Re-examining selective adaptation: Fatiguing feature detectors, or distributional learning?

Dave F. Kleinschmidt¹ · T. Florian Jaeger²

Published online: 5 October 2015
© Psychonomic Society, Inc. 2015

Abstract When a listener hears many good examples of a /b/ in a row, they are less likely to classify other sounds on, e.g., a /b/-to-/d/ continuum as /b/. This phenomenon is known as *selective adaptation* and is a well-studied property of speech perception. Traditionally, selective adaptation is seen as a *mechanistic* property of the speech perception system, and attributed to fatigue in acoustic-phonetic feature detectors. However, recent developments in our understanding of non-linguistic sensory adaptation and higher-level adaptive plasticity in speech perception and language comprehension suggest that it is time to re-visit the phenomenon of selective adaptation. We argue that selective adaptation is better thought of as a *computational* property of the speech perception system. Drawing on a common thread in recent work on both non-linguistic sensory adaptation and plasticity in language comprehension, we furthermore propose that selective adaptation can be seen as a consequence of *distributional learning* across multiple levels of representation. This proposal opens up new questions for research on selective adaptation itself, and also suggests that selective adaptation can be an important bridge between work on adaptation in low-level sensory systems and the complicated plasticity of the adult language comprehension system.

Keywords Speech perception · Computational models · Statistical inference · Perceptual learning

Repeated exposure to a prototypical example of a phonological category seems to ‘shrink’ that category. For example, exposure to a reasonably good example of a /b/ many times in a row makes listeners less likely to classify other sounds on a /b/-/d/ continuum as /b/ (Eimas & Corbit, 1973). The discovery of this effect, known as *selective adaptation*, led to a great deal of excitement among speech perception researchers. By analogy with similar adaptation effects in visual psychophysics, it promised a “psychophysicist’s microelectrode” (Frisby, 1979; Mollon, 1974), a powerful and general method of probing the neural code by which continuous acoustic input is transformed into discrete linguistic units, and hence potentially resolving fundamental debates about the psychological reality and properties of phonemes and phonetic features hypothesized by formal linguistic analysis. Selective adaptation effects in visual perception, for instance, were critical in establishing—based on behavior alone—the features used by the visual system to analyze visual input (e.g., Blakemore & Campbell, 1969), and it was hoped that the same sort of effects in speech perception would be able to establish the existence of feature detectors for abstract linguistic categories like phonemes and phonetic features like *voicing*. However, this excitement was quickly tempered for a variety of reasons. Chief among them were findings suggesting that selective adaptation reflects something more like fatigue of general sensory feature detectors, rather than *language-specific* processing of speech sounds (e.g., Remez, 1979; Samuel & Newport, 1979; Schouten, 1980; Roberts & Summerfield, 1981). More deeply problematic, the patterns of adaptation observed were not consistent with fatigue of the *kind* of

✉ Dave F. Kleinschmidt
dave.f.kleinschmidt@gmail.com

¹ Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY, USA

² Departments of Brain and Cognitive Sciences, Computer Science, and Linguistics, University of Rochester, Rochester, NY, USA

feature detectors that had been assumed (see Remez, 1987 for an excellent critique). It became clear that adaptation effects needed to be explained as phenomena in their own right, and subsequently selective adaptation fell out of favor in work on speech perception (cf. Remez, 1987; Samuel, 1986).

In the decades since the rapid rise and fall of selective adaptation, we have learned a great deal about adaptive properties of sensory systems generally, and speech perception in particular. We argue that taken together, these bodies of work suggest that it is time to re-examine the phenomenon selective adaptation. Two developments contribute to our argument. First, recent work challenges the idea that sensory adaptation is best thought of as the fatigue of feedforward feature detectors, or more broadly as a *mechanistic* property of sensory systems. Rather, an emerging perspective views adaptation as a *computational* property of sensory systems, where the sensitivity of the sensory system (with its limited neural resources) is re-aligned to information that is relevant to the task at hand and meaningful in the current environment (e.g., Gutnisky & Dragoi, 2008; Kohn, 2007; Stocker & Simoncelli, 2006; Webster et al., 2005). While in some cases this re-alignment might be achieved by detector fatigue, there is a growing body of work that cannot easily be accounted for in this way (which we will discuss below). Second, the last decade of speech perception research has revealed that, far from a series of fixed feature detectors, the speech perception system is adaptive and flexible in smart ways (e.g., Bertelson et al., 2003; Bradlow & Bent, 2008; Clayards et al., 2008; Kraljic et al., 2008; Maye et al., 2008; Norris et al., 2003). Plasticity and context-sensitivity are increasingly understood to be central features of a speech perception system that has to function in a world where the cue-to-category mapping depends on aspects of the local environment, like *who* is talking (Norris et al., 2003; Huang & Holt, 2012; Laing et al., 2012).

These developments, we argue, undermine two of the basic assumptions of early work on phonetic selective adaptation: that speech perception is (1) performed by acoustic-phonetic feature detectors that (2) become fatigued if they are stimulated too often. In this paper, we focus on the fatigue assumption (see Remez, 1987 for a rejection of the feature detector assumption). In part, we focus on fatigue because contemporary speech perception research views selective adaptation as a fatigue effect that is qualitatively different from other sorts of plasticity (e.g., Grabski et al., 2013; Vroomen et al., 2004; 2007; van der Zande et al., 2014; Zäske et al., 2013). More importantly, we focus on fatigue here because the developments mentioned above do more than just undermine the fatigue assumption. Additionally, we propose, they provide a productive lens through which to re-examine the phenomenon of selective adaptation, and speech perception more generally.

A common thread that runs through recent work on sensory adaptation and flexibility in language comprehension is the idea of *distributional learning across multiple levels*. The sensory information that is relevant or meaningful in the current environment depends on the distribution of sensory features in that environment. Sensory systems adapt to these properties, and to that extent they can be thought of as implicitly learning those distributions (Brenner et al., 2000; Fairhall et al., 2001; Gutnisky & Dragoi, 2008; Sharpee et al., 2006; Stocker & Simoncelli, 2006). At the same time, building on recent work that treats speech perception as statistical inference (e.g., Clayards et al., 2008; Feldman et al., 2009; Kleinschmidt & Jaeger, 2015b; Norris & McQueen, 2008; Sonderegger & Yu, 2010), much of the flexibility of speech perception—and language comprehension more broadly—can be seen as a form of distributional learning. In this view, listeners deal with talker-to-talkers variability by updating their beliefs about the distributions of cues (acoustic, phonetic, lexical, etc.) that result when a talker produces a linguistic structure (phonetic category, word, syntactic structure, pragmatic intention, etc.; Fine et al., 2013; Guediche et al., 2014; Holt, 2006; Idemaru & Holt, 2011; Kleinschmidt & Jaeger, 2015b; Yildirim et al., *in press*).

First, we briefly review the evidence that sensory adaptation in general is better viewed as a form of distributional learning, rather than the fatigue of static feature detectors. Second, we lay out a tentative proposal for how the logic of distributional learning applies to phonetic selective adaptation. Third, we show that much of the literature on selective adaptation which appears to contradict our proposal is actually *predicted* by it, given a modern understanding of distributional learning that occurs at multiple levels of representation. Fourth, and finally, we lay out what we see as some of the most pressing questions that this proposal raises for future work.

Our goal is not to show conclusively that selective adaptation *is*, in all cases, distributional learning. Rather, we hope that distributional learning provides a productive perspective for re-examining previous evidence, and renews interest in selective adaptation as an important phenomenon in need of explanation in its own right, rather than simply a methodological tool.

Sensory adaptation as distributional learning

Non-linguistic sensory adaptation (which usually means visual adaptation) has much in common with phonetic selective adaptation: repeated exposure to a stimulus attribute will reduce responsiveness to that attribute, both at the level of behavior (psychophysics) and neural responses (electrophysiology). We refer to this decrease in responsiveness as

a *negative after-effect*. For example, after extended viewing of a vertical grating, the contrast threshold for detecting another vertical grating goes up, and the firing rate of neurons in the visual cortex which are sensitive to vertical orientation is reduced when viewing another vertical grating (cf. Kohn, 2007). These findings were originally explained as fatigue of neuronal populations which are primarily sensitive to the adapted feature (Blakemore & Campbell, 1969), and when analogous negative after-effects of selective adaptation with speech were first discovered (Eimas & Corbit, 1973), they too were chalked up to a similar mechanism (although not without debate, some of which we revisit below; e.g., Cole et al., 1975; Ainsworth, 1977).

However, work in the decades since challenges the idea that adaptation is broadly the result of “dumb” fatigue of feature detectors (Kohn, 2007; Webster et al., 2005). We now know that adaptation does not just reduce the firing of neurons overall but actually *changes their tuning*, such that some stimuli might even elicit *higher* firing rates after adaptation (Dragoi et al., 2000; Gutfreund, 2012; Kohn & Movshon, 2004). Moreover, these changes in tuning have been shown to be coordinated across *populations* of neurons, and often in ways that increase the amount of information that is transmitted about stimuli in the local environment (Brenner et al., 2000; Fairhall et al., 2001; Gutfreund & Dragoi, 2008; Sharpee et al., 2006). Specifically, changes in tuning seem to serve an increase in the perceptual resolution for the range and distribution of stimuli observed in the current environment (e.g., in an experiment).

This and related findings suggest that selective adaptation has to be understood at least in part as a computational property of sensory systems that need to use limited neural resources to efficiently represent information across a variety of environments (Brenner et al., 2000; Fairhall et al., 2001; Gutfreund & Dragoi, 2008; Kohn, 2007; Sharpee et al., 2006; Wainwright, 1999; Webster et al., 2005). In order to allocate processing resources, sensory systems need to know, at a basic level, how the features they represent are statistically distributed in the current environment. The sensory information that is most *likely* and *relevant* changes from one environment to the next, and so sensory systems need to adapt to these changes in order to make the best use of their limited neural resources. One way to distinguish more stimuli using a fixed number of neurons is to change the stimulus features encoded by each neuron’s signal based on the current environment, because the environment itself provides some clues as to which stimuli are most likely and how they ought to be interpreted. Viewing adaptation purely as a mechanistic property like feature detector fatigue misses this important point.

How does this computational perspective help to understand the effects observed in a typical adaptation experiment? Consider the extreme deviation such an experiment

presents from everyday perception. The perceptual environments we encounter outside the lab tend to exhibit a lot of variability in stimulus features. In a typical adaptation experiment, however, the exact same stimulus is presented over and over again. This produces a highly concentrated, *non-variable* distribution. The typical behavioral and neural after-effects of adaptation can be attributed to the retuning of sensory systems to this new distribution, by, for instance, reducing noise or increasing selectivity to make finer-grained distinctions in the more narrow range of stimuli observed in the experiment (Gardner et al., 2004; Stocker & Simoncelli, 2006; Wainwright, 1999), or reducing sensitivity to highly predictable stimuli (Friston, 2005; Rao & Ballard, 1999). In order to re-tune itself based on the distribution of sensory features, the sensory system has to, at least implicitly, *learn* those distributions. In this sense, selective adaptation can be thought of as distributional learning.

Research on selective adaptation in the visual domain provides two pieces of evidence in support of this view that we discuss next. Both of these phenomena reveal that it is critical to consider distributional learning at *multiple levels*: short-, medium-, and long-term distributions in the first case, and different levels of features (part vs. whole) in the second.

First, to the extent that adaptation is a consequence of distributional learning, stronger adaptation is predicted to occur when the current distribution deviates more from the expected distribution, and more learning is required. Chopin and Mamassian (2012) showed that, as predicted by distributional learning, in an (experimental) visual world where right-leaning edges are more common overall, seeing a series of right-leaning bars produces less of a negative after-effect, compared to a world where left-leaning edges are more common. Subjects were asked to say whether a bar was leaning to the left or the right, in a sequence of trials that contained vertical, left-leaning, and right-leaning bars. Critically, Chopin and Mamassian (2012) manipulated the short-term and long-term frequency of left- vs. right-leaning trials before each of the vertical test trials. When there were more right-leaning bars in the short-term history (last 3 min), subjects showed the typical negative after-effect, and were more likely to say that a vertical bar was leaning *left*, but when there were more right-leaning bars in the *long-term* history (from 2 to 10 min in the past), subjects showed a *positive* after-effect, and were biased to say a vertical bar was right-leaning. This effect runs contrary to the basic prediction of the neural fatigue theory of adaptation: more exposure to right-leaning stimuli should *always* result in fewer right-leaning responses, regardless of the long-term statistics of the current environment.

Second, high-level features can change the expected distribution of lower-level features, and distributional learning predicts that such a change should lead to more or less

Fig. 1 Adaptor and test stimuli used by He et al. (2012). The adaptor stimulus was a *diamond shape* (a) that rotated behind *three occluder bars* (b). In the high-level condition (c), the lines indicating the edges of the occluders matched the portion actually occluded, leading to a percept of a *moving diamond shape*. In the low-level condition (d), the lines were moved, leading to a percept of four disconnected edges. The shape after-effect was tested by means of a tall-versus-wide diamond task (e), while adaptation to the edges themselves was tested via

a tilt after-effect, measured with a left-versus-right leaning bar task (f). Adaptation to the whole shape (c) produced strong shape after-effects (e) but much weaker tilt after-effects (f) than the non-fused edges (d). Reprinted from Current Biology, 22, He, D., Kersten, D., and Fang, F., Opposite Modulation of High- and Low-Level Visual Aftereffects by Perceptual Grouping, 1041, Copyright (2012), with permission from Elsevier.

adaptation to those lower-level features. For instance, the presence of a diamond shape predicts the presence of oriented bars corresponding to its edges. Sensory adaptation indeed appears to be sensitive to this dependence as well: adaptation to a higher-level feature—like a shape—reduces or eliminates the after-effects produced by its component features—like individual edges—when presented in isolation (He et al., 2012). By manipulating perceptual grouping of individual oriented bars, He et al. (2012) changed whether they were perceived as edges of a single figure (a diamond, Fig. 1, c) or as separate objects (four bars, Fig. 1, d). When the bars were perceived as separate, each feature individually produced the typical after-effect. However, when all of the features were perceptually grouped into a diamond, the after-effects of the individual features were substantially reduced. Instead, a higher-level *shape* after-effect was obtained. Again, this effect runs contrary to a fatigue theory of adaptation. Subjects were exposed to exactly the same low-level features in both cases (edges

with the same orientation). Thus, the corresponding edge detectors should be subject to the same fatigue, but only when there is no higher-level feature that explains the oriented edges is there any adaptation to the oriented edges themselves.

In both of these cases, attributing selective adaptation to fatigue fails to explain the observed effects. Distributional learning, however, offers a possible explanation. In a distributional learning theory of adaptation, negative after-effects happen because the sensory system has adjusted to an unusual distribution of stimuli. The degree of adjustment, and hence the strength of the negative after-effects, is determined by just how unusual that stimulus distribution is, compared to what is typical or expected. Critically, a stimulus distribution that would be highly unusual on its own might be perfectly typical if it is similar to the distributions encountered in medium-to-long-term history (e.g., as in Chopin and Mamassian, 2012) or predictable given the presence of higher-level features (e.g., as in He

Fig. 2 Distributional learning accounts for the negative after-effects observed in phonetic selective adaptation. Repeated exposure to a prototypical /b/ results in a distribution with much lower variance than normal (*left*). Learning this distribution means that previously ambiguous stimuli are now, to the listener, more likely to be occur if the

intended category was /d/, resulting in a negative shift in the /b/-/d/ category boundary (*right*). Incremental belief updating predicts that the more exposure to this distribution the listener has, the more their beliefs should reflect it (*left* vs. *middle*) and the stronger the after-effect (*right*, *dashed* vs. *solid*)

et al., 2012). To a first approximation, a distributional learning theory thus correctly predicts that such *predictable* or *expected* deviations in sensory statistics should produce weaker after-effects than unpredictable deviations.¹

A growing body of evidence thus suggests that non-linguistic sensory adaptation is intimately tied to the distribution of sensory features in the current environment. Rather than inducing the same adaptation regardless of the environment, the effect of repeated exposure to low-level features depends on higher-level statistics of the environment. This work shows that the distribution of sensory features in isolation are not enough for understanding adaptation. Rather, it is necessary to consider sensory feature distributions across *multiple levels*, both in the sense of part-whole feature relationships (He et al., 2012) and in the sense of short-, medium-, and long-run distributions of a single feature (Chopin & Mamassian, 2012). Next, we review how the same distributional learning logic can be applied to selective adaptation in speech perception.

Phonetic selective adaptation as distributional learning

At the most basic level, the distribution of acoustic cues determines the mapping from acoustic cues to the underlying phonetic categories. Moreover, these distributions can vary quite a bit from one talker to the next. This variability means that in order to robustly recognize phonetic categories, the speech perception system must, in some way, be sensitive to changes in these statistics (Kleinschmidt & Jaeger, 2015b; McMurray & Jongman,

2011). One way of achieving this sensitivity is by distributional learning (for a review of additional means, such as normalization, see Weatherholtz & Jaeger, 2015). Indeed, research on speech perception has accumulated an impressive body of evidence that adult listeners do engage in distributional learning during speech perception, both at the level of cues (e.g., recalibration/perceptual learning, Bertelson et al., 2003; Clayards et al., 2008; Feldman et al., 2013; Kraljic & Samuel, 2007; Munson, 2011; Norris et al., 2003) and at the level of phonetic categories themselves (e.g. learning novel phonotactic constraints, Dell & Warker, 2004; Warker & Dell, 2006; Warker et al., 2009). Recent work further suggests that adults also engage in distributional learning at higher levels of linguistic representation, like syntactic structures (Chang et al., 2006; Fine et al., 2013; Jaeger & Snider, 2013; Kamide, 2012) and pragmatics (Kurumada et al., 2014, 2012; Yildirim et al., *in press*). As with general sensory adaptation reviewed above, this learning seems to be a form of “smart” plasticity that increases the speed and accuracy of language processing in the face of uncertainty and variability across environments (for review, see Kleinschmidt & Jaeger, 2015b).

Distributional learning in speech perception is a prerequisite for a distributional learning account of selective adaptation, but how does the logic of distributional learning play out for selective adaptation to phonetic input? Intuitively, the logic is the same as for adaptation in the visual domain, discussed above. Consider a typical phonetic adaptation experiment, where a listener hears a prototypical /b/ repeated over and over again. This leads to a very odd distribution of acoustic cues: typically, there is some variability in the signal, even when the same category is repeated by the same talker in the same phonological context (Allen et al., 2003; Newman et al., 2001). Such a highly concentrated distribution might, intuitively, correspond to a talker who is *very* precise in how they produce their /b/s. Such a talker is less likely to produce a /b/ with a cue value that, for a normal talker, would be ambiguous between a /b/ and a /d/. Consequently, if the listener has learned this changed

¹An important caveat is that if the statistics deviate *too* much from what is expected, then learning may not be possible at all. One notable example from speech perception is that listeners do not adapt fully to highly unusual accents (e.g., Idemaru & Holt, 2011; Sumner, 2011; see also, Kleinschmidt & Jaeger, 2015a).

distribution, then we would predict that they classify fewer items by the same talker on a /b/-to-/d/ continuum as /b/ (Fig. 2).

This reduction in /b/ responses after /b/ exposure is exactly the negative after-effect (shift in category boundary) that characterizes selective adaptation in speech (Eimas & Corbit, 1973; Samuel, 1986). Thus, qualitatively, distributional learning provides an account for selective adaptation, at least as far as its effects on classification are concerned. Quantitatively, too, distributional learning provides a good fit against data from selective adaptation: the incremental effects of repeated exposure to the same stimulus are well described by a distribution learning model ($r^2 = 0.85$, Kleinschmidt & Jaeger, 2015b).

The “ideal adapter” distributional learning model developed in Kleinschmidt and Jaeger (2015b) predicts that listeners do not just track the variance of cue distributions, but the *whole* distribution, which includes at least both the *means* and variances of cue distributions. The same model can be applied to phonetic recalibration/perceptual learning, where listeners hear an *atypical* production of /b/ which is acoustically ambiguous between /b/ and /d/, but which is disambiguated by, e.g., a visual cue or lexical context. The resulting cue distribution is unusual both in terms of its mean and its variance. Recalibration is characterized by a *positive* after-effect: after such exposure, listeners classify *more* items from a /b/-/d/ continuum as /b/. Recalibration, unlike selective adaptation, has traditionally been described as a form of implicit learning (Norris et al., 2003; Vroomen et al., 2004, 2007). Nevertheless, the very same distributional learning model fits listeners’ classification behavior in a recalibration experiment just as well as for selective adaptation ($r^2 = 0.86$, Kleinschmidt & Jaeger, 2015b), using an identical set of parameters for both recalibration and selective adaptation. According to the ideal adapter model, the positive after-effect elicited by recalibration is typically due to the unusual *mean* of the cue values that listeners hear. By learning that the experimental talker produces their /b/ with a mean cue value that is more like /d/, listeners can infer that /d/-like cue values are more likely to be produced by this talker when they are saying /b/, resulting in a positive after-effect.

The ideal adapter makes a further, less intuitive prediction that has only recently been recognized. The distribution of cues in a typical recalibration experiment does not only have an unexpected mean; it also has unexpectedly low variance. This later property is shared with selective adaptation experiments. Distributional learning predicts that longer-term exposure to a distribution with lower-than-expected variance should produce a *negative* after-effect. That means that the positive after-effect that is observed in perceptual recalibration should eventually be undone or perhaps even reversed. Quantitative simulations of an ideal adapter

corroborate this intuitive prediction: with enough exposure to the same stimuli, the positive after-effects of recalibration can be canceled out, or even reversed (Kleinschmidt & Jaeger, 2015b, pp. 164–6). Moreover, this is exactly what Vroomen et al. (2007) found for human listeners: exposing listeners to 256 repetitions of an ambiguous /b/ lead to initial positive after-effects that were later canceled out. In recent work, we have replicated this effect (Kleinschmidt and Jaeger, 2012). In the same experiments, we extended exposure to cue distributions with a range of means, from fully prototypical (as in selective adaptation) to fully ambiguous (as in recalibration) and a number of steps in between. As predicted by a distributional learning account, the observed after-effects formed a continuum: initially, exposure to cue distributions with shifted means lead to positive after-effects; with increasing exposure, however, these effects were overcome, undoing the positive after-effects and eventually even reversing them into negative after-effects. An ideal adapter model fits all of these exposure distributions with a single set of parameters, suggesting that distributional learning can account for the whole range of exposure conditions, including selective adaptation (Kleinschmidt & Jaeger, 2015b).

Finally, we note that the ideal adapter model is only one example of a broad class of models that, implicitly or explicitly, incorporate distributional learning. For instance, episodic and exemplar models (e.g., Goldinger, 1998; Johnson, 1997) can be seen as implementing a form of implicit distributional learning (as kernel density estimation, cf. Sanborn et al., 2010; or importance sampling, Shi et al., 2010). Distributional models of phonetic category *acquisition* (e.g., Feldman et al., 2013; McMurray et al., 2009; Vallabha et al., 2007), if extended to adult phonetic adaptation data, would make similar predictions, possibly even using the same learning rates (Toscano et al., *submitted*). Finally, other models that treat speech perception as a process of inference under uncertainty implicitly assume sensitivity to the underlying distributions (Clayards et al., 2008; Feldman et al., 2009; Norris & McQueen, 2008; Sonderegger & Yu, 2010), even though they do not explicitly include distributional learning *per se*. While members of this broad class of models predict sensitivity to cue distributions, only the ideal adapter has been explicitly applied to selective adaptation data (Kleinschmidt & Jaeger, 2015b). It thus remains to be seen how well this *class* of models (rather than the ideal adapter model itself) might account for selective adaptation.

Challenges to this view

In summary, evidence from both speech perception and sensory adaptation in non-linguistic domains suggests that selective adaptation is better understood as distributional

learning by sensory systems that are constantly adapting to changes in the statistical properties of the sensory world. However, the literature on phonetic selective adaptation presents a number of immediate challenges to this view. In this section we review some of the most striking, and argue that in large part they are actually *predicted* by a distributional learning account.

Similar accounts, previously rejected

Several previous proposals have treated selective adaptation in ways that bear resemblance to the current proposal. These accounts are widely considered to be rejected by existing evidence. As we lay out below, these previous accounts differ from the current proposal in critical ways. We argue that, in fact, the evidence that rejects them is entirely consistent with the sort of distributional learning we are proposing.

Re-tuning, or changing category mean

Ainsworth (1977) investigated whether selective adaptation could be due to adaptation to changes in category means, a hypothesis he referred to as “re-tuning” of feature detectors. Ainsworth rejected this account because adaptation with sounds that are closer to the category boundary still produces a negative after-effect (Ainsworth, 1977). As Ainsworth pointed out, such changes cannot be accounted for by changes to only category means, which would produce a *positive* after-effect as in recalibration discussed above.

However, unlike Ainsworth’s (1977) proposal, the ideal adapter account we described in the previous section assumes that listeners implicitly learn not *only* the category means, but rather the *whole* distribution (or at least both the mean and variance). As discussed in the previous section, such an ideal adapter account *can* capture the negative after-effect elicited by a typical selective adaptation paradigm, both qualitatively and quantitatively (Kleinschmidt and Jaeger, 2015b). More specifically, it predicts that long-term exposure to the same stimulus will, in many cases, elicit a negative after-effect, even when the initial after-effect is positive. This is precisely the pattern that is observed in experiments where listeners are tested at different stages of adaptation (Kleinschmidt & Jaeger, 2015b; see also Vroomen et al., 2007 and the re-analysis of Samuel, 2001 therein).

These findings suggest an alternative explanation for the findings reported by Ainsworth (1977). As is typical for selective adaptation experiments, Ainsworth employed a paradigm in which a large number of adapting trials preceded each test trial. It is thus likely that Ainsworth effectively tested long-term effects of adaptation, in which case the negative after-effects Ainsworth observed are exactly

what the ideal adapter distributional learning account predicts.

Narrowing of selectivity, or changing category variance

In a similar vein, Cole and Cooper (1977) investigated another type of distributional learning account, essentially assuming that listeners adapt only to changes in *variance*. Cole & Cooper referred to this as a “narrowing of selectivity”. Based on data from several perception experiments, shown in Fig. 3, Cole & Cooper concluded that narrowing of selectivity alone is not sufficient to explain selective adaptation. If selective adaptation was caused by only adaptation to changes in category variance, then exposure to a range of adaptor stimuli with normal levels of variance should produce no selective adaptation at all. However, Cole and Cooper (1977) found that their “variable” adaptation condition (Fig. 3, squares) produced a negative after-effect just as large as a non-variable condition with a single, intermediate adaptor (Fig. 3, upward pointing triangles).

While this result cannot be explained in terms of adaptation to changes in variance alone, it is compatible with an ideal adapter. Recall that the ideal adapter predicts that listeners are sensitive to the whole distribution of cues

Fig. 3 Selective adaptation data from Cole and Cooper (1977). The *x*-axis shows the test stimuli on a 1 to 9, /dʒ/ (the “j” sound in *jump*) to /d/ continuum, and the *y* axis shows subjects’ average number of /dʒ/ responses out of 10 repetitions of each test stimulus. The “variable adapter” condition was exposure to a uniform distribution of stimuli, from *x* = 1 to 6. Reprinted with permission from Properties of friction analyzers for [j] Cole, Ronald A. and Cooper, William E., The Journal of the Acoustical Society of America, 62, 177-182 (1977). Copyright 1977, Acoustic Society of America.

(including the *mean*), not just the variance alone. All else being equal, distributional learning predicts that a high-variance adaptor distribution will produce a smaller negative after-effect than a low-variance distribution, but all else is not equal in Cole and Cooper (1977). The intermediate /dʒ/ adaptor ($x = 5$) was substantially farther from the /dʒ/ endpoint ($x = 1$) than the average of the variable-/dʒ/ condition ($\bar{x} = 3.5$, a difference of 2.5 from the endpoint, vs. a difference of 4 for the intermediate adaptor).² An adaptor closer to the category prototype is, in a distributional learning model, predicted to produce a *stronger* negative after-effect. This would mitigate the effect of the higher variance in the range condition, which could explain why the intermediate and range conditions produce comparable effects. In fact, the ideal adapter model qualitatively predicts that the low-variance, intermediate adaptor should lead to a slightly steeper category boundary, which is in fact what Cole and Cooper (1977) appear to have found (see Fig. 3, $x = 5, 6, 7$).

Selective adaptation to audiovisual adaptors

One prediction of a distributional learning theory of selective adaptation is that adaptation should depend on the joint distribution of all relevant cues, including visual speech cues. However, studies using audiovisual adaptors (Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994) have been argued to support the view that selective adaptation is due to the fatigue of specifically *auditory* feature detectors. These audio-visual adaptors had large, categorical mismatches, which are intended to produce a McGurk Effect percept, where the perceived category of the audio-visual stimulus is different than the auditory component alone (McGurk & MacDonald, 1976). Saldaña and Rosenblum (1994) used an audio-/ba/, visual-/va/ adaptor stimulus which was consistently identified as /va/ by participants. This adaptor produced an after-effect on a /ba/-/va/ continuum the same size as when the audio /ba/ component was presented separately. Saldaña and Rosenblum (1994) interpreted this after-effect as evidence that listeners adapted to the audio component alone, rather than the integrated percept of /va/.³ As a result, Saldaña and Rosenblum (1994)

rejected explanations of selective adaptation as learning at the level of phonetic categories—as in distributional learning—in favor of a process of purely low-level acoustic feature detector fatigue.

However, it is not possible to tell whether the observed effect was due to selective adaptation of /b/ (lower /b/ variance), or *recalibration* of /v/ (shift in /v/ mean), since both would produce a shift in the category boundary towards /b/. Unless they completely ignored the audio component, subjects in this experiment likely perceived the audio-visual adaptor not as a perfect /va/, but rather a somewhat /ba/-like /va/. Kleinschmidt and Jaeger (2011) presented evidence that phonetic distributional learning operates on this kind of combined audio-visual percept (also cf. Bejjanki et al., 2011; Ernst & Bühlhoff, 2004). If this is in fact the case, then listeners in Saldaña & Rosenblum's experiment should update their beliefs about the /v/ cue distribution, shifting it to be more /b/ like. The result of this recalibration of /v/ would be more /v/ and hence fewer /b/ responses, which is the after-effect that Saldaña and Rosenblum (1994) observed. Thus, the results of Saldaña and Rosenblum (1994) are not *necessarily* incompatible with distributional learning at the phonetic category level.

We believe that future work addressing this question will prove particularly valuable in evaluating the feasibility of an explanation of selective adaptation purely in terms of distributional learning. For instance, we have suggested that the after-effects observed by Saldaña and Rosenblum (1994) are due to *recalibration* to an integrated percept that combines audio and visual cues. This account predicts that changing the reliability of the *visual* cue should change the integrated percept (as observed by Bejjanki et al., 2011; Ernst & Bühlhoff, 2004), and thus the resulting after-effect, while the audio-only selective adaptation account offered by Saldaña and Rosenblum (1994) predicts that changing the visual cue should have no effect at all.

Discussion

Since selective adaptation for speech was first discovered, our understanding of both adaptation in sensory systems in general, and flexibility in the language system in particular has come a long way. First, work on non-linguistic sensory adaptation has established that adaptation goes far beyond fatigue of feedforward feature detectors (Brenner et al., 2000; Dragoi et al., 2000; Fairhall et al., 2001; Gutfreund, 2012; Kohn & Movshon, 2004; Kohn, 2007; Sharpee et al., 2006; Webster et al., 2005). In particular, there is tentative evidence that sensory adaptation serves to increase the efficiency of the representation of sensory information, based on the statistical or distributional properties of recent sensory input. Second, the speech perception system is now

²The same argument holds when using frication duration—the actual physical dimension manipulated to construct the continuum—instead of continuum step, with one caveat: due to an error in constructing the stimuli (Cole & Cooper, 1977, Footnote 1), steps $x = 4$ and $x = 5$ both had frication durations of 32 ms, which means that the variability of the range condition was actually even lower than intended.

³An earlier study (Roberts & Summerfield, 1981) used an audio-/ba/, visual-/ga/ adaptor. This was intended to produce a percept of /da/ (as in McGurk & MacDonald, 1976), but only half of their subjects perceived an alveolar sound at all, with the others reporting /kl/, /m/, or /fl/ (Saldaña & Rosenblum, 1994). This makes it difficult to interpret their results and so we do not discuss them further here.

understood to be flexible in smart ways, for example, by recalibrating phonetic categories based on unusual pronunciations (Bertelson et al., 2003; Kraljic & Samuel, 2005; Norris et al., 2003).

Both of these developments are rooted in a growing awareness of the importance of distributional or statistical properties of the sensory world for understanding sensory systems. While these two developments may appear rather dissimilar—one concerned with the allocation of scarce neural resources, and the other with robust language comprehension in the presence of variability—they come together in the phenomenon of phonetic selective adaptation. Although for different reasons than originally anticipated, selective adaptation thus rightly deserves the spotlight it enjoyed in the early era of cognitive research on speech perception. Moreover, we have tentatively proposed that selective adaptation can be understood as one consequence of the more general process of distributional learning, along with phonetic recalibration and other forms of smart adaptation in language processing (Kleinschmidt & Jaeger, 2015b). This proposal provides good coverage of the data that exists on how selective adaptation changes listeners' classification of speech sounds.

More importantly, however, we hope that this proposal will provide a road map for how specifically to go about re-evaluating selective adaptation. There is a great deal of work that is required to flesh out and critically evaluate such a distributional-learning theory. In closing, we discuss four important directions for future research.

Future directions

First, distributional learning is a computational-level account, in the sense of Marr (1982). Such accounts focus on the in-principle constraints on a cognitive system that come from the information that is available and the task the system is carrying out. While these considerations guide and constrain cognitive models, they are not cognitive models per se. There are many different possible algorithms, processes, or representations that can carry out the *computation* of distributional learning, and many possible neural implementations for each of those. Fleshing out the relationships between these is an important next step in formulating and evaluating a complete distributional learning theory of selective adaptation. Critically, mechanistic- and computational-level explanations are not necessarily in opposition. It may well be the case, for instance, that a population of fatiguing feature detectors can approximate distributional learning in some circumstances. These links are, however, largely un-explored, and call for future work.

Second, and relatedly, we have focused on the effects of selective adaptation on listeners' *classification* decisions, and have not discussed reaction time effects of selective

adaptation. Such effects are well documented (Samuel, 1986; Samuel and Kat, 1996) and historically one of the most important sources of insight into selective adaptation. Across the cognitive sciences, the *outcome* of a decision can often be dissociated from the *time course* of that decision (see, e.g., the related discussion for lexical selection in speech production; Mahon et al., 2007; Oppenheim et al., 2010). The same is true in phonetic selective adaptation: two adaptors can produce the same boundary shifts, but different changes in reaction time (Samuel & Kat, 1996). A computational-level account like distributional learning naturally captures the decisions that listeners make, but the time course falls more under the purview of process- or implementation-level accounts. Thus, a critical next step in evaluating distributional learning as a theory of adaptation in speech perception (and perception more broadly) is to develop process- and implementation-level models that carry out the computations of distributional learning. There is some promising if preliminary work that links increased probability of a stimulus to an increase in signal-to-noise ratio in sensory representations (Stocker & Simoncelli, 2006; Wei & Stocker, 2012). Increased reaction times might be a consequence of achieving such increased signal-to-noise ratio through, for instance, stronger lateral inhibition leading to sparser responses by effectively narrowing receptive fields (Gardner et al., 2004). We emphasize that this is only one possibility, and that in general linking distributional learning to reaction time data at all is a critical direction for future work.

Third, as we have outlined it here, distributional learning accounts for selective adaptation through changes in listeners' beliefs about the *variance* of the adapted phonetic category. There is some evidence that listeners are, in fact sensitive to category variance (Clayards et al., 2008; Cole & Cooper, 1977; Newman et al., 2001; Schreiber et al., 2013), but only a few studies have explored this systematically. Only one study that we know of has specifically investigated the link between category variance and selective adaptation (Cole & Cooper, 1977), and while, as we have argued, their results are consistent with a distributional learning theory, their design confounds category variance and category *mean*. More work is required, first to determine what distributional learning predicts the effect of different degrees of category variance is, and second to see whether listeners actually behave in the predicted ways.

Fourth, distributional learning provides a bridge between selective adaptation—typically thought of as a low-level process—and higher-level processes like talker-specific recalibration and accent adaptation. A hallmark of these processes is that listeners often apply what they have previously learned to future situations in smart ways. This can mean *recognizing* a previously encountered talker, as

demonstrated by recalibration that persists even after 12 hours outside the lab (Eisner & McQueen, 2006). It can also mean *generalizing* learning to an unfamiliar talker that is similar to previously-encountered talkers, as demonstrated by talker-independent accent adaptation (Bradlow & Bent, 2008). Elsewhere (Kleinschmidt & Jaeger, 2015b), we have proposed that such cross-situational learning effects can be modeled as distributional learning that is indexed to particular indexical variables like a talker, groups of talkers, or environment. Such hierarchical distributional learning enables previously learned cue distributions to be re-learned very quickly when the associated indexical variable is encountered again (for similar arguments applied to domain-general sensory/motor learning, see Qian et al., 2012, 2015). If selective adaptation is due to the same sort of distributional learning, then it follows that listeners should re-adapt to previously encountered distributions more quickly than the initial adaptation. To our knowledge, this prediction has not been addressed in the literature on phonetic selective adaptation.

Whither feature detectors?

We have focused on—and argued against—the assumption that phonetic selective adaptation is the *fatigue* of auditory-phonetic feature detectors, without discussing the nature or existence of these feature detectors themselves. The early work on selective adaptation was rightly criticized for relying on problematic assumptions about these feature detectors (Remez, 1987): that there were a small number of them, specified largely a priori, which respond largely independently of context, and are organized into clearly delineated layers (simple acoustic, complex acoustic, phonetic, etc.). Our proposal that selective adaptation be seen as a form of distributional learning does not rely on these same assumptions. Nevertheless, a fully fleshed out theory of adaptation as distributional learning has to say something about what those distributions are *of*. The things that these distributions cover may be seen as a sort of features, and expressing those distributions in terms of some relatively low-dimensional compression of the full acoustic input information makes sense for the same reasons that distributional learning makes sense: sensory systems have limited representational resources, and care about some aspects of the sensory world more than others. However, there is no guarantee that these “features” will line up with our intuitions about what the relevant features are, or that they will be organized in clearly delineated levels.

There is also no guarantee that adaptation itself provides a straightforward means of probing these features. In fact, the work on visual adaptation we reviewed here suggests that adaptation has to be considered in terms of

distributions across *multiple levels* simultaneously. We take these levels to refer not only to part-whole feature relationships (He et al., 2012), but also to the statistical sense of distributions over the same features but at different time scales (seconds vs. minutes) (Chopin & Mamassian, 2012) or groups of situations (talker, accent, language, etc.). There is analogous work in speech perception, including selective adaptation, that makes the same point (e.g., adaptation of single segments depends on the syllabic context, Bryant, 1978; and others summarized in Remez, 1987), but the consequences of adaptation across multiple levels have not been fully explored, either in vision or speech perception.

Speech as a model organism for perception

Finally, we close by arguing that selective adaptation provides a good bridge between speech perception and the study of perception more generally. As we have discussed, recent work on general sensory adaptation has revealed that adaptation has a variety of functional properties that make it interesting as an object of study in its own right, rather than just a methodological tool. The emerging computational understanding of adaptation puts it squarely at the intersection of many fundamental and interesting questions in the study of perception, such as the interaction between bottom-up and top-down information, the ways that innate constraints and information from the world shape perceptual systems, the role that probabilistic predictions play in perception (e.g., He et al., 2012), and how the brain manages to process sensory information efficiently in a variable world with a limited stock of representational resources, among others.

Speech perception exemplifies many of these issues as well as, or better than, other perceptual processes. As a sensory signal, speech is structured in multiple, intersecting ways. On the one hand, there is the *linguistic* structure of speech: individual sounds combine into phonetic categories, categories into words, words into phrases, phrases into sentences, etc. On the other hand, there is the *indexical* structure of speech: every talker maps phonetic categories to sounds in a different way, often dramatically so, and talkers themselves can be clustered based on dialect, native language background, etc. While neither sort of structure is exhaustively understood, they are understood well enough to manipulate them experimentally in ways that carefully control their statistical properties but are reasonably ecologically valid (Allen & Miller, 2004; Clayards et al., 2008; Newman et al., 2001). This stands in contrast to visual psychophysics, which in large part relies on simple, artificial stimuli in order to achieve careful control. Obviously, much has been learned from these techniques. However, for investigating questions of how sensory systems deal with

changes in the statistics of the sensory world, more leverage might be provided by studying sensory domains where such deviations occur frequently and are typically managed successfully by human observers. Speech perception is just such a domain: differences between talkers are well studied, often to the extent of quantifying the statistical deviations from one talker to another. Likewise, the last decade has seen an explosion in research showing listeners can rapidly adapt to different talkers.

What role does phonetic selective adaptation play in all of this? It sits at the intersection of the study of low-level adaptation in vision and more recent work on smart plasticity in speech perception and language comprehension. It is our hope that, by bringing new attention to the long literature on phonetic selective adaptation, researchers studying speech perception will be more aware of developments in how sensory adaptation more broadly is understood, and that researchers who study perception in general will consider both the potential offered and the challenges posed by speech perception as a perceptual process.

Conclusions

The literature on phonetic selective adaptation is long, complex, and often apparently contradictory (as summarized by Samuel, 1986 and Remez, 1987). Over time, interest in the theoretical basis for this fundamental phenomenon has dwindled. We believe, though, that adaptation is an important property of perception in general, and of speech perception in particular, and that this literature deserves more attention. Recent developments in the understanding of sensory adaptation more broadly and flexibility in speech perception in particular challenge some of the basic assumptions of early work on phonetic selective adaptation. However, as we have argued, they also provide a set of conceptual tools for understanding phonetic selective adaptation in the broader context of language comprehension and sensory adaptation. We have proposed, to start, that selective adaptation can be seen as a form of distributional learning. Distributional learning provides a coherent perspective on the existing literature on phonetic selective adaptation, highlights parallels with non-speech adaptation, and suggests a unifying perspective on flexibility in adult language comprehension. It also, most importantly, raises a number of interesting questions for future work.

Acknowledgments This work was partially funded by a National Science Foundation (NSF) Graduate Research Fellowship and National Institute of Child Health and Human Development (NICHD) F31 HD082893 to Dave F. Kleinschmidt and NICHD R01 HD075797 to T. Florian Jaeger. We thank Sarah Bibyk, Esteban Buz, Kodi Weatherholtz, other members of the HLP Lab, Arthur Samuel, Robert

Remez, Kenneth Paap and one anonymous reviewer for comments on previous versions of this work. The views expressed here are those of the authors and not necessarily those of the funding agencies.

References

- Ainsworth, W. A. (1977). Mechanisms of selective feature adaptation. *Perception & Psychophysics*, 21(4), 365–370. doi:[10.3758/BF03199488](https://doi.org/10.3758/BF03199488)
- Allen, J. S., & Miller, J. L. (2004). Listener sensitivity to individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America*, 115(6), 3171.
- Allen, J. S., Miller, J. L., & DeSteno, D. (2003). Individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America*, 113(1), 544. doi:[10.1121/1.1701898](https://doi.org/10.1121/1.1701898)
- Bejjanki, V. R., Clayards, M., Knill, D. C., & Aslin, R. N. (2011). Cue integration in categorical tasks: Insights from audio-visual speech perception. *PLoS ONE*, 6(5), e19812. doi:[10.1121/1.1528172](https://doi.org/10.1121/1.1528172)
- Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual recalibration of auditory speech identification: a McGurk aftereffect. *Psychological Science*, 14(6), 592–597. doi:[10.1371/journal.pone.0019812](https://doi.org/10.1371/journal.pone.0019812)
- Blakemore, C., & Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *The Journal of Physiology*, 203(1), 237–260. doi:[10.1046/j.0956-7976.2003.psci.1470.x](https://doi.org/10.1046/j.0956-7976.2003.psci.1470.x)
- Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, 106(2), 707–29. doi:[10.1016/j.cognition.2007.04.005](https://doi.org/10.1016/j.cognition.2007.04.005)
- Brenner, N., Bialek, W., & de Ruyter Van Steveninck, R. (2000). Adaptive rescaling maximizes information transmission. *Neuron*, 26(3), 695–702. doi:[10.1016/S0896-6273\(00\)81205-2](https://doi.org/10.1016/S0896-6273(00)81205-2)
- Bryant, J. S. (1978). Feature detection process in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 4(4), 610–620. doi:[10.1037/0096-1523.4.4.610](https://doi.org/10.1037/0096-1523.4.4.610)
- Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, 113(2), 234–72. doi:[10.1037/0033-295X.113.2.234](https://doi.org/10.1037/0033-295X.113.2.234)
- Chopin, A., & Mamassian, P. (2012). Predictive properties of visual adaptation. *Current Biology*, 22(7), 622–6. doi:[10.1016/j.cub.2012.02.021](https://doi.org/10.1016/j.cub.2012.02.021)
- Clayards, M. A., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, 108(3), 804–9. doi:[10.1016/j.cognition.2008.04.004](https://doi.org/10.1016/j.cognition.2008.04.004)
- Cole, R. A., & Cooper, W. E. (1977). Properties of friction analyzers for [j]. *The Journal of the Acoustical Society of America*, 62(1), 177. doi:[10.1121/1.381479](https://doi.org/10.1121/1.381479)
- Cole, R. A., Cooper, W. E., Singer, J., & Allard, F. (1975). Selective adaptation of English consonants using real speech. *Perception & Psychophysics*, 18(3), 227–244. doi:[10.3758/BF03205973](https://doi.org/10.3758/BF03205973)
- Dell, G. S., & Warker, J. A. (2004). The tongue slips into (recently learned) patterns. In Quene, H., & van Heuven, V. (Eds.) *On speech and language: Studies for Sieb G. Nooteboom* (pp. 45–56). Utrecht: Netherlands Graduate School of Linguistics.
- Dragoi, V., Sharma, J., & Sur, M. (2000). Adaptation-induced plasticity of orientation tuning in adult visual cortex. *Neuron*, 28(1), 287–98.
- Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, 4(1), 99–109. doi:[10.1016/0010-0285\(73\)90006-6](https://doi.org/10.1016/0010-0285(73)90006-6)

- Eisner, F., & McQueen, J. M. (2006). Perceptual learning in speech: Stability over time. *The Journal of the Acoustical Society of America*, 119(4), 1950–3. doi:[10.1121/1.2178721](https://doi.org/10.1121/1.2178721)
- Ernst, M. O., & Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, 8(4), 162–9. doi:[10.1016/j.tics.2004.02.002](https://doi.org/10.1016/j.tics.2004.02.002)
- Fairhall, A. L., Lewen, G. D., Bialek, W., & de Ruyter Van Steveninck, R. R. (2001). *Efficiency and ambiguity in an adaptive neural code*. *Nature*, 412(6849), 787–92. doi:[10.1038/35090500](https://doi.org/10.1038/35090500)
- Feldman, N. H., Griffiths, T. L., Goldwater, S., & Morgan, J. L. (2013). A role for the developing lexicon in phonetic category acquisition. *Psychological Review*, 120(4), 751–778. doi:[10.1037/a0034245](https://doi.org/10.1037/a0034245)
- Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review*, 116(4), 752–82. doi:[10.1037/a0017196](https://doi.org/10.1037/a0017196)
- Feldman, N. H., Myers, E. B., White, K. S., Griffiths, T. L., & Morgan, J. L. (2013). Word-level information influences phonetic learning in adults and infants. *Cognition*, 127(3), 427–438.
- Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE*, 8(10), e77661. doi:[10.1371/journal.pone.0077661](https://doi.org/10.1371/journal.pone.0077661)
- Frisby, J. (1979). *Seeing: Illusion*. Oxford: Oxford University Press.
- Friston, K. J. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 360(1456), 815–36. doi:[10.1098/rstb.2005.1622](https://doi.org/10.1098/rstb.2005.1622)
- Gardner, J. L., Tokiyama, S. N., & Lisberger, S. G. (2004). A population decoding framework for motion aftereffects on smooth pursuit eye movements. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 24(41), 9035–48. doi:[10.1523/JNEUROSCI.0337-04.2004](https://doi.org/10.1523/JNEUROSCI.0337-04.2004)
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251–79.
- Grabski, K., Tremblay, P., Gracco, V. L., Girin, L., & Sato, M. (2013). A mediating role of the auditory dorsal pathway in selective adaptation to speech: A state-dependent transcranial magnetic stimulation study. *Brain Research*, 1515, 55–65. doi:[10.1016/j.brainres.2013.03.024](https://doi.org/10.1016/j.brainres.2013.03.024)
- Guediche, S., Blumstein, S. E., Fiez, J. A., & Holt, L. L. (2014). Speech perception under adverse conditions: insights from behavioral, computational, and neuroscience research. *Frontiers in Systems Neuroscience*, 7, 126. doi:[10.3389/fnsys.2013.00126](https://doi.org/10.3389/fnsys.2013.00126)
- Gutfreund, Y. (2012). Stimulus-specific adaptation, habituation and change detection in the gaze control system. *Biological Cybernetics*. doi:[10.1007/s00422-012-0497-3](https://doi.org/10.1007/s00422-012-0497-3)
- Gutnisky, D. A., & Dragoi, V. (2008). Adaptive coding of visual information in neural populations. *Nature*, 452(7184), 220–4. doi:[10.1038/nature06563](https://doi.org/10.1038/nature06563)
- He, D., Kersten, D., & Fang, F. (2012). Opposite modulation of high- and low-level visual aftereffects by perceptual grouping. *Current Biology: CB*, 22(11), 1040–5. doi:[10.1016/j.cub.2012.04.026](https://doi.org/10.1016/j.cub.2012.04.026)
- Holt, L. L. (2006). The mean matters: effects of statistically defined nonspeech spectral distributions on speech categorization. *The Journal of the Acoustical Society of America*, 120(5 Pt 1), 2801–17. doi:[10.1121/1.2354071](https://doi.org/10.1121/1.2354071)
- Huang, J., & Holt, L. L. (2012). Listening for the norm: adaptive coding in speech categorization. *Frontiers in Psychology*, 3, 10. doi:[10.3389/fpsyg.2012.00010](https://doi.org/10.3389/fpsyg.2012.00010)
- Idemaru, K., & Holt, L. L. (2011). Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology: Human Perception and Performance*, 37(6), 1939–56. doi:[10.1037/a0025641](https://doi.org/10.1037/a0025641)
- Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, 127(1), 57–83. doi:[10.1016/j.cognition.2012.10.013](https://doi.org/10.1016/j.cognition.2012.10.013)
- Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In Johnson, & Mullenix (Eds.) *Talker Variability in Speech Processing* (pp. 145–165). San Diego: Academic Press.
- Kamide, Y. (2012). Learning individual talkers' structural preferences. *Cognition*, 124(1), 66–71. doi:[10.1016/j.cognition.2012.03.001](https://doi.org/10.1016/j.cognition.2012.03.001)
- Kleinschmidt, D. F., & Jaeger, T. F. (2011). A Bayesian belief updating model of phonetic recalibration and selective adaptation. In *Proceedings of the 2nd ACL Workshop on Cognitive Modeling and Computational Linguistics*. Stroudsburg, PA: Association for Computational Linguistics. Talk.
- Kleinschmidt, D. F., & Jaeger, T. F. (2012). A continuum of phonetic adaptation: Evaluating an incremental belief-updating model of recalibration and selective adaptation. In Miyake, N., Peebles, D., & Cooper, R. P. (Eds.) *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 605–10). Austin, TX: Cognitive Science Society.
- Kleinschmidt, D. F., & Jaeger, T. F. (2015a). *Inferring listeners' prior beliefs about unfamiliar talkers*. Manuscript submitted for publication. doi:[10.13140/RG.2.1.3803.4405](https://doi.org/10.13140/RG.2.1.3803.4405)
- Kleinschmidt, D. F., & Jaeger, T. F. (2015b). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, 122(2). doi:[10.1037/a0038695](https://doi.org/10.1037/a0038695)
- Kohn, A. (2007). Visual adaptation: physiology, mechanisms, and functional benefits. *Journal of Neurophysiology*, 97(5), 3155–64. doi:[10.1152/jn.00086.2007](https://doi.org/10.1152/jn.00086.2007)
- Kohn, A., & Movshon, J. A. (2004). Adaptation changes the direction tuning of macaque MT neurons. *Nature Neuroscience*, 7(7), 764–72. doi:[10.1038/nn1267](https://doi.org/10.1038/nn1267)
- Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, 51(2), 141–78. doi:[10.1016/j.cogpsych.2005.05.001](https://doi.org/10.1016/j.cogpsych.2005.05.001)
- Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, 56(1), 1–15. doi:[10.1016/j.jml.2006.07.010](https://doi.org/10.1016/j.jml.2006.07.010)
- Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First impressions and last resorts: how listeners adjust to speaker variability. *Psychological Science*, 19(4), 332–8. doi:[10.1111/j.1467-9280.2008.02090.x](https://doi.org/10.1111/j.1467-9280.2008.02090.x)
- Kurumada, C., Brown, M., Bibyk, S., Pontillo, F., & Tanenhaus, M. K. (2014). Rapid adaptation in online pragmatic interpretation of contrastive prosody. In Bello, P., Guarini, M., McShane, M., & Scassellati, B. (Eds.) *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 791–796). Austin, TX: Cognitive Science Society.
- Kurumada, C., Brown, M., & Tanenhaus, M. K. (2012). Pragmatic interpretation of contrastive prosody : It looks like speech adaptation. In Miyake, N., Peebles, D., & Cooper, R.P. (Eds.) *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 647–652). Austin, TX: Cognitive Science Society.
- Laing, E. J. C., Liu, R., Lotto, A. J., & Holt, L. L. (2012). Tuned with a tune: Talker normalization via general auditory processes. *Frontiers in Psychology*, 3, 203. doi:[10.3389/fpsyg.2012.00203](https://doi.org/10.3389/fpsyg.2012.00203)
- Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: a reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(3), 503–535. doi:[10.1037/0278-7393.33.3.503](https://doi.org/10.1037/0278-7393.33.3.503)
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: Henry Holt and Co., Inc.

- Maye, J., Aslin, R. N., & Tanenhaus, M. (2008). The Weckud Wetch of the Wast: Lexical adaptation to a novel accent. *Cognitive Science*, 32(3), 543–562. doi:[10.1080/03640210802035357](https://doi.org/10.1080/03640210802035357)
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746–748. doi:[10.1038/264746a0](https://doi.org/10.1038/264746a0)
- McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: insights from a computational approach. *Developmental Science*, 12(3), 369–78. doi:[10.1111/j.1467-7687.2009.00822.x](https://doi.org/10.1111/j.1467-7687.2009.00822.x)
- McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review*, 118(2), 219–46. doi:[10.1037/a0022325](https://doi.org/10.1037/a0022325)
- Mollon, J. (1974). After-effects and the brain. *New Scientist*, 479–482.
- Munson, C. M. (2011). Perceptual learning in speech reveals pathways of processing. Unpublished doctoral dissertation, University of Iowa.
- Newman, R. S., Clouse, S. A., & Burnham, J. L. (2001). The perceptual consequences of within-talker variability in fricative production. *The Journal of the Acoustical Society of America*, 109(3), 1181–1196. doi:[10.1121/1.1348009](https://doi.org/10.1121/1.1348009)
- Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115(2), 357–95. doi:[10.1037/0033-295X.115.2.357](https://doi.org/10.1037/0033-295X.115.2.357)
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47(2), 204–238. doi:[10.1016/S0010-0285\(03\)00006-9](https://doi.org/10.1016/S0010-0285(03)00006-9)
- Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition*, 114(2), 227–252. doi:[10.1016/j.cognition.2009.09.007](https://doi.org/10.1016/j.cognition.2009.09.007)
- Qian, T., Jaeger, T. F., & Aslin, R. N. (2012). Learning to represent a multi-context environment: more than detecting changes. *Frontiers in Psychology*, 3, 228. doi:[10.3389/fpsyg.2012.00228](https://doi.org/10.3389/fpsyg.2012.00228)
- Qian, T., Jaeger, T. F., & Aslin, R. N. (2015). *Implicit Learning of Bundles of Statistical Patterns in an Incremental Task*. Manuscript submitted for publication.
- Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1), 79–87. doi:[10.1038/4580](https://doi.org/10.1038/4580)
- Remez, R. E. (1979). Adaptation of the category boundary between speech and nonspeech: A case against feature detectors. *Cognitive Psychology*, 11(1), 38–57. doi:[10.1016/0010-0285\(79\)90003-3](https://doi.org/10.1016/0010-0285(79)90003-3)
- Remez, R. E. (1987). Neural models of speech perception: {A} case history. In Harnad, S. (Ed.) *Categorical Perception* (pp. 199–224). New York: Cambridge University Press.
- Roberts, M., & Summerfield, Q. (1981). Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Perception & Psychophysics*, 30(4), 309–14.
- Saldaña, H. M., & Rosenblum, L. D. (1994). Selective adaptation in speech perception using a compelling audiovisual adaptor. *The Journal of the Acoustical Society of America*, 95(6), 3658–61.
- Samuel, A. G. (1986). Red herring detectors and speech perception: in defense of selective adaptation. *Cognitive Psychology*, 18(4), 452–99.
- Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science*, 12(4), 348–351. doi:[10.1111/1467-9280.00364](https://doi.org/10.1111/1467-9280.00364)
- Samuel, A. G., & Kat, D. (1996). Early levels of analysis of speech. *Journal of Experimental Psychology: Human Perception and Performance*, 22(3), 676.
- Samuel, A. G., & Newport, E. L. (1979). Adaptation of speech by non-speech: evidence for complex acoustic cue detectors. *Journal of Experimental Psychology: Human Perception and Performance*, 5(3), 563–78.
- Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review*, 117(4), 1144–67. doi:[10.1037/a0020511](https://doi.org/10.1037/a0020511)
- Schouten, M. (1980). The case against a speech mode of perception. *Acta Psychologica*, 44(1), 71–98. doi:[10.1016/0001-6918\(80\)90077-3](https://doi.org/10.1016/0001-6918(80)90077-3)
- Schreiber, E., Onishi, K., & Clayards, M. (2013). Manipulating phonological boundaries using distributional cues. In *Proceedings of Meetings on Acoustics*. doi:[10.1121/1.4801082](https://doi.org/10.1121/1.4801082), Vol. 19: Acoustical Society of America.
- Sharpee, T. O., Sugihara, H., Kurgansky, A. V., Rebrik, S. P., Stryker, M. P., & Miller, K. D. (2006). Adaptive filtering enhances information transmission in visual cortex. *Nature*, 439(7079), 936–42. doi:[10.1038/nature04519](https://doi.org/10.1038/nature04519)
- Shi, L., Griffiths, T. L., Feldman, N. H., & Sanborn, A. N. (2010). Exemplar models as a mechanism for performing Bayesian inference. *Psychonomic Bulletin & Review*, 17(4), 443–64. doi:[10.3758/PBR.17.4.443](https://doi.org/10.3758/PBR.17.4.443)
- Sonderegger, M., & Yu, A. (2010). A rational account of perceptual compensation for coarticulation. In Ohlsson, S., & Catrambone, R. (Eds.) *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 375–380). Austin, TX: Cognitive Science Society.
- Stocker, A. A., & Simoncelli, E. P. (2006). Sensory Adaptation within a Bayesian Framework for Perception. In Weiss, Y., Schölkopf, B., & Platt, J. (Eds.) *Advances in Neural Information Processing Systems*, (Vol. 18 pp. 1291–1298). Cambridge, MA: MIT Press.
- Sumner, M. (2011). The role of variation in the perception of accented speech. *Cognition*, 119(1), 131–6. doi:[10.1016/j.cognition.2010.10.018](https://doi.org/10.1016/j.cognition.2010.10.018)
- Toscano, J. C., Munson, C. M., Kleinschmidt, D. F., & Jaeger, T. F. (2015). *A single mechanism for language learning across the lifespan*. Manuscript submitted for publication.
- Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences of the United States of America*, 104(33), 13273–8. doi:[10.1073/pnas.0705369104](https://doi.org/10.1073/pnas.0705369104)
- van der Zande, P., Jesse, A., & Cutler, A. (2014). Cross-speaker generalisation in two phoneme-level perceptual adaptation processes. *Journal of Phonetics*, 43, 38–46. doi:[10.1016/j.wocn.2014.01.003](https://doi.org/10.1016/j.wocn.2014.01.003)
- Vroomen, J., van Linden, S., de Gelder, B., & Bertelson, P. (2007). Visual recalibration and selective adaptation in auditory-speech perception: Contrasting build-up courses. *Neuropsychologia*, 45(3), 572–7. doi:[10.1016/j.neuropsychologia.2006.01.031](https://doi.org/10.1016/j.neuropsychologia.2006.01.031)
- Vroomen, J., van Linden, S., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Selective adaptation and recalibration of auditory speech by lipread information: dissipation. *Speech Communication*, 44(1–4), 55–61. doi:[10.1016/j.specom.2004.03.009](https://doi.org/10.1016/j.specom.2004.03.009)
- Wainwright, M. J. (1999). Visual adaptation as optimal information transmission. *Vision Research*, 39(23), 3960–3974. doi:[10.1016/S0042-6989\(99\)00101-7](https://doi.org/10.1016/S0042-6989(99)00101-7)
- Warker, J. A., & Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(2), 387–98. doi:[10.1037/0278-7393.32.2.387](https://doi.org/10.1037/0278-7393.32.2.387)
- Warker, J. A., Xu, Y., Dell, G. S., & Fisher, C. (2009). Speech errors reflect the phonotactic constraints in recently spoken syllables,

- but not in recently heard syllables. *Cognition*, 112(1), 81–96. doi:[10.1016/j.cognition.2009.03.009](https://doi.org/10.1016/j.cognition.2009.03.009)
- Weatherholtz, K., & Jaeger, T. F. (2015). *Speech perception and generalization across talkers and accents*. Manuscript submitted for publication.
- Webster, M. A., Werner, J. S., & Field, D. J. (2005). Adaptation and the Phenomenology of Perception. In Clifford, C., & Rhodes, G. (Eds.) *Fitting the mind to the world: Adaptation and after-effects in high-level vision (Advances in visual cognition)*, (Vol. 2 pp. 241–277): Oxford University Press.
- Wei, X., & Stocker, A. A. (2012). Efficient coding provides a direct link between prior and likelihood in perceptual Bayesian inference. *Advances in Neural Information Processing Systems*, 25, 1–9.
- Yildirim, I., Degen, J., Tanenhaus, M. K., & Jaeger, T. F. (in press). Talker-specificity and adaptation in quantifier interpretation. *Journal of Memory and Language*.
- Zäske, R., Fritz, C., & Schweinberger, S. R. (2013). Spatial inattention abolishes voice adaptation. *Attention, Perception & Psychophysics*, 75(3), 603–13. doi:[10.3758/s13414-012-0420-y](https://doi.org/10.3758/s13414-012-0420-y)