

Prior experience with negative spectral correlations promotes information integration during auditory category learning

Mathias Scharinger · Molly J. Henry · Jonas Obleser

© Psychonomic Society, Inc. 2013

Abstract Complex sounds vary along a number of acoustic dimensions. These dimensions may exhibit correlations that are familiar to listeners due to their frequent occurrence in natural sounds—namely, speech. However, the precise mechanisms that enable the integration of these dimensions are not well understood. In this study, we examined the categorization of novel auditory stimuli that differed in the correlations of their acoustic dimensions, using *decision bound theory*. Decision bound theory assumes that stimuli are categorized on the basis of either a single dimension (*rule based*) or the combination of more than one dimension (*information integration*) and provides tools for assessing successful integration across multiple acoustic dimensions. In two experiments, we manipulated the stimulus distributions such that in Experiment 1, optimal categorization could be accomplished by either a rule-based or an information integration strategy, while in Experiment 2, optimal categorization was possible only by using an information integration strategy. In both experiments, the pattern of results demonstrated that unidimensional strategies were strongly preferred. Listeners focused on the acoustic dimension most closely related to pitch, suggesting that pitch-based categorization was given preference over timbre-based categorization. Importantly, in Experiment 2, listeners also relied on a two-dimensional information integration strategy, if there was immediate feedback. Furthermore, this strategy was used more often for distributions defined by a negative spectral correlation between stimulus dimensions, as compared with distributions with a positive correlation. These results suggest that prior experience with such correlations might shape short-term auditory category learning.

Keywords Audition · Categorization · Implicit learning · Implicit memory · Perception

Introduction

Categorization of auditory sensory information is vital for making rapid decisions in an acoustic environment. For instance, the ability to correctly map a complex harmonic tone to the category *car horn* is advantageous when one is crossing a road. Recent decades have brought about a considerable body of research on how categories are formed and maintained (Ashby & Waldron, 1999; McQueen, 1996; Nosofsky, 1988; Rosch, 1973, 1978; Russ, Lee, & Cohen, 2007; Sloutsky, 2003; Spiering & Ashby, 2008; Verbeemen, Vanpaemel, Pattyn, Storms, & Verguts, 2007; Yamauchi, Love, & Markman, 2002; for reviews, see Ashby & Maddox, 2005, 2011). This work has become increasingly focused on auditory categories (Goudbeek, Cutler, & Smits, 2008; Guenther & Bohland, 2002; Guenther, Nieto-Castanon, Ghosh, & Tourville, 2004; Holt & Lotto, 2006; Mirman, Holt, & McClelland, 2004).

There is a consensus that auditory categorization involves the utilization and integration of different acoustic dimensions (e.g., spectrum [pitch, timbre], duration; Goudbeek, Swingley, & Smits, 2009; Holt & Lotto, 2006), but it is less clear how integration of information from integral (nonseparable; Goudbeek et al., 2009) dimensions might differ from integration of information from nonintegral (separable) dimensions. Furthermore, integration across dimensions might be influenced by prior knowledge of specific correlations between dimensions, particularly from speech. The purpose of this study was to assess the degree to which integration of information from two dimensions is influenced by (1) the integrality of acoustic dimensions (here, location of spectral peaks in frequency space), (2) the correlation between these dimensions (positive vs. negative correlation of first [S1] and second

M. Scharinger (✉) · M. J. Henry · J. Obleser
Max Planck Research Group “Auditory Cognition”,
Max Planck Institute for Human Cognitive and Brain Sciences,
Stephanstrasse 1A,
04103 Leipzig, Germany
e-mail: mscharinger@cbs.mpg.de

[S2] spectral peak), and (3) the presence or absence of immediate corrective feedback.

To that end, we focused on two modeling approaches that are particularly apt to these purposes: *decision bound theory* and *logistic regression*. Logistic regression has previously been applied to auditory categorization and can be used to predict category membership decisions on the basis of stimulus properties (Hosmer & Lemeshow, 2000). Values along a stimulus dimension are entered into the regressions as continuous variables, and the β -weight for this regressor reflects the predictive power of the stimulus dimension with respect to categorization responses. Comparisons of β -weights for different stimulus dimensions allow assessment of the degree to which individual stimulus properties are used for categorization. For instance, Goudbeek et al. (2009) assessed strategy use in two conditions where either frequency or duration was the relevant dimension. Responses were predicted from frequency and duration with logistic regressions that yielded β -weights for either dimension. A higher β -weight for frequency, as compared with duration, in the condition where frequency was the relevant dimension showed that participants indeed used this dimension for their response.

On the other hand, decision bound theory (Ashby & Gott, 1988) assumes that category acquisition involves learning to divide perceptual space, corresponding to an internal representation of stimulus space, into response regions according to a linear or nonlinear boundary (Ashby & Waldron, 1999). The position of a novel stimulus in perceptual space is compared with the location of the boundary, and the corresponding response is assigned. Thus, from this perspective, category learning is a signal detection problem where the decision bound separating categories corresponds to the response criterion. Decision bound models for auditory categorization have been extensively studied in the visual domain and, thus, provide a framework from which we can generate specific predictions about strategy use in novel category learning.

Thus far, no study of which we are aware has combined these two approaches (i.e., logistic regression and decision bound models). In this regard, the present study complements and goes beyond previous research. In what follows, we will describe in detail two strategies for novel category learning that emerge from decision bound theory and that we assess in the context of auditory category learning in the present study.

Rule-based and information integration category learning

The distinction between rule-based and information integration category learning comes from a neuropsychological model called *competition between verbal and implicit systems* (COVIS; Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Waldron, 1999). The model assumes that learning involves two systems that compete or interact

(Ashby & Crossley, 2010) and differ in their functions and neural underpinnings. *Rule-based* learning involves categorization based on an explicit rule that is frequently relatively easy to verbalize (e.g., *if a tone is high in frequency, respond "category A"; otherwise, respond "category B"*). Generally, rule-based decision bounds are orthogonal to the dimension on which the decision should be based. Rule-based learning is assumed to predominantly depend on an explicit hypothesis-testing system (Maddox, Filoteo, Lauritzen, Connally, & Hejl, 2005), subserved by the dorsolateral prefrontal cortex, the anterior cingulate, and the caudate nucleus (Ashby & Ell, 2001; Rao et al., 1997).

On the other hand, *information integration* category learning tasks require a predecisional combination of information from more than one dimension. Usually, the optimal rule in information integration tasks is not easily verbalized (e.g. *if a tone is higher in frequency than it is long in duration, respond "A"; otherwise, respond "category B"*). Decision bounds are not orthogonal to the dimension on which the decision should be based but, rather, are represented as a diagonal—for example, in a two-dimensional stimulus space. Successful learning of an information integration task is proposed to rely on an implicit procedural learning system that depends on feedback processes (Ashby & Waldron, 1999; Maddox, Filoteo, Hejl, & Ing, 2004; Maddox et al., 2005). This system is claimed to be subserved by the body and tail of the caudate (Nomura et al., 2007; Seger & Cincotta, 2005). In several studies, it has been shown that information integration is indeed dependent on (immediate) feedback in categorization or discrimination tasks (Ashby, Queller, & Berretty, 1999; Ashby & Waldron, 1999).

In the present study, we applied decision bound modeling of rule-based and information integration category learning to an auditory categorization task. Decision bound theory provided us with an optimal tool with which to evaluate the success of information integration across two acoustic dimensions in making category membership decisions.

Auditory category learning

A number of studies have attempted to distinguish between rule-based and information integration strategies in auditory categorization performance (e.g., Goudbeek et al., 2008; Goudbeek et al., 2009; Holt & Lotto, 2006, 2008; Maddox, Ing, & Lauritzen, 2006; Mirman et al., 2004; Smits, Sereno, & Jongman, 2006), although relatively few of them directly applied decision bound models to describe learning. For instance, in studies by Goudbeek et al. (2009) and Smits et al., participants learned to categorize inharmonic complex tones that varied along the dimensions of duration and spectral filter location (analogous to the first formant frequency, $F1$, in speech). Goudbeek et al. (2009)

examined performance for category distributions that were best separated by a unidimensional duration-based boundary, a unidimensional frequency (i.e., spectral filter location) boundary, and a diagonal (information integration) boundary that required categorizing stimuli on the basis of a combination of duration and frequency values. The authors observed much poorer performance for the information integration condition, relative to either of the rule-based conditions. On this basis, they hypothesized that rule-based category learning may be the default strategy in audition (cf. Maddox et al., 2006).

However, it is also possible that rule-based learning may have been used more often due to use of duration and frequency dimensions, which have been suggested to be *separable*, rather than *integral* (Grau & Kemler Nelson, 1988; Silbert, Townsend, & Lentz, 2009). Prior research on acoustic-phonetic processing has suggested that dimensions that vary in the same domain (e.g., frequency) are more likely to be integral (Kingston, Diehl, Kirk, & Castleman, 2008; Kingston & Macmillan, 1995; Kingston, Macmillan, Dickey, Thorburn, & Bartels, 1997). Consistent with this suggestion, Maddox and colleagues (Maddox, Molis, & Diehl, 2002) showed that a model of categorization performance assuming information integration accounted well for performance when stimuli varied in their second and third resonance frequencies; however, they did not examine learning per se, since the categories were already highly learned.

Furthermore, previous research also stressed the importance of how acoustic dimensions are related to each other. In this respect, *negative correlations* between spectral filter dimensions are relevant with respect to *intrinsic pitch* of vowels, reflecting the impression that high vowels (with a low $F1$) have slightly higher pitch (higher f_0) than do low vowels (with a higher $F1$; Lehiste & Peterson, 1961), and *vowel nasalization*, showing that the more nasalized a vowel, the lower its $F1$ frequency (Diehl, Kluender, & Walsh, 1990). Given that these correlations hold cross-linguistically (likely to be based on articulatory constraints; Carre, 2009), listeners should be familiar with them and, correspondingly, benefit in novel categorization situations that employ such correlations.

Our two experiments sought to assess whether (1) information is integrated across two integral acoustic dimensions, (2) information integration depends on the correlation of the acoustic dimensions, and (3) information integration requires immediate feedback. For these reasons, both experiments used stimuli that differed in the location of two spectral peaks (analogous to the first formant frequencies of vowels, $F1$ and $F2$) and comprised a learning phase with immediate feedback as well as a maintenance phase without feedback. In Experiment 1, we examined distributions whose decision bounds would similarly allow for rule-based or information integration categorization, thereby assessing the natural

inclination for a particular strategy during the categorization of auditory stimuli with integral dimensions (Fig. 1a). In contrast, Experiment 2 used stimulus distributions that required information integration for optimal performance (Fig. 2a).

Our hypotheses are as follows:

1. On the basis of Maddox et al. (2006), we assume that rule-based categorization is the preferred strategy in audition. Therefore, in both experiments, we should see substantial evidence for rule-based behavior.
2. The long-term experience with negative correlations in speech (and corresponding decision bounds) should shape the short-term categorization of nonspeech stimuli. As a result, more ready use of an information integration strategy, and consequently, better category-learning performance should be observed for distributions with negative correlations between spectral peak frequencies. Since Experiment 2 was designed such that information integration would yield optimal performance (Fig. 1b), we expect differences between correlations to particularly manifest themselves in this experiment.
3. Finally, information integration seems to require immediate feedback (Ashby et al., 1999; Ashby & Waldron, 1999). We therefore expect more information integration in the learning than in the maintenance phase of our experiments.

Experiment 1

Experiment 1 extended the work of Goudbeek and colleagues (Goudbeek et al., 2008; Goudbeek et al., 2009) to stimuli varying along integral acoustic dimensions. Two distribution types were examined. In the *falling* condition, stimulus distributions were characterized by a *negative* correlation between spectral filter locations, while the *rising* condition distributions were characterized by a *positive* correlation. In both conditions in Experiment 1, rule-based and information integration strategy use would have yielded little performance difference so that we could assess the natural inclinations of participants.

Accuracy analyses were supplemented by fitting a number of decision bound models (Ashby, 1992) to individual participant data in order to assess strategy use during category learning. As was outlined above, we also calculated logistic regressions with the dependent variable *category A vs. B* for both learning and maintenance phases (Hilbe, 2009) in order to quantify the contributions of each spectral dimension to single-trial category membership decisions. On the basis of previous findings (little information integration use; Goudbeek et al., 2009; Maddox et al., 2006) and as

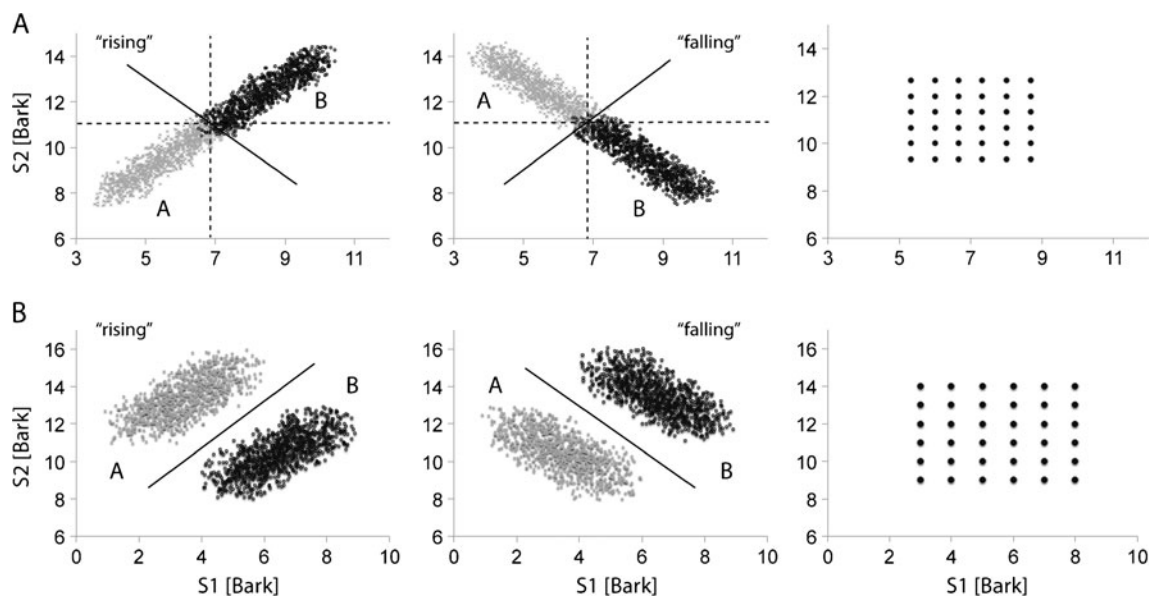


Fig. 1 **a** Stimulus distributions in the S1–S2 space for Experiment 1. The dotted lines indicate unidimensional decision bounds for either S1 (vertical) or S2 (horizontal). The diagonal solid line represents a decision bound for information integration. The “rising” distribution (left) is characterized by a positive correlation between S1 and S2, while the “falling” distribution (middle) yields a negative S1–S2 correlation. The right panel shows the equidistantly spaced grid for the

maintenance phase. **b** Stimulus distributions for Experiment 2. Information integration decision bounds are represented by the solid diagonal lines. Again, the rightmost panel illustrates the stimulus arrangement for the maintenance phase. Note that stimuli in the maintenance phase were matched to the corresponding distributions in the learning phase, such that there exist small differences between Experiments 1 and 2

a consequence of the stimulus materials in Experiment 1 (no strong bias for either strategy), we predicted a bias toward using rule-based strategies.

We did not expect differences between the rising and falling distributions. This is because observing a difference between rising and falling distributions should have depended on adoption of an information integration strategy, which we did not predict to observe in Experiment 1. For the same reason, we predicted that immediate feedback in Experiment 1 would play no or only a negligible role, since it is claimed to be important for information integration, but not rule-based learning (Ashby et al., 1999; Ashby & Waldron, 1999).

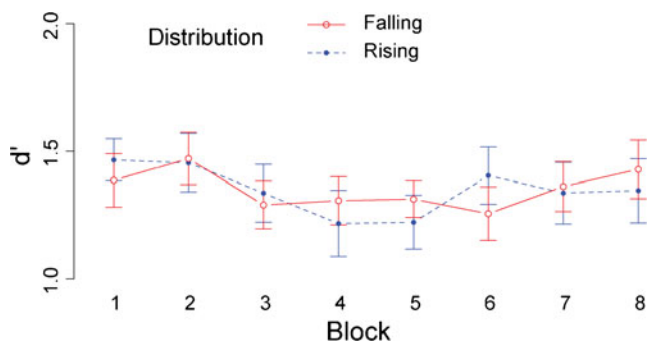


Fig. 2 Performance measure per block in the learning phase of Experiment 1. Error bars indicate across-subjects standard errors of the means. Note the absence of any effect of *falling* versus *rising* stimulus distribution on performance

Method

Materials

Stimuli were created with PRAAT (Boersma & Weenink, 2011) in two steps. First, a 90-ms white noise was generated; the duration was chosen on the basis of previous studies (Goudbeek et al., 2008; Goudbeek et al., 2009). Second, the white noise was filtered in two frequency bands that approximated the location of the first and second formant frequencies of naturally produced vowels—that is, F1 and F2, respectively. Target filter frequencies are referred to as spectral filter frequencies, S1 and S2, throughout this article, and are normalized to Bark (Zwicker, 1961). The Bark conversion is commonly applied in acoustic-phonetic research and accounts for the nonlinearity of the frequency resolution by the human auditory system.

The same original white noise token was used as the basis for all 1,000 stimuli that were generated for each category (A and B) and for each distribution (falling and rising), with different S1 and S2 filter frequencies in each case. Filter frequencies were drawn from the distributions shown in Fig. 1a. In order to arrive at the stretched distributions along the falling and rising diagonals in the S1/S2 space (with a slope of -1 and $+1$), the linear equation for the diagonal running through the distribution center was calculated. Then individual bivariate normal distributions were generated that had means, μ , at 40 (x,y) locations along the

diagonal and equal standard deviations, σ (see Table 1 for details). Twenty-five tokens per distribution were randomly generated, yielding a total of 1,000 stimuli per distribution.

Noises were filtered with fast IIR filters, comprising two recursive filter coefficients. Filter bandwidths were 0.2 times the target filter frequency. Stimuli were normalized to an equal average intensity that approximated 60 dB SPL (Boersma & Weenink, 2011). Onsets were multiplied with the first half period of a $[1 - \cos(x)] * 0.5$ function, and offsets with the first half period of a $[1 + \cos(x)] * 0.5$ function, over a duration of 10 ms in each case, in order to eliminate acoustic artifacts.

In order to arrive at a stimulus-based measure for the likelihood of strategy preference, we determined the normalized distance between the means of category A and category B distributions, here referred to as δ' , along both the S1 and S2 dimensions. The information integration δ' (Euclidean distance between distribution means) was of comparable size to the rule-based δ' values (difference between means for each dimension, S1 and S2). We expect that larger category distances would lead to better categorization performance, such that if participants were to utilize optimal strategies, they ought to prefer those for which δ' is largest. In Experiment 1, the similarity of the distances therefore suggested equal strategy preference.

Stimuli for the nonfeedback maintenance phase were arranged in an equidistantly spaced grid with step sizes of 2/3 Bark in either dimension (S1, 5–9 Bark; S2, 9–13 Bark). They thus described a 6×6 grid that evenly covered the critical region of the original stimulus space (Fig. 2a).

Participants and procedure

Thirty-three native speakers of German (all right-handed) participated in Experiment 1 (16 males; mean age 25.76, *SD* 2.42). They were drawn from the participant pool of the Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, and received monetary compensation for their participation. None of them reported a history of hearing problems.

Participants were randomly assigned to either the rising ($n = 17$; 7 males; mean age 25.29, *SD* 2.08) or the falling ($n = 16$; 9 males, mean age 26.25, *SD* 2.72) distribution condition. Participants first completed eight blocks (36 trials each) during the learning phase. On each trial, a single stimulus was randomly selected from category A (1,000 exemplars) or B (1,000 exemplars), with the following restrictions: (1) No stimulus could be selected more than once for a given participant in the learning phase of the experiment; (2) within each block, category A and B stimuli were equally probable ($p = .5$). After stimulus presentation, participants indicated whether it belonged to category A or category B by pressing one of two keys on a computer keyboard (button assignments for the two categories were counterbalanced across participants).

Following the response, participants received corrective feedback, which was displayed for 1 s in the middle of a CRT screen (Sony Multiscan E430). Correct feedback was given in bold green font (24 points), while incorrect feedback was given in bold red font (24 points). Participants were allowed a short break following each block.

Participants then completed two *maintenance* blocks (also 36 trials each). On each trial, participants were presented with a stimulus sampled from the equidistantly spaced grid described above. Critically, during the maintenance phase, participants did not receive feedback about their responses. The entire experiment lasted for about 20 min.

Stimuli were presented on a Windows-based PC, using the stimulation software PRESENTATION (Neurobehavioral Systems, Inc., version 13.9), and were transmitted through a Creative Labs Audigy II sound card onto Sennheiser HD 201 headphones.

Results

Accuracy results

Overall, performance differed significantly from chance ($d' = 1.35$, *SD* = 0.43), $t(32) = 32.94$, $p < .001$. Accuracy in

Table 1 Stimulus details in Experiment 1: Distribution ranges (first value, starting point; second value, end point), means (μ) and standard deviations (σ) are given for both spectral filter frequencies (S1, S2) in Bark. Physical δ' is calculated on the basis of distribution means and standard deviations. For rule-based (RB) optimal boundaries, δ' can be calculated for both dimensions, while for information integration (II)

Distribution	Band	Range A	Range B	μ_A	μ_B	σ_A	σ_B	RB δ'	II δ'
Rising	S1	3.5–7.4	6.5–10.4	5.47	8.48	0.94	0.92	3.24	4.41
	S2	7.5–11.4	10.5–14.4	9.47	12.47	0.93	0.94	3.23	
Falling	S1	10.5–6.4	7.5–3.4	5.47	8.46	0.93	0.94	3.22	4.41
	S2	11.5–7.4	14.5–10.4	12.55	9.52	0.92	0.92	3.26	

optimal boundaries, δ' can be calculated only for the integrated dimension and is based on the Euclidean distance between the distribution means in the two-dimensional space. A and B are arbitrary labels and refer to stimulus sets in the S1/S2 space, where the leftmost set (with low S1) is A and the rightmost (with high S2) is B

the learning phase was assessed by d' , a signal detection measure of perceptual sensitivity that is independent of response bias (Macmillan & Creelman, 2005). Figure 2 shows d' as a function of block separately for the rising and falling distributions. In order to assess learning, d' values were entered into a mixed-measures analysis of variance (ANOVA) with block as a within-subjects and distribution (rising vs. falling) as a between-subjects variable. For all ANOVAs, we report partial eta squared (η_p^2) as a measure of effect size and Greenhouse–Geisser-corrected p -values and degrees of freedom in cases of sphericity violations. There were no significant main effects [block, $F(6.1, 188.9) = 1.14$, $\eta_p^2 = .04$, $p = .34$; distribution, $F(1, 31) = 0.002$, $\eta_p^2 = .00001$, $p = .97$] and also no block \times distribution interaction, $F(6.1, 188.9) = 0.43$, $\eta_p^2 = .04$, $p = .89$. Hence, performance did not differ across blocks, and performance across blocks did not differ as a function of distribution condition.

Logistic regression

In order to assess the degree to which category membership judgments (i.e., A vs. B) depended on the acoustic dimensions under investigation (i.e., $S1$, $S2$), logistic regressions were calculated in order to predict category A responses from $S1$ and $S2$ and their interaction. Note that a significant β -weight indicates the importance of the dimension in determining category membership. Logistic regression models were calculated separately for the learning and the maintenance phases.

Learning phase The model comprised the regressors $S1$ and $S2$ and the factors *block* and *distribution* (rising, falling). The following interactions were also included in the model: $S1 \times S2$, $S1 \times \text{block}$, $S2 \times \text{block}$, $S1 \times \text{distribution}$, and $S2 \times \text{distribution}$. $S1$ values per trial significantly predicted category judgments, $\beta = -1.27$, $z = -4.82$, $p < .001$, but there was no interaction with block, $z = -0.02$, $p = .24$, indicating that participants similarly weighted $S1$ information in making category judgments over the course of the learning phase. The $S1 \times \text{distribution}$ interaction reached significance, $\beta = -0.35$, $z = -2.21$, $p < .05$, indicating that category A responses were better predicted by $S1$ in the falling than in the rising distribution. None of the other factors or interactions were significant, $z < 2$, n.s.

Maintenance phase The model included the same predictor variables listed above for the learning phase model, with the exception of block, which was not included here. There was no significant $S1$ effect, even though there was a trend for more category A responses at lower $S1$ values, $\beta = -1.82$, $z = -1.38$, $p < .16$; the $S1$ effect reached significance if the $S1 \times S2$ interaction was

removed from the model, $\beta = -2.34$, $z = -13.08$, $p < .001$. In the full model, there was a further significant effect of distribution, $\beta = 3.97$, $z = 2.06$, $p < .05$, reflecting that category A responses were better predicted in the falling than in the rising condition, and this effect was qualified by the significant $S1 \times \text{distribution}$ interaction, $\beta = -0.53$, $z = -1.92$, $p = .05$.

Learning versus maintenance phases Cue utilization during the learning versus maintenance phases was compared directly by entering the absolute values of β -weights for $S1$ and $S2$ into a mixed-measures ANOVA with the within-subjects factors *phase* (learning, maintenance) and *filter* ($S1$, $S2$) and the between-subjects factor *distribution* (rising, falling). Individual β -weights stemmed from the learning and maintenance models reported above, except that we did not include block in the learning phase model (parallel to the maintenance phase model). Note that for these analyses, only the magnitude (but not the sign) of the β -weights provided interesting information regarding the importance of each dimension to category judgments. There were significant main effects of filter, $F(1, 93) = 217.99$, $\eta_p^2 = .701$, $p < .001$, and phase, $F(1, 93) = 43.52$, $\eta_p^2 = .319$, $p < .01$. Higher β s were observed for $S1$ (2.55, $SD = 1.20$) than for $S2$ (0.62, $SD = 0.53$), and β s were higher in the maintenance (2.02, $SD = 1.57$) than in the learning (1.15, $SD = 0.90$) phase. Furthermore, there was a trend for a distribution \times filter interaction, $F(1, 93) = 2.03$, $\eta_p^2 = .021$, $p = .14$, reflecting that within the falling distribution, the difference between β s for $S1$ and $S2$ (2.70, $SD = 1.18$, vs. 0.55, $SD = 0.55$) was greater than within the rising distribution (2.43, $SD = 1.24$, vs. 0.67, $SD = 0.52$). There was also a filter \times phase interaction, $F(1, 93) = 16.50$, $\eta_p^2 = .151$, $p < .01$, indicating that β s for $S1$ and $S2$ differed more in the maintenance (3.25, $SD = 1.20$, vs. 0.78, $SD = 0.61$) than in the learning (1.86, $SD = 0.70$, vs. 0.45, $SD = 0.39$) phase.

In order to visualize the degree to which participants relied on the individual dimensions, $S1$ and $S2$, we plotted β s for $S2$ (ordinate) against β s for $S1$ (abscissa) in the learning and in the maintenance phases. In these scatterplots, participants are coded according to whether they significantly used $S1$ and $S2$ ($S1 + S2$; blue diamonds), $S1$ only ($S1$; red squares), $S2$ only ($S2$; green triangles), or none of the dimensions (purple circles) for categorization. Significant usage was determined by β s that significantly differed from zero on the basis of the single-subject logistic regression models ($\alpha = .05$). In these plots, participants who used both dimensions tended to fall on a diagonal. Participants with a preference for $S2$ are clustered near the ordinate, and participants with a preference for $S1$ are clustered near the abscissa (Fig. 3a). It can be seen from the figure that most participants relied on $S1$ in

both the learning and the maintenance phases. The percentages of significant β s did not significantly differ between the falling and rising distributions (all χ^2 s < 2.5, n.s.), even though we observed a trend for participants to be more likely to rely on both S1 and S2 in the falling, as compared with the rising, stimulus distribution.

In sum, the logistic regressions indicated that most participants relied on the first filter frequency, S1, during categorization and more strongly in the maintenance phase (without feedback) than in the learning phase (with feedback). On the other hand, some participants used information from both S1 and S2, while only very few exclusively used S2 for categorization.

Furthermore, categorization also somewhat differed between the falling and the rising distributions in that the reliance on S1 was greater in the falling than in the rising stimulus distribution and in that more participants tended to use both S1 and S2 in the falling than in the rising distribution.

Modeling results

Three families of decision bound models (e.g., Ashby & Gott, 1988; Maddox & Ashby, 1993) were fit to the data for each individual participant on a block-by-block basis to determine the decision strategy that best accounted for performance (Fig. 3b; see the Appendix for details): unidimensional rule-based, information integration, and random-response models. The two rule-based models assumed that listeners made use of unidimensional rules based on either S1 or S2. Two information integration models assumed an optimal decision bound or allowed decision bound slope and intercept to be free parameters but are summarized as one model for the remainder of this article. Finally, the random-response model presumes that participants guessed randomly on every trial. In order to assess whether the decision bound models provided substantial evidence, we transformed the respective Bayesian information criterion (BIC) scores to Bayes factors (Kass & Raftery, 1995;

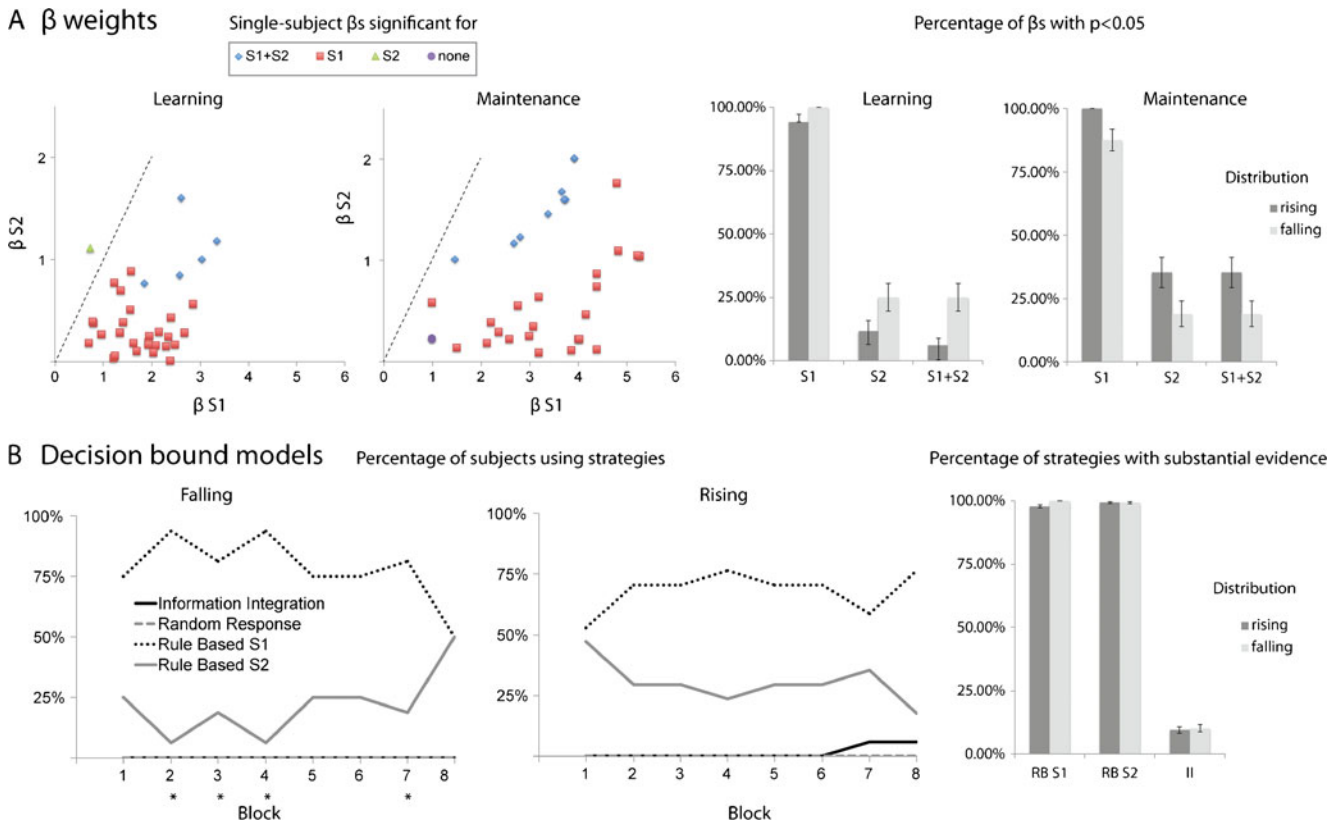


Fig. 3 Illustration of strategy use in Experiment 1. **a** Left: Scatterplots of β weights (per subject) for S1 and S2 in the learning phase (left) and in the maintenance phase (right). Participants are coded as follows: blue diamonds, both S1 β s and S2 β s significantly differ from zero; red squares, only S1 β s significantly differ from zero; green triangles, only S2 β s significantly differ from zero; purple circles, none of the β s significantly differ from zero. The diagonals illustrate the location of β s that would correspond equally to an S1- and S2-based response. Right: Percentage of significant β s in the learning and maintenance

phases, for the falling and rising distributions. Percentages are based on whether β s were significant for S1, S2, or both S1 and S2. **b** Left: Proportions of participants whose responses were best accounted for by the individual decision bound models. Block numbers marked with an asterisk indicate a significant difference in the proportion of participants using an S1 versus an S2 rule. Right: Percentage of rule-based (S1 and S2) and information integration models whose Bayes factors exceeded 3

Rafferty, 1986; see the Appendix) and subsequently used Jeffrey's suggested scale of evidence. According to this scale, Bayes factors greater than 3 indicate substantial evidence for model use.

Almost all model fits (per participant and block) for the rule-based S1 and S2 models provided substantial evidence, while only 10 % of the model fits for information integration exceeded this threshold (Fig. 3b, right). The percentages did not differ between distributions (i.e., falling vs. rising; all χ^2 s > 2, n.s.). Figure 3b (left) gives the proportion of listeners in the rising and falling distributions whose data were best fit by each of the tested models across blocks. All winning models had Bayes factors of >3.

Consistent with the results of the logistic regression analysis, participants in both the rising and falling distribution conditions made almost exclusive use of unidimensional rules, and participants were more likely to use a rule based on S1 than one based on S2. Chi-squared tests indicated that, overall, more participants relied on a unidimensional S1 rule, as compared with a unidimensional S2 rule, in six of the eight blocks. In order to account for multiple comparisons, we corrected our statistics with the false-discovery-rate (FDR) method (Benjamini & Hochberg, 1995; FDR-corrected α -level = .05). Taking the distribution conditions separately, participants in the falling distribution condition exhibited this pattern more strongly, using a unidimensional S1 rule more often than a unidimensional S2 rule on four of the eight blocks ($ps < .05$), whereas this difference was not significant in any block for the rising distribution condition.

Convergence of logistic regression and decision bound models

To our knowledge, no study has assessed the degree to which the two approaches, logistic regressions and decision bound models, converge. For this reason, we explored the relationship between block-averaged β -weights and goodness-of-fit measures (i.e., BICs) separately for the unidimensional S1 and S2 decision bounds in two ANOVAs with block-averaged BIC scores as dependent variables. We were effectively asking to what degree BIC-scores supporting a rule-based S1 or S2 strategy could be predicted from β -weights of S1 or S2 logistic regressions. Note that information integration models were not included in these analyses, since the proportion of participants using information integration in Experiment 1 was too small for meaningful comparisons.

Both models included the between-subjects factor *distribution* (rising, falling), the regressor *β -weight*, as well as the *β -weight \times distribution* interaction. The S1 model (with S1 BIC score as dependent variable) showed a significant effect

of S1 β -weight, $F(1, 29) = 28.16$, $\eta_p^2 = .492$, $p < .001$, reflecting a negative correlation between β -weights and BIC scores (i.e., higher β -weights for lower BIC scores). However, the correlation was not modulated by distribution, as evidenced by no other significant main effects or interactions (all F s < 3, all $ps > .15$). In the S2 model (with S2 BIC scores as dependent variable), the negative correlation between β -weights was not significant, $F(1, 29) = 2.33$, $\eta_p^2 = .074$, $p < .15$, overall, but depended on distribution [β -weight \times distribution: $F(1, 29) = 3.07$, $\eta_p^2 = .113$, $p = .05$]. The β -weight/BIC score correlation was significant for the falling, $F(1, 15) = 4.88$, $\eta_p^2 = .245$, $p < .05$, but not for the rising, $F(1, 14) = 0.39$, $\eta_p^2 = .027$, $p = .54$, distribution.

Overall, BIC scores supporting either S1 or S2 rule-based strategies negatively correlated with the corresponding absolute β -weights from the S1 and S2 logistic regression effects. Thus, β -weights and decision bound model BIC scores converged.

Prediction of performance by decision bound models

Finally, we tried to predict performance from strategy use; that is, we tested whether the likelihood of using a rule-based S1 or S2 categorization strategy was associated with better performance, as indexed by two separate mixed-measures ANOVAs with d' as the dependent variable, the proportions of rule-based S1 and S2 strategy use and distribution (falling, rising) as independent variables. Since proportions of rule-based S1 and S2 strategy use are necessarily highly correlated, the two factors were investigated in separate ANOVAs.

None of the ANOVAs showed significant main effects or interactions (all F s > 1, n.s.). Thus, performance in Experiment 1 did not depend on either rule-based S1 or S2 strategy use.

Discussion

Participants in Experiment 1 showed a strong preference for using a unidimensional rule-based decision bound for auditory categorization and primarily relied on the first filter frequency (S1). Thus, our prediction was borne out: Participants preferred a rule-based approach and did not exhibit differences in performance, as indexed by d' , as a function of block for either the rising or the falling distribution condition. Performance was overall high beginning from block 1, indicating that listeners in both conditions discovered a strategy yielding good categorization performance immediately. However, additional learning was not apparent over blocks. This is likely because our category distributions overlapped, causing some stimuli to be ambiguous and performance to plateau below ceiling levels.

Our observation that participants generally tended to adopt a unidimensional strategy when categorizing auditory stimuli is in line with the study of Goudbeek et al. (2009). However, here the stimulus dimensions were based on the same acoustic dimension (i.e., frequency), suggesting that rule-based categorization strategies are preferred even if dimensions are integral.

Without an S1–S2 correlation and without feedback (i.e., in the maintenance phase), the magnitude of β -values for S1 was larger than in the learning phase; this S1 preference was more pronounced in the falling than in the rising distribution. Hence, in the maintenance phase, participants seemed to rely on S1 to a greater extent for the falling than for the rising distribution. The assumed special role of the falling distribution condition was further analyzed in Experiment 2, where optimal performance depended on the usage of an information integration strategy and where the decision bound was a falling (with a negative slope) or rising (with a positive slope) diagonal.

Crucially, Experiment 2 was motivated by the observation that negatively correlated acoustic dimensions seemed to be preferred in speech (and more general, in audition), for which reason we assume that if dimensions are required to be integrated, this is done more readily for those that show a negative, as compared with a positive, correlation.

Experiment 2

Experiment 2 made use of stimulus distributions that required a predecisional integration of information from two spectral dimensions—that is, for which rule-based categorization was a suboptimal strategy. We were interested in whether a substantial proportion would use an information integration strategy, and the degree to which this choice depended on the nature of the correlation between spectral filter locations, S1 and S2 (i.e., falling vs. rising). Due to the assumed familiarity with negative acoustic correlations, we expected that information integration would be more readily used in the falling, as compared with the rising, stimulus distribution. We also assumed that if rule-based categorization is indeed predominant in audition (Goudbeek et al., 2009; Maddox et al., 2006), some participants in Experiment 2 would continue using this strategy. Finally, the use of information integration strategies in Experiment 2 should also depend on the availability of immediate feedback (Maddox, Ashby, & Bohil, 2003), for which reason we did not expect indications of information integration in the maintenance phase (designed as in Experiment 1).

Method

Materials

The stimuli were similar to those in Experiment 1, with the exception that (1) category A and B distributions were parallel in the S1–S2 space, rather than lying on the same diagonal as in Experiment 1, and (2) the spread was increased in both dimensions, S1 and S2, in order to render categorization more difficult. S1/S2 ranges and standard deviations are illustrated in Table 2. In contrast to Experiment 1, the normalized distance, δ' , was considerably higher for the information integration bound than for either of the rule-based bounds; thus, best performance would be attainable by an information integration strategy. Parallel to Experiment 1, stimuli for the maintenance phase consisted of a 6×6 grid that covered the critical region of the stimulus space;

Participants and procedure

Thirty-six native speakers of German (all right-handed) participated in Experiment 2 (19 males; mean age 25.14, SD 4.02); participants were drawn from the same pool as in Experiment 1, although none of the participants had been recruited for Experiment 1. As before, participants were assigned to either the rising (11 males; mean age 24.67, SD 3.09) or falling stimulus (8 males; mean age 25.61, SD 4.82) distribution.

Participants received monetary compensation for their participation. No participant reported hearing problems. The procedure was identical to that in Experiment 1.

Results

Accuracy results

Accuracy as measured by d' differed significantly from chance ($d' = 1.33$, $SD = 0.56$), $t(35) = 19.78$, $p < .001$. As in Experiment 1, learning was assessed in a mixed-measures ANOVA on d' with *block* and *distribution* as independent variables. The main effect of block was significant, $F(5.5, 188) = 2.09$, $\eta_p^2 = .060$, $p < .05$, indicating that performance increased over time. There was also a significant main effect of distribution, $F(1, 34) = 35.03$, $\eta_p^2 = .507$, $p < .001$, with participants in the falling distribution condition ($d' = 1.61$) outperforming participants in the rising distribution condition ($d' = 1.04$). Learning rate, however, did not depend on distribution, as indicated by a nonsignificant block \times distribution interaction, $F(5.5, 188) = 0.50$, $\eta_p^2 = .014$, $p = .84$. The accuracy results are illustrated in Fig. 4.

Table 2 Stimulus details in Experiment 2. Distribution ranges, means (μ) and standard deviations (σ) are given for both spectral bands (S1, S2). δ' is calculated on the basis of distribution means and standard deviations. For rule-based (RB) optimal boundaries, δ' can be

Distribution	Band	Range A	Range B	μ_A	μ_B	σ_A	σ_B	RB δ'	II δ'
Rising	S1	1.5–5.4	4.5–8.4	3.49	6.48	1.04	1.03	2.86	4.05
	S2	11.5–15.4	8.5–12.4	13.49	10.49	1.06	1.05	2.88	
Falling	S1	1.5–5.4	4.5–8.4	3.47	6.49	1.05	1.05	2.88	4.06
	S2	12.5–8.4	15.5–11.4	10.52	13.52	1.06	1.03	2.87	

Logistic regression

Logistic regression analyses were conducted as in Experiment 1, which predicted the likelihood of a category A response from S1 and S2 separately for the learning and maintenance phases.

Learning phase The model comprised the regressors S1 and S2, the factors *block* and *distribution* (rising, falling), and the interactions S1 \times S2, S1 \times block, S2 \times block, S1 \times distribution, and S2 \times distribution. Notably, there was a main effect of S1, $\beta = -1.38$, $z = -5.60$, $p < .001$; that is, S1 values per trial significantly predicted category judgments. The significant block main effect, $\beta = 0.33$, $z = -3.23$, $p < .01$, and the S1 \times block interaction, $\beta = -0.04$, $z = -4.32$, $p < .001$, together indicated that category A responses could be increasingly better predicted over the course of the experiment, especially on the basis of S1. Furthermore, category A responses were generally

predicted better in the falling than in the rising distribution, $\beta = 9.41$, $z = 20.41$, $p < .001$. The distribution factor furthermore interacted with both S1, $\beta = -0.50$, $z = -9.00$, $p < .01$, and S2, $\beta = -0.59$, $z = -18.23$, $p < .001$, reflecting larger S1 and S2 effects for the falling than for the rising distribution. Finally, there was a significant interaction of S1 and S2, $\beta = 0.60$, $z = 2.75$, $p < .01$.

Maintenance phase In the model without a block factor, both S1, $\beta = -2.29$, $z = -7.37$, $p < .001$, and S2, $\beta = -0.54$, $z = -3.69$, $p < .01$, were significant and interacted with distribution (S1 \times distribution, $\beta = -0.96$, $z = -8.88$, $p < .001$; S2 \times distribution, $\beta = -0.44$, $z = -6.32$, $p < .001$). S1 and S2 were significant predictors for category A responses, and more so for the falling than for the rising distribution. In general, category A responses were predicted better in the falling than in the rising distribution, $\beta = 10.07$, $z = 9.26$, $p < .001$. Finally, as in the learning phase, the S1 \times S2 interaction was significant, $\beta = 0.72$, $z = 4.79$, $p < .001$.

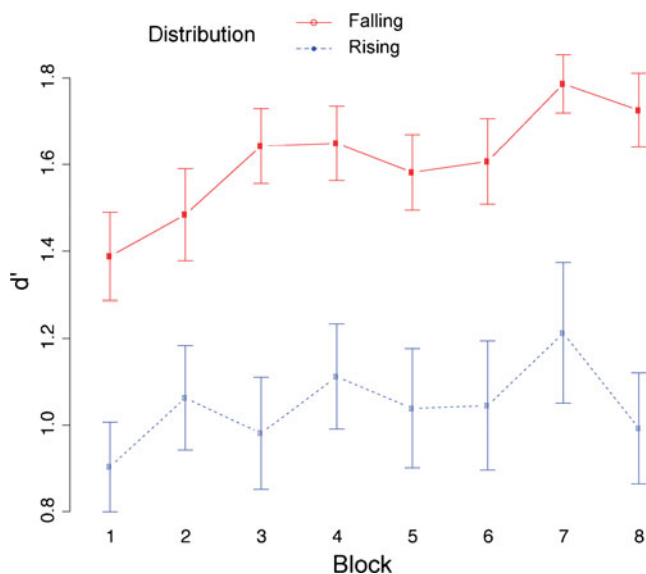


Fig. 4 Performance measure per block in the learning phase of Experiment 2. Whiskers indicate across-participants standard errors of the means

Learning versus maintenance phase The absolute values of the single-subject β -weights (from the same models as those reported in Experiment 1) were used as the dependent variable in an ANOVA with the factors *phase* (learning, maintenance), *distribution* (rising, falling), and *filter* (S1, S2). The ANOVA revealed main effects of filter, $F(1, 102) = 184.21$, $\eta_p^2 = .644$, $p < .001$, phase, $F(1, 102) = 28.66$, $\eta_p^2 = .219$, $p < .001$, and distribution, $F(1, 34) = 29.23$, $\eta_p^2 = .462$, $p < .001$. Weights were higher for S1 than for S2 (1.77, $SD = 1.06$, vs. 0.47, $SD = 0.37$), and higher in the maintenance phase than in the learning phase (1.38, $SD = 1.26$, vs. 0.86, $SD = 0.64$). β -weights were larger in the falling distribution than in the rising distribution (1.44, $SD = 1.21$ vs. 0.80, $SD = 0.67$). Furthermore, the filter \times distribution, $F(1, 102) = 13.85$, $\eta_p^2 = .119$, $p < .001$, phase \times distribution, $F(1, 102) = 7.90$, $\eta_p^2 = .072$, $p < .01$, and filter \times phase, $F(1, 102) = 13.86$, $\eta_p^2 = .120$, $p < .001$, interactions reached significance, reflecting larger β differences between S1 and S2 in the falling (2.27, $S = 1.19$, vs. 0.61, $SD = 0.38$) than in the rising (1.28, $SD =$

0.60, vs. 0.33, $SD = 0.30$) distribution and in the maintenance (2.21, $SD = 1.26$, vs. 0.55, $SD = 0.43$) than in the learning (1.34, $SD = 0.55$, vs. 0.39, $SD = 0.28$) phase. Importantly, β differences between the maintenance and learning phases were larger for the falling (1.85, $SD = 1.50$, vs. 1.04, $SD = 0.65$) than for the rising (0.93, $SD = 0.73$, vs. 0.68, $SD = 0.59$) distribution. The three-way filter \times distribution \times phase interaction was significant as well, $F(1, 102) = 10.02$, $\eta_p^2 = .090$, $p < .01$, motivating separate analyses for the learning and the maintenance phases.

These analyses showed a significant filter \times distribution interaction in the maintenance phase, $F(1, 34) = 15.07$, $\eta_p^2 = .307$, $p < .01$, but not in the learning phase, $F(1, 34) = 0.31$, $\eta_p^2 = .009$, $p = .58$. In order to visualize the degree to which participants relied on the individual dimensions, S1 and S2, we plotted β s for S1 (abscissa) against β s for S2 (ordinate) in the learning and in the maintenance phases, separately for the rising and the falling distributions (Fig. 5a). The plots illustrate that β s were larger in the maintenance than in the learning phase and also larger in the falling than in the rising distribution.

Notably, more participants used both dimensions, S1 and S2, in the falling than in the rising distribution, as reflected by significant differences in proportions of significant β s from the single-subject logistic regressions, $\chi^2 = 8.86$, p (FDR) $< .05$ (Fig. 5a, right). This distinction was more pronounced in the learning than in the maintenance phase, where the proportions did not differ, $\chi^2 < 2$, n.s.. In the same vein, participants used S2 more often in the falling than in the rising distribution of the learning phase, $\chi^2 = 7.26$, p (FDR) $< .05$, while again, proportions did not differ in the maintenance phase, $\chi^2 < 2$, n.s.

Modeling results

As in Experiment 1, three families of decision bound models (e.g., Ashby & Gott, 1988; Maddox & Ashby, 1993) were fit to the learning-phase data—that is, unidimensional rule-based (S1, S2), information integration, and random-response models (Fig. 5b, left). Again, we calculated Bayes factors for each model fit. In contrast to Experiment 1, the proportion of models that received substantial evidence differed significantly between the falling and the rising distributions (Fig. 5b, right) all χ^2 s > 3 , p (FDR) $< .05$. All winning models had Bayes factors of > 3 .

The proportions of participants fit best by each model are shown in Fig. 5b (left). Overall, participants most often adopted a unidimensional rule-based strategy based on S1, as in Experiment 1. However, strategy use crucially differed between the rising and falling distribution conditions. For the rising distribution, the unidimensional S1 rule was used

in the majority of cases (six of eight blocks; $ps < .05$, FDR-corrected). In the falling condition, participants used an information integration strategy as often as the rule-based (S1) strategy in all eight blocks. Thus, participants trained on the falling distributions were more likely to adopt the optimal strategy that involved integrating S1 and S2 information before making a category membership decision.

Convergence of logistic regression and decision bound models

As before, the convergence of β -weights and BIC scores was assessed in two ANOVAs. The first ANOVA comprised the dependent measure *rule-based S1 BIC score* and the independent variables *S1- β -weight* and *distribution*. Importantly, there was a main effect of β -weight, $F(1, 32) = 130.53$, $\eta_p^2 = .803$, $p < .001$, reflecting a negative correlation between S1- β s and RB S1 BIC scores, and a main effect of distribution, $F(1, 32) = 20.82$, $\eta_p^2 = .394$, $p < .001$, showing lower BIC scores (better fits) for the falling than for the rising distribution.

The β -weight \times distribution interaction, $F(1, 32) = 6.73$, $\eta_p^2 = .803$, $p < .05$, indicated a stronger β -BIC score correlation in the falling than in the rising condition.

The second ANOVA comprised the dependent measure *RB S2 BIC score* and the independent variables *S2- β* and *distribution* and showed an effect of β -weight, $F(1, 32) = 133.67$, $\eta_p^2 = .806$, $p < .001$, as well as an effect of distribution, $F(1, 32) = 11.23$, $\eta_p^2 = .260$, $p < .01$, but no β -weight \times distribution interaction, $F(1, 32) = 1.14$, $\eta_p^2 = .034$, $p = .29$. Again, as in Experiment 1, β -weights and BIC scores converged.

Prediction of performance by decision bound models

In order to directly assess the performance benefit of using an information integration strategy, we carried out a correlation analysis between the proportions of rule-based S1 and information integration strategy use and d' . Notably, there was a positive correlation of proportion of information integration use and d' , $r = .48$, $t = 3.20$, $p < .01$, suggesting that using an information integration strategy was indeed beneficial for performance. By contrast, the correlation of proportion of rule-based S1 use and d' was negative, $r = -0.23$, $t = -1.39$, n.s.; that is, participants using a rule-based S1 strategy tended to perform worse.

Discussion

The important result of Experiment 2 is that, as compared to Experiment 1, more participants used an information

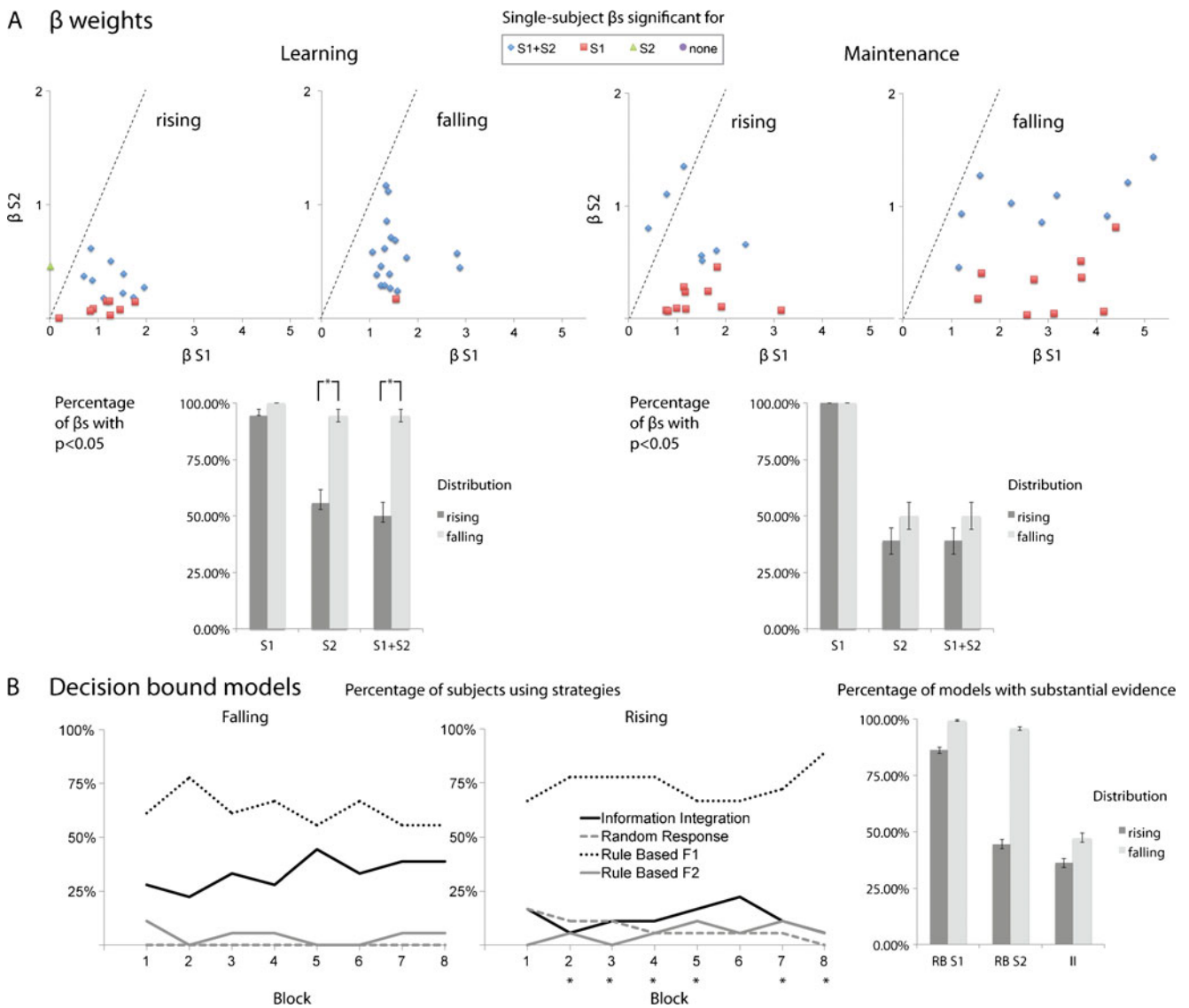


Fig. 5 Illustration of strategy use in Experiment 2. **a** Scatterplots of β weights (per participant) for S1 and S2, for the rising and falling distributions in the learning phase (left) and in the maintenance phase (right). Participants are coded as follows: blue diamonds, both S1 β s and S2 β s significantly differ from zero; red squares, only S1 β s significantly differ from zero; green triangles, only S2 β s significantly differ from zero; purple circles, none of the β s significantly differ from zero. The diagonals illustrate the location of β s that would correspond equally to an S1- and S2-based response. The percentage of significant

β s ($p < .05$) for the dimensions S1 and S2 and the combination of S1 and S2 are illustrated below the scatterplots. Significant proportion differences are indicated with an asterisk. **b** Left: Proportions of decision bound models used to categorize stimuli from the rising (left) and falling (right) distributions. Block numbers marked with an asterisk indicate a significant difference in the proportion of participants using an S1 versus an information integration rule. Right: Percentage of rule-based (S1 and S2) and information integration models whose Bayes factors exceeded 3

integration strategy, and more so in the falling than in the rising distribution condition. Generally, participants who were more likely to use information integration performed better than those who were more likely to focus on an S1 rule-based strategy.

Intriguingly, despite being disadvantageous, participants still used the S1 dimension to a high degree, as evidenced by both logistic regressions and decision bound models. That is, although Experiment 2 examined classification performance for auditory stimuli that were optimally separated by

an information integration (diagonal) decision bound, participants still frequently used a rule-based strategy based on S1 (cf. Goudbeek et al., 2009).

Crucially, the use of the optimal information integration strategy depended on whether stimulus distributions were rising or falling (see Fig. 1b). The falling condition was associated more strongly with use of an information integration decision strategy, and the resulting performance was shown to be better for individuals adopting an information integration strategy. This result was predicted, since we

hypothesized that familiarity with negative acoustic (here, spectral) correlations between dimensions would promote their integration.

Finally, a comparison of β -weights for the learning and maintenance phases suggested that the correlation of the stimulus dimensions in the learning phase, as well as immediate feedback, promoted the use of both S1 and S2 dimensions for categorization. This is consistent with previous findings in vision research (cf. Maddox et al., 2003). Our analyses suggested that information integration is characterized by an equal usage of S1 and S2 and that it was present in the learning phase, but not in the maintenance phase. There, without an S1–S2 correlation and without feedback, we observed a significant shift to an almost exclusive rule-based use of S1.

General discussion

Two experiments examined auditory category formation for stimuli varying along two spectral dimensions (i.e., S1 and S2), which exhibited either positive or negative correlations. Decision bound modeling and logistic regressions yielded three main results: (1) better information integration for negative spectral correlations, (2) tendency toward rule-based S1 categorization overall, and (3) dependency of immediate corrective feedback for information integration.

Promotion of information integration by negative correlations

The most important result of the two experiments is that the use of information integration (confined to Experiment 2) depended on the nature of the correlation between dimensions: Negative correlations promoted information integration, while positive correlations inhibited it. Overall, information integration in Experiment 2 predicted better performance.

Regarding the special status of negative acoustic correlations, our study extends the phonetic work by Kingston and colleagues (Kingston et al., 2008; Kingston & Macmillan, 1995). Kingston and colleagues characterized the interaction of fundamental frequency (f_0 , pitch) and first resonance frequency ($F1$) in the human oral cavity, as well as the interaction of $F1$ and nasalization (i.e., the resonance frequencies in the human nasal cavity) in speech vowel and nonspeech vowel-like sounds, and observed that both the f_0 – $F1$ and $F1$ –nasalization relations approximate a negative correlation. With respect to nasalization, vowels with a higher degree of nasalization tend to have lower $F1$ frequencies (Diehl et al., 1990). Kingston and colleagues demonstrated that both of the discussed negative correlations (f_0 – $F1$ and $F1$ –nasalization), in comparison with their positive

counterparts, led to better categorization performance for speech and nonspeech stimuli (Kingston et al., 2008; Kingston & Macmillan, 1995). On the basis of these findings, we suggest that there is a general inclination toward encountering negative acoustic correlations in speech (resulting from articulatory bases as discussed in Carre, 2009), possibly shaping the learning of nonspeech stimuli with similar negative correlations.

Preference of rule-based strategies

In both experiments, participants predominantly based their categorization on the first spectral filter, S1. This inclination, which is in line with previous research (Goudbeek et al., 2009; Maddox et al., 2006), seems remarkable in the light of Experiment 2, where rule-based categorization was clearly suboptimal and where participants using this strategy performed worse than those employing information integration. In general, this inclination may have a neural explanation: Previous neuroimaging studies on visual categorization have shown that cortico–striatal connections are vital for information integration (Ashby & Ell, 2001; Ashby & Ennis, 2006; Ashby & Spiering, 2004; Helie, Roeder, & Ashby, 2010; Nomura et al., 2007; Seger, 2008). Moreover, cortico–striatal connections between the auditory cortex and the caudate have been argued to be more diffuse than cortico–striatal connections between the visual cortex and the caudate (Maddox et al., 2006). Thus, information integration may be relatively less likely in auditory categorization than in vision, due to anatomical constraints.

A second possible explanation for the reliance on rule-based strategies is developmental in nature. We have argued that information integration depends on acoustic correlations familiar from speech; thus, speech itself should presumably be acquired predominantly by information integration. A potential bias toward information integration learning in early life is related to the observation that brain structures such as the prefrontal and medial cortices that support rule-based learning (Gabrieli, Brewer, Desmond, & Glover, 1997; Schacter & Wagner, 1999) develop relatively late (Diamond, 2002). As a result, rule-based learning in early life does not compete with information integration learning as strongly as in adolescence or adulthood (Huang-Pollock, Maddox, & Karalunas, 2011).

The preference for using specifically the S1 dimension during categorization may reflect that pitch changes (presumably, the perceptual dimension conveyed by S1 variation; Remez & Rubin, 1993) are easier to verbalize than timbre changes (as potentially reflected by S2; cf. Smits et al., 2006). That is, pitch changes can be easily verbally described as being “high” or “low,” while timbral differences are more difficult to label verbally. The easier-to-verbalize dimension (i.e., S1) was then not surprisingly

more often used for rule-based categorization, which, by definition, is based on easy-to-verbalize rules (Ashby & Maddox, 2005; Ashby & Maddox, 2011).

The influence of immediate feedback

The third main finding from the present experiments concerns the comparison of the learning phase with the maintenance phase. First, the learning phase contained distributional information from which the correlations between dimensions could be extracted, while in the maintenance phase, no such information was available. Second, there was immediate feedback in the learning phase, but not in the maintenance phase. Previous research indicates that both of these aspects may contribute to the likelihood that information integration strategies are used for categorization. Research from vision provides evidence that immediate feedback is crucial for adopting an information integration strategy (cf. Ashby et al., 1999; Ashby & Waldron, 1999). Additionally, β -weights in the maintenance phase of both experiments were larger, and more so for S1 than for S2. The falling condition in Experiment 2 furthermore showed that even though both dimensions were used during learning, concomitant with information integration, participants reverted to using S1 in the maintenance phase.

Potential limitations of decision bound models

Decision bound models are not unequivocally accepted; they implicitly assume dissociations of multiple memory systems subserving rule-based and information integration learning. In particular, they assume that rule-based learning requires working memory, while information integration does not (Filoteo, Lauritzen, & Maddox, 2010). Challenging this claim, (Lewandowsky, Yang, Newell, & Kalish, 2012; Newell & Dunn, 2008; Newell, Dunn, & Kalish, 2010), showed that rule-based as well as information integration learning tax working memory.

Furthermore, decision bound models are not the only means by which auditory categorization can be modeled. On the one hand, *prototype models* (Rosch, 1973) assume that a novel stimulus is assigned to the category whose average or most representative member (i.e., the prototype) it is most similar to. Different formulations of prototype theory suggest that categorization decisions are also based in part on the spread of category members around the prototype (i.e., category variance; Nearey & Assmann, 1986). Exemplar models (Nosofsky, 1986), on the other hand, assume (in their most extreme formulation) that categorization involves comparing novel acoustic items with all previously encountered members of relevant existing categories and then making a category membership decision on the basis of the maximum of the summed similarities to the

members of the relevant categories. Prototype and exemplar models have been extensively compared elsewhere (Nosofsky & Stanton, 2005; Tunney & Fernie, 2012) and have also found to be somewhat inferior to decision bound models (Smits et al., 2006).

The present data show that decision bound models provide a valuable tool for assessing the contribution of different acoustic dimensions to auditory categorization. Furthermore, since Experiment 2, in particular, provides converging results from logistic regressions and decision bound models regarding the use of specific stimulus dimensions for categorization, we are confident that decision bound models accurately account for participants' response behavior.

Conclusions

In sum, the present experiments provided evidence that listeners tend to use a rule-based—that is, an explicit—hypothesis-testing approach when they are categorizing novel auditory stimuli. Our results further suggest that long-term experience with sound distributions characterized by a negative spectro-spectral (S1–S2) correlation shapes the categorization of novel auditory stimuli. This is in line with experiments on speech sound categorization where long-term experience with correlations among auditory dimensions could not be easily overridden by short-term exposure to contrasting dimension correlations (Idemaru & Holt, 2011). Even though this long-term experience promoted information integration strategies and better performance in auditory categorization, our experiments challenge the view that the most successful strategy is necessarily the one participants use most frequently.

Appendix: Methods

To model the learning process, we fit a number of decision bound models (DBMs) to each listener's data on a block-by-block basis. DBMs assume that a single stimulus presentation is represented in a multidimensional perceptual space and that each stimulus can be mapped to perceptual (i.e., internal) space by a transformation corresponding to a psychophysical function:

$$\Psi(\mathbf{y}_i) = \mathbf{x}_{pi} + \mathbf{e}_{pi}, \quad (1)$$

where \mathbf{e}_{pi} is a random vector representing perceptual noise. Here, we assume a one-to-one mapping of physical to perceptual coordinates but allow for trial-by-trial (unbiased) variability in the percept.

According to decision bound theory, participants make categorization decisions on the basis of division of the

psychophysical space by a response criterion. We thus fit a number of DBMs to the data of each listener for each block in order to estimate the response criterion that best accounted for the listener’s pattern of responses. For each experiment, we fit two unidimensional models, two information integration models, and one random-response model.

Unidimensional rules assume that a listener makes a categorization decision on the basis of one dimension only by setting a response criterion, λ_1 , at a location along the relevant dimension. Given this criterion location, the probability of responding “category A”, $P(R_A)$, is

$$P(R_A|\mathbf{x}) = P[x_1 + e_{p1} < \lambda_1 + e_{c1}], \tag{2}$$

and the probability of responding “category B,” $P(R_B)$, is

$$P(R_B|\mathbf{x}) = 1 - P(R_A|\mathbf{x}), \tag{3}$$

where λ_1 is the response criterion location, e_{c1} is criterial error, and e_{p1} is perceptual noise on the relevant dimension. In the model, e_{c1} and e_{p1} are assumed to be independent and identically distributed. Equation 2 can be rewritten as

$$P(R_A|\mathbf{x}) = \Phi\left(\frac{\lambda_1 - x_1}{\sqrt{(\sigma_p^2 + \sigma_c^2)}}\right), \tag{4}$$

where Φ is the normal cumulative distribution function. In the model, σ_p and σ_c cannot be separately determined, so we fit only one noise parameter, $\sigma^2 = \sigma_p^2 + \sigma_c^2$. Thus, each unidimensional model has two free parameters: the variability parameter, σ , and the response criterion location, λ_1 . We fit unidimensional models based on both S1 and S2.

Information integration models assume that the listener integrates the S1 and S2 values before making a decision about category membership. The response criterion location, λ_{12} , can then be described in two-dimensional space by assuming a slope, b , and intercept, c_0 . The probability of making a “category A” response for a two-dimensional stimulus is then

$$P(R_A|x) = \Phi\left(\frac{\lambda_{12} - x_1x_2}{\sqrt{\sigma_p^2 + \sigma_c^2}}\right) \tag{5}$$

Two versions of the information integration model were fit to each listener’s data. The first assumed that the response criterion location was oriented optimally; this model thus had only one free parameter, σ . The second information integration model thus allowed the slope and intercept of the response criterion location to vary, and so had three free parameters: b , c_0 , and σ .

The random-response rule modeled the probability of responding “category A,” $P(R_A)$, as the frequency of actual “category A” responses for each listener in each block,

ignoring the value of the stimulus on either dimension. This model had one free parameter that was estimated from the data, the observed frequency of “category A” responses.

In order to determine which DBM best fit each participant’s data on a block-by-block basis, all DBMs were fit to responses using maximum likelihood methods. Best-fitting parameters were found with MATLAB’s constrained non-linear optimization routine based on a quasi-Newton approximation of the Hessian function. We used the BIC (Kass & Raftery, 1995) for model comparisons. The BIC is calculated for each model according to

$$BIC = -2ML_i + j_i \log(n), \tag{6}$$

where ML_i is the maximum log-likelihood of model i , j is the number of parameters in the model, and n is the number of observations. The number of parameters in the expression serves as a handicap for model complexity; the model with the smallest BIC is selected as the best-fitting model.

Bayes factors were derived from BIC scores on the basis of approximation formulae provided in Raftery (1986) and rewritten as

$$B_1 = e^{\frac{-1(M_1 - M_0)}{2}}, \tag{7}$$

where M_1 is the BIC score to be converted into the Bayes factor and M_0 is the BIC score of the alternative model—here, the random-response model.

References

- Ashby, F. G. (1992). Multidimensional models of categorization. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 449–484). Hillsdale, NJ: Lawrence Erlbaum.
- Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*(3), 442–481.
- Ashby, F. G., & Crossley, M. J. (2010). Interactions between declarative and procedural-learning categorization systems. *Neurobiology of Learning and Memory*, *94*(1), 1–12.
- Ashby, F. G., & Ell, S. W. (2001). The neurobiology of human category learning. *Trends in Cognitive Sciences*, *5*(5), 204–210.
- Ashby, F. G., & Ennis, J. M. (2006). The role of the basal ganglia in category learning. *Psychology of Learning and Motivation*, *46*, 1–36.
- Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 33–53.
- Ashby, E. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, *56*, 149–178.
- Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences*, *1224*(1), 147–161.
- Ashby, F. G., Queller, S., & Berretty, P. M. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Perception & Psychophysics*, *61*(6), 1178–1199.
- Ashby, F. G., & Spiering, B. J. (2004). The neurobiology of category learning. *Behavioral and Cognitive Neuroscience Reviews*, *3*(2), 101–113.

- Ashby, F. G., & Waldron, E. M. (1999). On the nature of implicit categorization. *Psychonomic Bulletin & Review*, 6(3), 363–378.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B, Methodological*, 57(1), 289–300.
- Boersma, P., & Weenink, D. (2011). PRAAT: Doing Phonetics by Computer (ver. 5.2.24). Amsterdam: Institut for Phonetic Sciences.
- Carre, R. (2009). Dynamic properties of an acoustic tube: Prediction of vowel systems. *Speech Communication*, 51(1), 26–41.
- Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. Stuss & R. Knight (Eds.), *Principles of Frontal Lobe Function* (pp. 466–503). New York: Oxford University.
- Diehl, R. L., Kluender, K. R., & Walsh, M. A. (1990). Some auditory bases of speech perception and production. In W. A. Ainsworth (Ed.), *Advances in speech, hearing and language processing* (pp. 243–268). London: JAI Press.
- Filoteo, J. V., Lauritzen, S., & Maddox, W. T. (2010). Removing the frontal lobes: the effects of engaging executive functions on perceptual category learning. *Psychological Science*, 21(3), 415–423.
- Gabrieli, J. D. E., Brewer, J. B., Desmond, J. E., & Glover, G. H. (1997). Separate neural bases of two fundamental memory processes in the human medial temporal lobe. *Science*, 276(5310), 264–266.
- Goudbeek, M., Cutler, A., & Smits, R. (2008). Supervised and unsupervised learning of multidimensionally varying non-native speech categories. *Speech Communication*, 50(2), 109–125.
- Goudbeek, M., Swingle, D., & Smits, R. (2009). Supervised and unsupervised learning of multidimensional acoustic categories. *Journal of Experimental Psychology: Human Perception and Performance*, 35(6), 1913–1933.
- Grau, J. W., & Kemler Nelson, D. G. (1988). The distinction between integral and separable dimensions: evidence for the integrality of pitch and loudness. *Journal of Experimental Psychology: General*, 117(4), 347–370.
- Guenther, F. H., & Bohland, J. W. (2002). Learning sound categories: a neural model and supporting experiments. *Acoustical Science and Technology*, 23(4), 213–220.
- Guenther, F. H., Nieto-Castanon, A., Ghosh, S. S., & Tourville, J. A. (2004). Representation of sound categories in auditory cortical maps. *Journal of Speech, Language, and Hearing Research*, 47(1), 46–57.
- Helie, S., Roeder, J. L., & Ashby, F. G. (2010). Evidence for cortical automaticity in rule-based categorization. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 30(42), 14225–14234.
- Hilbe, J. M. (2009). *Logistic Regression Models*. New York: Chapman & Hall.
- Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America*, 119(5), 3059–3071.
- Holt, L. L., & Lotto, A. J. (2008). Speech perception within an auditory cognitive science framework. *Current Directions in Psychological Science*, 17(1), 42–46.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (Vol. 354). New York: Wiley.
- Huang-Pollock, C. L., Maddox, W. T., & Karalunas, S. L. (2011). Development of implicit and explicit category learning. *Journal of Experimental Child Psychology*, 109(3), 321–335.
- Idemaru, K., & Holt, L. L. (2011). Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology: Human Perception and Performance*, 37(6), 1939–1956.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795.
- Kingston, J., Diehl, R. L., Kirk, C. J., & Castleman, W. A. (2008). On the internal perceptual structure of distinctive features: The [voice] contrast. *Journal of Phonetics*, 36(1), 28–54.
- Kingston, J., & Macmillan, N. A. (1995). Integrality of nasalization and F1 in vowels in isolation and before oral and nasal consonants: a detection-theoretic application of the Garner paradigm. *Journal of the Acoustical Society of America*, 97(2), 1261–1285.
- Kingston, J., Macmillan, N. A., Dickey, L. W., Thorburn, R., & Bartels, C. (1997). Integrality in the perception of tongue root position and voice quality in vowels. *Journal of the Acoustical Society of America*, 101(3), 1696–1709.
- Lehiste, I., & Peterson, G. E. (1961). Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America*, 33(4), 419–419.
- Lewandowsky, S., Yang, L.-X., Newell, B. R., & Kalish, M. L. (2012). Working memory does not dissociate between different perceptual categorization tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(4), 881–904.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide*. Mahwah, NJ: Erlbaum.
- Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, 53(1), 49–70.
- Maddox, W. T., Ashby, F. G., & Bohil, C. J. (2003). Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(4), 650–662.
- Maddox, W. T., Filoteo, J. V., Hejl, K. D., & Ing, A. D. (2004). Category number impacts rule-based but not information-integration category learning: Further evidence for dissociable category-learning systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(1), 227–245.
- Maddox, W. T., Filoteo, J. V., Lauritzen, J. S., Connally, E., & Hejl, K. D. (2005). Discontinuous categories affect information-integration but not rule-based category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(4), 654–669.
- Maddox, W. T., Ing, A. D., & Lauritzen, J. S. (2006). Stimulus modality interacts with category structure in perceptual category learning. *Perception & Psychophysics*, 68(7), 1176–1190.
- Maddox, W. T., Molis, M. R., & Diehl, R. L. (2002). Generalizing a neuropsychological model of visual categorization to auditory categorization of vowels. *Perception & Psychophysics*, 64(4), 584–597.
- McQueen, J. (1996). Phonetic categorisation. *Language & Cognitive Processes*, 11(6), 655–664.
- Mirman, D., Holt, L. L., & McClelland, J. L. (2004). Categorization and discrimination of nonspeech sounds: Differences between steady-state and rapidly-changing acoustic cues. *Journal of the Acoustical Society of America*, 116(2), 1198–1207.
- Nearey, T. M., & Assmann, P. F. (1986). Modeling the role of inherent spectral change in vowel identification. *Journal of the Acoustical Society of America*, 80(5), 1297–1297.
- Newell, B. R., & Dunn, J. C. (2008). Dimensions in data: testing psychological models using state-trace analysis. *Trends in Cognitive Sciences*, 12(8), 285–290.
- Newell, B. R., Dunn, J. C., & Kalish, M. (2010). The dimensionality of perceptual category learning: a state-trace analysis. *Memory and Cognition*, 38(5), 563–581.
- Nomura, E. M., Maddox, W. T., Filoteo, J. V., Ing, A. D., Gitelman, D. R., Parrish, T. B., Mesulam, M. M., & Reber, P. J. (2007). Neural

- correlates of rule-based and information-integration visual category learning. *Cerebral Cortex*, 17(1), 37–43.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, 115(1), 39–39.
- Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(4), 700–708.
- Nosofsky, R. M., & Stanton, R. D. (2005). Speeded classification in a probabilistic category structure: contrasting exemplar-retrieval, decision-boundary, and prototype models. *Journal of Experimental Psychology: Human Perception and Performance*, 31(3), 608–629.
- Raftery, A. E. (1986). A note on Bayes factors for log-linear contingency table models with vague prior information. *Journal of the Royal Statistical Society, Series B*, 48, 249–250.
- Rao, S. M., Bobholz, J. A., Hammeke, T. A., Rosen, A. C., Woodley, S. J., Cunningham, J. M., Cox, R. W., Stein, E. A., & Binder, J. R. (1997). Functional MRI evidence for subcortical participation in conceptual reasoning skills. *NeuroReport*, 8(8), 1987–1993.
- Remez, R. E., & Rubin, P. E. (1993). On the intonation of sinusoidal sentences: contour and pitch height. *Journal of the Acoustical Society of America*, 94(4), 1983–1988.
- Rosch, E. (1973). Natural categories. *Cognitive Psychology*, 4(3), 328–350.
- Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and Categorization* (pp. 27–48). Hillsdale: Lawrence Erlbaum Associates.
- Russ, B. E., Lee, Y. S., & Cohen, Y. E. (2007). Neural and behavioral correlates of auditory categorization. *Hearing Research*, 229(1–2), 204–212.
- Schacter, D. L., & Wagner, A. D. (1999). Medial temporal lobe activations in fMRI and PET studies of episodic encoding and retrieval. *Hippocampus*, 9(1), 7–24.
- Seger, C. A. (2008). How do the basal ganglia contribute to categorization? Their roles in generalization, response selection, and learning via feedback. *Neuroscience and Biobehavioral Reviews*, 32(2), 265–278.
- Seger, C. A., & Cincotta, C. M. (2005). The roles of the caudate nucleus in human classification learning. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 25(11), 2941–2951.
- Silbert, N. H., Townsend, J. T., & Lentz, J. J. (2009). Independence and separability in the perception of complex nonspeech sounds. *Attention, Perception, & Psychophysics*, 71(8), 1900–1915.
- Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences*, 7(6), 246–251.
- Smits, R., Sereno, J., & Jongman, A. (2006). Categorization of sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 32(3), 733–754.
- Spiering, B. J., & Ashby, F. G. (2008). Response processes in information-integration category learning. *Neurobiology of Learning and Memory*, 90(2), 330–338.
- Tunney, R. J., & Fernie, G. (2012). Episodic and prototype models of category learning. *Cognitive Processing*, 13(1), 41–54.
- Verbeemen, T., Vanpaemel, W., Pattyn, S., Storms, G., & Verguts, T. (2007). Beyond exemplars and prototypes as memory representations of natural concepts: A clustering approach. *Journal of Memory and Language*, 56(4), 537–554.
- Yamauchi, T., Love, B. C., & Markman, A. B. (2002). Learning nonlinearity separable categories by inference and classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 585–593.
- Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands. *Journal of the Acoustical Society of America*, 33(2), 248–248.