frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 01 February 2012
doi: 10.3389/fpsyg.2012.00010

Listening for the norm: adaptive coding in speech categorization

Jingyuan Huang* and Lori L. Holt

Department of Psychology, Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA, USA

Edited by:
Peter Neri, University of Aberdeen, UK

Reviewed by:
Simon Baumann, Newcastle University, UK
Emily Myers, University of Connecticut, USA

****Correspondence:***
Jingyuan Huang, Department of Psychology, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA.
e-mail: jingyuan@andrew.cmu.edu

Perceptual aftereffects have been referred to as "the psychologist's microelectrode" because they can expose dimensions of representation through the residual effect of a context stimulus upon perception of a subsequent target. The present study uses such context-dependence to examine the dimensions of representation involved in a classic demonstration of "talker normalization" in speech perception. Whereas most accounts of talker normalization have emphasized talker-, speech-, or articulatory-specific dimensions' significance, the present work tests an alternative hypothesis: that the long-term average spectrum (LTAS) of speech context is responsible for patterns of context-dependent perception considered to be evidence for talker normalization. In support of this hypothesis, listeners' vowel categorization was equivalently influenced by speech contexts manipulated to sound as though they were spoken by different talkers and non-speech analogs matched in LTAS to the speech contexts. Since the non-speech contexts did not possess talker, speech, or articulatory information, general perceptual mechanisms are implicated. Results are described in terms of adaptive perceptual coding.

Keywords: talker normalization, LTAS, speech perception

INTRODUCTION

Perceptual systems adjust rapidly to changes in the environment, with neural and behavioral responses dynamically changing to mirror changes in the input (Sharpee et al., 2006; Gutinsky and Dragoi, 2008). Such adaptive codes provide efficient representations because they direct computational resources toward uncommon inputs and provide information about potentially important changes in the world (Barlow, 1990).

Although adaptive coding is less-well-studied in audition than vision, behavioral demonstrations of context-dependence, particularly in speech perception, resonate with the perspective that context adaptively tunes perceptual codes. The identity (Ladefoged and Broadbent, 1957) or accent (Evans and Iverson, 2004) of a talker, the rate of the utterance (Liberman et al., 1956), and the phonetic make-up of a preceding utterance (Mann, 1986) all influence perception of subsequent speech targets. From an adaptive coding perspective, the perceptual system's response to preceding speech lingers to affect subsequent processing.

Significantly, however, this can only be true to the extent that context and target share common neural resources. In this way, perceptual aftereffects of context have been described as "the psychologist's microelectrode" (Frisby, 1980) because they can expose neural coding related to perceptual experience through the residual effect of one stimulus upon perception of another (Clifford and Rhodes, 2005). Thus, context-dependent speech perception perhaps can serve to reveal the underlying representation of speech.

A simple model clarifies the approach (**Figure 1**, after Rhodes et al., 2005). Imagine two populations of units coding a dimension of auditory representational space (e.g., acoustic frequency, or a higher-order feature like talker identity). Pool 1 units best code below-average values along the dimension whereas Pool 2 units better code above-average values; within each pool, more extreme values are coded more robustly. The average value along the dimension is encoded implicitly in the neutral cross-over point at which pools respond equivalently. An input with a higher value along the dimension would result in strong Pool 2 response, thus reducing Pool 2 responsiveness in the short-term due to adaptation. In this way, the context stimulus "lingers" in the perceptual system to affect the resources available to process subsequent stimuli. As is evident in **Figure 1**, this reduction shifts the neutral point where the two pools respond equivalently toward higher values and the formerly "average" value is now more robustly coded by Pool 1 neurons. Overall, adaptive encoding results in a contrastive shift in target encoding as a function of whether the context was better-coded by Pool 1 or Pool 2 neurons.

This toy model is simple, to be sure, but the general principle may extend to higher-dimensional perceptual spaces and more complex representations. In vision, adaptive coding has been successful in predicting patterns of interactions among low-level representations for brightness and hue (see Frisby, 1980) as well as higher-level representations for faces and bodies (see Clifford and Rhodes, 2005). What is implicit in this approach is the assumption that context and target stimuli are encoded along common dimension(s) of representation. Here, we pursue context-dependent speech categorization to examine significant representational dimensions of speech categories.

For these purposes, Ladefoged and Broadbent's (1957) classic demonstration of talker normalization is relevant. In their study, listeners heard a constant target-word with a relatively ambiguous

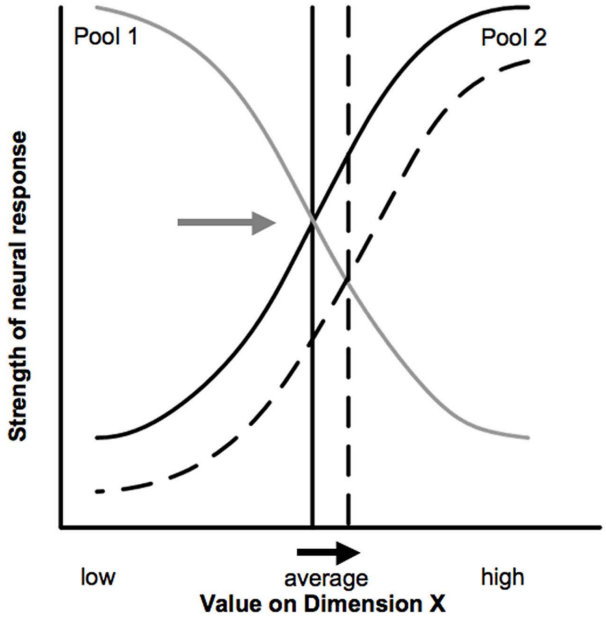

FIGURE 1 | Model of adaptive coding adapted from Rhodes et al. (2005).

vowel at the end of a context phrase "*Please say what this word is*. . .." The first (F1) and/or second (F2) formant frequencies (peaks in energy of a voice spectrum; Fant, 1960) of the context phrase were increased or decreased. These shifts can be conceptualized, respectively, as decreasing and increasing the talker's vocal tract length and, correspondingly, as a change in talker. When the resulting phrases preceded the speech targets, listeners' categorization shifted in a manner suggesting that they were compensating, or normalizing, for the change in vocal tract length or talker. A constant vowel was more often heard as "bit" when it followed a phrase synthesized as though spoken by a shorter vocal tract (higher formant frequencies in the phrase), but more often as "bet" following the same phrase modeling speech from a longer vocal tract (lower frequencies).

A central and enduring theoretical issue has been the representational dimension across which listeners "normalize" speech categorization in this way. Is the relevant representational dimension talker identity, vocal tract shape/anatomy, or acoustic phonetic space (Joos, 1948; Ladefoged and Broadbent, 1957; Halle and Stevens, 1962; Nordstrom and Lindblom, 1975; McGowan, 1997; McGowan and Cushing, 1999; Poeppel et al., 2008)?

Here, we investigate the extent to which Ladefoged and Broadbent's classic results can be explained by adaptive coding along a representational dimension that has a general auditory, rather that talker-, or speech-specific basis: the long-term average spectrum (LTAS) of the preceding sound. Recent research has suggested that listeners are sensitive to the LTAS of sound stimuli and adjust perception of subsequent sounds contrastively opposing context LTAS. Holt (2005, 2006a,b) found that sequences of 21 non-speech sine-wave tones, each with a unique frequency sampling a 1000 Hz range affect speech categorization of /ga/–/da/ as a function of the mean frequency of the tones forming the sequence. The influence of these contexts on speech categorization cannot be attributed to any particular acoustic segment of the sequences because tones were randomly ordered on a trial-by-trial basis. Instead, the pattern of context-dependent speech categorization is predicted only by the tone sequences' LTAS. Perception of the subsequent speech targets was relative to, and contrastive with, the LTAS consistent with the adaptive coding scheme sketched above if LTAS serves as the common representational dimension linking non-speech contexts and speech targets.

Here, we seek to directly replicate the Ladefoged and Broadbent (1957) results with speech contexts and to explicitly test whether non-speech contexts modeling critical characteristics of the speech contexts' LTAS produce the same effects on vowel categorization. Directly mimicking the methods of Ladefoged and Broadbent (1957), listeners categorized a series of speech sounds varying perceptually from "bet" to "but" in the context of a preceding phrase ("Please say what this word is. . ."). This phrase was synthesized to model two different "talkers": one with a larger vocal tract and the other with a smaller vocal tract. The same listeners also categorized the same "bet" to "but" targets preceded by sequences of non-speech sine-wave tones with frequencies modeling the LTAS of the context phrases. These extremely simple acoustic contexts carried no talker- or speech-specific information. Should the speech and non-speech contexts similarly influence vowel categorization, it would suggest that speech and non-speech contexts draw upon common neural resources. Rather than articulatory or talker-specific dimensions that typically have been proposed to account for talker normalization, listeners may rely on sounds' LTAS to tune speech perception.

MATERIALS AND METHODS

PARTICIPANTS

Twenty-six adult native-English speakers from the Carnegie Mellon University campus with no reported speech or hearing disabilities were recruited for the experiment. All received written informed consent in accord with Carnegie Mellon University ethics approval, and course credit for their time.

STIMULI

Figure 2 illustrates stimulus design. Each stimulus had a 1600 ms context segment, followed by a 50 ms silent interval and a 250 ms speech categorization target drawn from a six-step series of speech syllables varying perceptually from /bɛt/ to /b∧t/ (*bet* to *but*). The /bV/ segment was created by varying the second formant (F2) frequency of the main vowel portion in equal steps from 1300 Hz/b∧/ to 1700 Hz/bɛ/ using Klattworks (McMurray, in preparation). The onset F2 frequency was 1100 Hz and gradually changed to the target frequency across 50 ms. Similarly, the first formant was 150 Hz and linearly transitioned to 600 Hz over 50 ms. The fundamental frequency and the third formant frequency were held constant at 120 and 2600 Hz, respectively. The /t/ segment was taken from a natural utterance of "*whit*" recorded from a male native-English talker and appended to each speech target.

Two types of context preceded these speech targets, one speech and the other a sequence of tones. The speech context was generated by extracting formant frequencies and bandwidths from

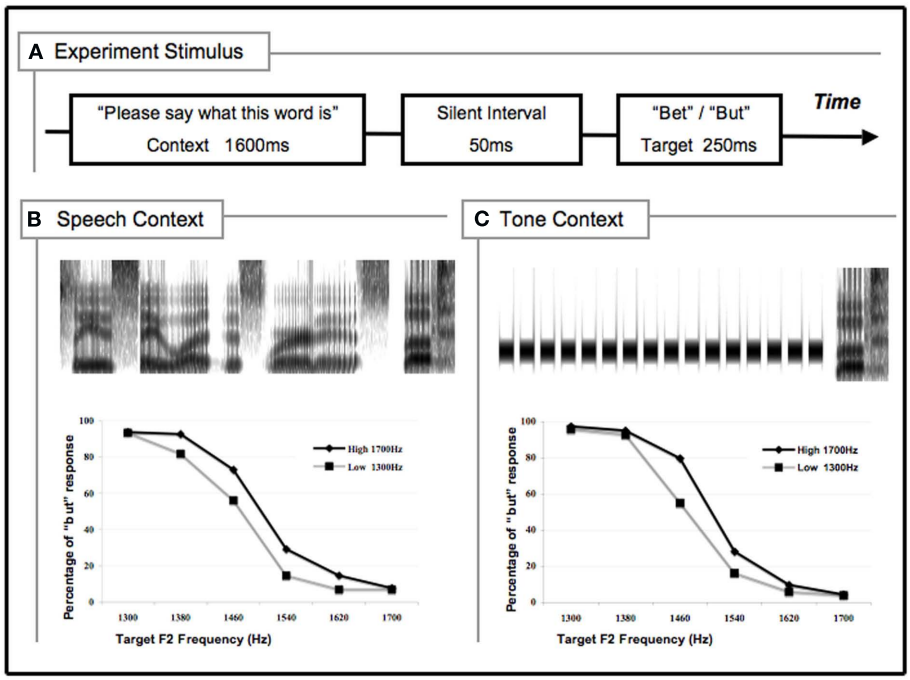

FIGURE 2 | Stimulus design and results of the experiment. (A) Schematic illustration of stimulus components; **(B)** spectrogram in time x frequency dimensions for the high mean speech context (top panel) and mean percentage of "but" responses in speech contexts (bottom panel); **(C)** spectrogram in time x frequency dimensions for a representative high mean tone context (top panel) and mean percentage of "but" responses in tone context (bottom panel). Preceded by both speech **(B)** and tone **(C)** contexts, higher-frequency contexts led to more low-frequency target responses ("but"), and vice versa.

a recording a male voice uttering the sentence "*Please say what this word is. . .*," and using these values to synthetically reproduce the sentence in the parallel branch of the Klatt and Klatt (1990) synthesizer. Following the methods of Ladefoged and Broadbent with modern techniques, this 1600 ms base phrase was spectrally manipulated by adjusting formant center frequencies and bandwidths to create different "talkers." To mimic a longer vocal tract, a voice with relatively lower frequencies in the region of F2 was created (across the phrase, F2 frequencies ranged from 390 to 1868 Hz with an average of 1300 Hz). A voice with a relatively shorter vocal tract was mimicked by increasing these base frequencies by 400 Hz. Thus, the mean acoustic energy in the range of F2 approximated the energy varying across the /b∧t/–/bɛt/ speech target stimuli.

The non-speech tone contexts were composed of a sequence of 16 repeated 70 ms sine-wave tones (5 ms linear amplitude onset/offset ramps) with 30 ms silent intervals separating them as in Holt (2005; 1600 ms total duration). The tone contexts modeled the mean F2 frequency of the speech contexts (1300 and 1700 Hz for the long and short vocal tracts, respectively) and each tone was a single harmonic without variation. As such, the non-speech contexts did not sound like speech or possess information about talker identity, vocal tract anatomy, or phonetic space. Thus, these non-speech contexts eliminated shared information between context and target along talker- and speech-specific dimensions while preserving a similar frequency-specific peak in the LTAS. Stimuli were RMS-matched in amplitude to the "bet" endpoint of the target-word series. All stimuli were sampled at 11025 Hz.

PROCEDURE

Participants categorized each target as "bet" or "but" using labeled keyboard buttons across a 1 h experiment under the control of E-prime (Schneider et al., 2002). Participants first categorized 10 randomly ordered repetitions of each speech target in isolation to assure that targets were well-categorized as the intended vowels. They then categorized the same speech targets preceded by high and low versions of speech and non-speech contexts, blocked by context type with block order counterbalanced across participants. Each context/target pairing was presented 10 times in a random order. Sounds were presented diotically over linear headphones (Beyer Dt-150) at approximately 70 dB SPL(A).

RESULTS

Participants' categorization of speech targets in isolation was orderly, as indicated by a significant main effect target F2 frequency, $F\,(5, 25) = 139.43$, $p < 0.01$. Individuals' data conformed to this average pattern.

Figures 2B,C illustrate the influence of context on vowel categorization. A 2 (context type, speech/non-speech) X 2 (LTAS, low/high) X 6 (target F2 frequency) repeated-measures ANOVA of listeners' percent "but" responses reveals that, as expected from vowel categorization in isolation, responses varied reliably as a function of the target F2 frequency, $F\,(5, 25) = 290.50$, $p < 0.001$. Moreover, there was no main effect of context type indicating no overall bias in vowel categorizations a function of the type of context that preceded targets, $F\,(1, 25) = 0.274$, $p = 0.61$.

Of greater interest, there was a significant main effect of LTAS on vowel categorization, $F\ (1,\ 25) = 27.05$, $p < 0.001$. The vowels were categorized as /ʌ/ significantly more often following high-frequency contexts whereas the same vowels were more often labeled as /ɛ/ following low-frequency contexts. Thus, the context-dependent effect was spectrally contrastive and consistent with previous studies of the influence of sentence-length contexts on speech categorization (Ladefoged and Broadbent, 1957; Watkins and Makin, 1994, 1996; Holt, 2005, 2006a,b; Huang and Holt, 2009). The interaction between LTAS and target F2 frequency was significant, $F\ (5,\ 25) = 11.65$, $p < 0.001$, indicating that context had a greater influence on perceptually ambiguous targets.

This pattern of contrastive context-dependent vowel categorization was evident for both speech, $F\ (1, 25) = 21.62$, $p < 0.001$, and non-speech contexts, $F\ (1,\ 25) = 17.38$, $p < 0.001$. Of primary interest, there was no significant interaction between context type and LTAS, $F\ (1, 25) = 0.324$, $p = 0.574$. It is interesting that although the LTAS contrast was larger in the non-speech contexts compared with speech context condition (**Figure 3**; see detail explanation in discussion), the magnitude of the influence of speech and non-speech contexts on speech target categorization was statistically indistinguishable. Neither the interaction between context type and target frequency, $F\ (1,\ 25) = 1.22$, $p = 0.30$, nor the three-way interaction was significant, $F\ (5,\ 25) = 1.424$, $p = 0.22$. In sum, "talker" is not an essential element of talker normalization as it appears even in the absence of a talker when context is merely a sequence of sine-wave tones.

DISCUSSION

We exploited context-dependent speech categorization as "the psychologist's microelectrode" (Frisby, 1980) to reveal characteristics of the underlying representation of speech. The residual effect of one stimulus upon perception of another demands that the two share common neural processing and/or representation. Thus, the comparable influence of speech and non-speech contexts on speech categorization indicates a common substrate of interaction. Importantly, since the two context types did not share linguistic, articulatory gestural, or talker-specific information, but yet produced equivalent effects on speech categorization, it does not appear that speech-, vocal tract-, or talker-specific information is essential in eliciting the patterns of context-dependent perception that have been described in the literature as "talker normalization."

What the two context types in the present experiment shared was a similar pattern of spectral energy across their time course. **Figure 3** illustrates the LTAS of the speech-target endpoints (3A) and speech (3B) and non-speech contexts (3C). Gray shading highlights the region of acoustic energy that critically distinguishes the speech targets. We suggest that the context-dependent speech categorization observed here (and in Ladefoged and Broadbent, 1957) arises because the auditory system is sensitive to the context LTAS and the speech target is encoded relative to, and contrastive with, that long-term average.

Specifically, we propose that speech is adaptively coded according to the LTAS of ambient sound. The simple model described in the introduction can clarify one means by which this might be accomplished. Imagine pools of neurons sensitive to energy in the range of the second formant with one pool better coding relatively lower frequencies and the other better coding higher frequencies. By this model, presentation of the speech context modeling a longer vocal tract with lower-frequency energy within this frequency range would result in greater activity among the pool of neurons that better code lower frequencies. Subsequent adaptation among this pool of responsive neurons would result in a shift toward the opposite, higher-frequency neural pool at the time of speech target presentation. The formerly "neutral" frequencies would be now more robustly encoded by the higher-frequency neural pool, shifting representation contrastively away from the lower-frequency context. This adaptive coding serves to exaggerate differences between the LTAS of the context and target (Holt, 2006a).

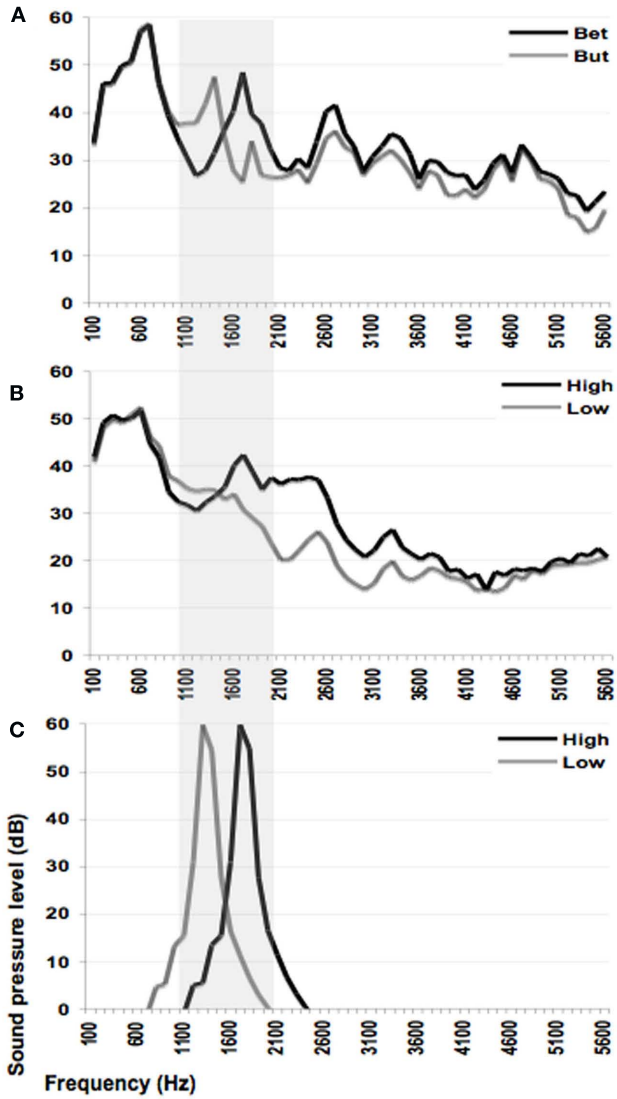

FIGURE 3 | Long-term average spectrum of target stimuli (A), speech contexts (B), and tone contexts (C). The shaded area indicates F2 frequencies critical to distinguishing the speech targets and two context types (speech and tone) in present experiments.

By this model, "talker normalization" effects on speech targets are predicted and obtained even when no speech information is available in the context. Of note, the LTAS model makes no reference to specific linguistic units, such as phonemes. This generality makes the adaptive coding approach in general, and the LTAS model in particular, extend straightforwardly to other normalization phenomena in speech perception. Huang and Holt (2009) report that shifts in the peak energy of the LTAS in the region of the fundamental frequency (f0) of a Mandarin Chinese sentence predict patterns of context-dependent Mandarin lexical tone normalization (Leather, 1983; Fox and Qi, 1990; Moore and Jongman, 1997). Further, non-speech precursors with matched LTAS produce the same effects. Similarly, although the adaptive coding approach to context-dependent speech categorization reveals LTAS as an important dimension of representation, the model's application is more general. Wade and Holt (2005), for example, investigated rate-dependent normalization effects whereby the rate of a precursor sentence affects how listeners categorize a rate-dependent speech distinction like /ba/ versus /wa/, finding that the rate of presentation of a sequence of non-speech tones evokes a similar contrastive influence on speech categorization.

If demonstrations of "talker normalization" can be accounted for by general perceptual processes that are not talker- or speech-specific then there remains the question of whether the context-dependent speech categorization taken as evidence of normalization really accommodates the acoustic variability in speech that arises from different talkers. We argue that it does. Across context sentences like those of the current study, speech maps the scope of a talker's articulatory space and the mean of that space resembles a talker's neutral vowel, the shape of the non-articulating vocal tract (Story, 2005). This neutral vowel serves as an effective normalization referent because, as Story (2005) has demonstrated, most of the variability across talkers can be accounted for by differences in the shape the vocal air space of talkers' neutral vowels. Thus, if listeners were able to extract an estimate of the neutral vowel, the mean of the articulatory space, they would have an excellent referent for talker normalization. However, tracking back from acoustics to articulator requires negotiating the inverse problem. The results we describe here suggest an alternative.

As talkers produce a variety of consonants and vowels, speech maps the articulatory space but it also produces a sound spectrum that maps the acoustic space and, across time, samples an LTAS. Instead of solving the inverse problem to recover the actual neutral vocal tract shape as an articulatory referent for normalization, listeners may use the average spectrum LTAS as an auditory referent. The observation that non-speech tones modeling the LTAS of a talker are as effective in shifting speech categorization as speech contexts supports the viability of a general auditory referent. However, it should be noted that LTAS is unlikely to be the only contributing factor in talker normalization. Speaker identities, for example, may mediate listeners' attention and expectation and influence the way listeners tune their perception to the preceding contexts (Magnuson and Nusbaum, 2007). Talker normalization is likely to be a multi-facet phenomenon. Nonetheless LTAS, which provides sufficient context information via general perceptual processes, is an important factor in the adaptive coding of speech perception processing that contributes to talker normalization.

The lingering influence of a sequence of non-speech tones on listeners' response to speech targets indicates a shared substrate, the level of which remains to be evaluated. Relevant to this, Holt (2005) reported that non-speech contexts influence on speech categorization persists across 1300 ms of silence. Moreover, the aftereffect of the tones was present even when 13 neutral-frequency tones intervened between tone contexts and speech targets. The influence of temporally non-adjacent tone sequences can even override the influence of temporally adjacent speech contexts on speech targets (Holt, 2006b). These observations argue for a central, rather than peripheral, substrate.

In fitting the mind to the world, the ambient context plays a large role in how input is coded. The present results demonstrate that patterns of context-dependent speech categorization long taken to be evidence of talker-specific normalization for articulatory referents or rescaling of phonetic space may arise, instead, from general principles of adaptive perceptual coding. In "listening for the norm," listeners appear to adjust perception to regularities of the ambient environment.

ACKNOWLEDGMENTS

This work was supported by NIH grants R01DC004674.

REFERENCES

Barlow, H. B. (1990). "A theory about the functional role and synaptic mechanism of visual aftereffects," in *Vision: Coding and Efficiency*, ed. C. Blakemore (Cambridge: Cambridge University Press), 363–375.

Clifford, C. W. G., and Rhodes, G. (eds). (2005). *Fitting the Mind to the World: Adaptation and After Effects in High-Level Vision*. Oxford: Oxford University Press.

Evans, B. G., and Iverson, P. (2004). Vowel normalization for accent: an investigation of best exemplar locations in northern and southern British English sentences. *J. Acoust. Soc. Am.* 115, 352–261.

Fant, G. (1960). *Acoustic Theory of Speech Production*. The Hague: Mouton & Co.

Fox, R., and Qi, Y. (1990). Contextual effects in the perception of lexical tone. *J. Chin. Ling.* 18, 261–283.

Frisby, J. P. (1980). *Seeing: Illusion, Mind and Brain*. Oxford: OUP.

Gutinsky, D. A., and Dragoi, V. (2008). Adaptive coding of visual information in neural populations. *Nature* 452, 220–224.

Halle, M., and Stevens, K. N. (1962). Speech recognition: a model and a program for research. *IEEE Trans. Inf. Theory* 8, 155–159.

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

Holt, L. L. (2006a). The mean matters: effects of statistically-defined nonspeech spectral distributions on speech categorization. *J. Acoust. Soc. Am.* 120, 2801–2817.

Holt, L. L. (2006b). Speech categorization in context: joint effects of nonspeech and speech precursors. *J. Acoust. Soc. Am.* 119, 4016–4026.

Huang, J., and Holt, L. L. (2009). General perceptual contributions to lexical tone normalization. *J. Acoust. Soc. Am.*125, 3983–3994.

Joos, M. (1948). Acoustic phonetics. *Language* 24, 1–136.

Klatt, D. H., and Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am.* 87, 820–857.

Ladefoged, P., and Broadbent, D. E. (1957). Information conveyed by vowels. *J. Acoust. Soc. Am.* 29, 98–104.

Leather, J. (1983). Speaker normalization in perception of lexical tone. *J. Phon.* 11, 373–382.

Liberman, A. M., Delattre, P. C., Gerstman, L. J., and Cooper, F. S. (1956). Tempo of frequency change as a cue for distinguishing classes of speech sounds. *J. Exp. Psychol.* 52, 127–137.

Magnuson, J. S., and Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 391–409.

Mann, V. A. (1986). Distinguishing universal language-specific factors in speech perception: evidence from Japanese listeners' perception of /l/ and /r/. *Cognition* 24, 169–196.

McGowan, R. S. (1997). Normalization for articulatory recovery. *J. Acoust. Soc. Am.* 101, 3175.

McGowan, R. S., and Cushing, S. (1999). Vocal tract normalization for midsagittal articulatory recovery with analysis-by-synthesis. *J. Acoust. Soc. Am.* 106, 1090–1105.

Moore, C., and Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *J. Acoust. Soc. Am.* 102, 1864–1877.

Nordstrom, P. E., and Lindblom, B. (1975). "A normalization procedure for vowel formant data," in *Paper presented at the 8th International Congress of Phonetic Sciences*, Leed, 17–23.

Poeppel, D., Idsardi, W. J., and van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 1071–1086.

Rhodes, G., Robbins, R., Jaquet, E., McKone, E., Jeffery, L., and Clifford, C. W. G. (2005). "Adaptation and face perception – how aftereffects implicate norm based coding of faces," in *Fitting the Mind to the World: Aftereffects in High-Level Vision*, eds C. W. G. Clifford and G. Rhodes (Oxford: Oxford University Press), 213–240.

Schneider, W., Eschman, A., and Zuccolotto, A. (2002). *E-Prime User's Guide*. Pittsburgh: Psychology Software Tools Inc.

Sharpee, T. O., Sugihara, H., Kurgansky, A. V., Rebrik, S. P., Stryker, M. P., and Miller, K. D. (2006). Adaptive filtering enhances information transmission in visual cortex. *Nature* 439, 936–942.

Story, B. H. (2005). A parametric model of the vocal tract area function for vowel and consonant simulation. *J. Acoust. Soc. Am.* 117, 3231–3234.

Wade, T., and Holt, L. L. (2005). Perceptual effects of preceding nonspeech rate on temporal properties of speech categories. *Percept. Psychophys.* 67, 939–950.

Watkins, A. J., and Makin, S. J. (1994). Perceptual compensation for speaker differences and for spectral-envelope distortion. *J. Acoust. Soc. Am.* 96, 1263–1282.

Watkins, A. J., and Makin, S. J. (1996). Effects of spectral contrast on perceptual compensation for spectral-envelope distortions. *J. Acoust. Soc. Am.* 99, 3749–3757.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 19 October 2011; accepted: 10 January 2012; published online: 01 February 2012.

Citation: Huang J and Holt LL (2012) Listening for the norm: adaptive coding in speech categorization. Front. Psychology ***3****:10. doi: 10.3389/fpsyg.2012.00010*

This article was submitted to Frontiers in Perception Science, a specialty of Frontiers in Psychology.