oxenham, A. J. (2005). Comparing *F0* discrimination in sequential and simultaneous conditions. *Journal of the Acoustical Society of America, 118,* 41–44.

Mitterer, H. (2006). On the causes of compensation for coarticulation: Evidence for phonological mediation. *Perception & Psychophysics, 68,* 1227–1240.

Mitterer, H., Csépe, V., Honbolygo, F., & Blomert, L. (2006a). The recognition of phonologically assimilated words does not depend on specific language experience. *Cognitive Science, 30,* 451–479.

Mitterer, H., Csépe, V., & Blomert, L. (2006b). The role of perceptual integration in the recognition of assimilated word forms. *The Quarterly Journal of Psychology, 59,* 1395–1424.

Moreton, E. (1999). Evidence for phonological grammar in speech perception. In J.J. Ohala, Y.Hasegawa, M.Ohala, D.Granville, and A. C. Bailey, editors, *Proceedings of the 14th International Congress of Phonetic Sciences* , pages 2215–2217. San Francisco.

Moreton, E. (2002). Structural constraints in the perception of English stop-sonorant clusters. *Cognition, 84,* 55–71.

Moreton, E. and Amano, S. (1999). Phonotactics in the perception of Japanese vowel length: Evidence for long distance dependencies. Proceedings of the 6th European Conference on Speech Communication and Technology.

Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: A problem of significance. *Nature Neuroscience, 14,* 1105–1107.

Norris, D. (1995). Signal detection theory and modularity: On being sensitive to the power of bias models of semantic priming. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 935–939.

Pardo, J. S., & Fowler, C. A. (1997). Perceiving the causes of coarticulatory acoustic variation: Consonant voicing and vowel pitch. *Perception & Psychophysics, 59,* 1141–1152.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics, 13,* 253–260.

Pisoni, D. B., Carrell, T. D., & Gans, S. J. (1983). Perception of rapid spectrum changes in speech and nonspeech signals. *Perception & Psychophysics, 34,* 314–322.

Pitt, M. A., & McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language, 39,* 347–370.

Pitt, M. A., & Samuel, A. G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 699–725.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin, 92*, 81–110.

Samuel, A., & Pitt, M. A. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language, 48,* 416–434.

Silverman, K. E.A. (1987). *The Structure and Processing of Fundamental Frequency Contours* . Ph.D. dissertation, Cambridge University

Smits, R. (2001a). Evidence for hierarchical organization of coarticulated phonemes. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 1145–1162.

Smits, R. (2001b). Hierarchical organization of coarticulated phonemes: A theoretical analysis. *Perception & Psychophysics, 63,* 1109–1139.

Stephens, J., & Holt, L. (2003). Preceding phonetic context affects perception of non-speech sounds. *Journal of Acoustical Society of America, 114,* 3036–3039.

Stevens, K. (1998). *Acoustic Phonetics*. Cambridge, MA: MIT Press.

Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America, 79*(4), 1086–1100.

Ulanovsky, N., Las, L., & Nelken, I. (2003). Processing of low probability sounds by cortical neurons. *Nature Neuroscience, 6,* 391–398.

Ulanovsky, N., Las, L., Farkas, D., & Nelken, I. (2004). Multiple time scales of adaptation in auditory cortex neurons. *Journal of Neuroscience, 24,* 10440–10453.

Viswanathan, N. (2009). *Perceptual compensation for coarticulation: Implications for theories of speech perception. Talk delivered to the Phonology Group, Linguistics Department.* Amherst: University of Massachusetts.

Viswanathan, N., Fowler, C. A., & Magnuson, J. S. (2009). A critical examination of the spectral contrast account of compensation for coarticulation. *Psychonomic Bulletin and Review, 16,* 74–79.

Viswanathan, N., Fowler, C. A., & Magnuson, J. S. (2010). Compensation for coarticulation: Disentangling auditory and gestural theories of perception of coarticulatory effects in speech. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 1005–1015.

Wade, T., & Holt, L. (2005). Effects of later occurring nonlinguistic sounds on speech categorization. *Journal of Acoustical Society of America, 118,* 1701–1710.

Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception & Psychophysics, 35,* 49–64.

Whalen, D. H., Gick, B., Kumada, M., & Honda, K. (1999). Cricothyroid activity in high and low vowels: Exploring the automaticity of intrinsic F0. *Journal of Phonetics, 27,* 125–142.