Cognition 117 (2010) 217–223

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Brief article

Auditory adaptation in vocal affect perception

Patricia E.G. Bestelmeyer [a,*], Julien Rouger [b], Lisa M. DeBruine [c], Pascal Belin [a]

[a] Institute of Neuroscience and Psychology, University of Glasgow, UK
[b] Department of Cognitive Neuroscience, University of Maastricht, The Netherlands
[c] Department of Psychology, University of Aberdeen, UK

ARTICLE INFO

Article history:
Received 9 December 2009
Revised 4 August 2010
Accepted 6 August 2010

Keywords:
Auditory adaptation
Voice perception
Emotion
Vocal affect

ABSTRACT

Previous research has demonstrated perceptual aftereffects for emotionally expressive faces, but the extent to which they can also be obtained in a different modality is unknown. In two experiments we show for the first time that adaptation to affective, non-linguistic vocalisations elicits significant auditory aftereffects. Adaptation to angry vocalisations caused voices drawn from an anger–fear morphed continuum to be perceived as less angry and more fearful, while adaptation to fearful vocalisations elicited opposite aftereffects (Experiment 1). We then tested the link between these aftereffects and the underlying acoustics by using caricatured adaptors. Although caricatures exaggerated the acoustical and affective properties of the vocalisations, the caricatured adaptors resulted in aftereffects which were comparable to those obtained with natural vocalisations (Experiment 2). Our findings suggest that these aftereffects cannot be solely explained by low-level adaptation to acoustical characteristics of the adaptors but are likely to depend on higher-level adaptation of neural representations of vocal affect.

1. Introduction

Perception and recognition of signals conveying affect, such as from faces or voices, is important for everyday social functioning. In the auditory domain, nonverbal signals in particular have been found to be crucial in communicating emotional information (Wallbott & Scherer, 1986). This affective information is easily recognised (Elfenbein & Ambady, 2002) and is perceived categorically. For example, using synthetic continua ranging from one discrete emotion to another (e.g. from happiness to fear), Laukka (2005) found that these continua are divided by a clear category boundary; that is, the morphs were perceived as belonging to either one category or the other with the fifty percent morph being randomly categorised. This category boundary was particularly clear for anger–fear or anger–sadness continua.

Here we were interested in the perceptual representation of vocal affect and in particular the plasticity of this perceptual boundary. While research on the representation of facial affect has been extensive investigations regarding the representations of vocal affect have been largely neglected. To this end we employed adaptation paradigms which have been helpful in furthering our understanding of how sensory signals are coded and organised in the brain (Webster, Kaping, Mizokami, & Duhamel, 2004). Adaptation refers to a process during which continued stimulation results in a biased perception towards opposite features of the adapting stimulus (Grill-Spector et al., 1999). Research using adaptation has revealed neural populations tuned to respond to specific stimulus attributes by isolating and subsequently distorting the perception of these attributes (Grill-Spector et al., 1999; Winston, Henson, Fine-Goulden, & Dolan, 2004). Traditionally, studies have investigated low-level properties of impoverished stimuli by varying one of the stimulus dimensions such

* Corresponding author. Address: Institute of Neuroscience and Psychology, Centre for Cognitive Neuroimaging, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, UK. Tel.: +44 141 3304965; fax: +44 141 3304606.
E-mail address: p.bestelmeyer@psy.gla.ac.uk (P.E.G. Bestelmeyer).

0010-0277/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2010.08.008

as colour, shape or motion. More recently, adaptation paradigms have been used to understand more complex aftereffects, particularly in the domain of face perception, and have been crucial in evaluating models in the field (Bruce & Young, 1986; Haxby, Hoffman, & Gobbini, 2000).

Webster and MacLin (1999) were the first to show that extended exposure to faces can generate aftereffects. Adaptation to consistently distorted faces (e.g. expanded features) caused subsequently viewed unmanipulated faces to appear distorted in the opposite direction of the adapting stimulus (e.g. compressed features). This effect transferred to faces of different identities. Later studies demonstrated that these contrastive aftereffects are robust to changes in retinal position (Leopold, O'Toole, Vetter, & Blanz, 2001), size (Zhao & Chubb, 2001) and angular orientation (Watson & Clifford, 2003) of the adapting stimulus, suggesting that adaptation to those low-level features cannot fully explain face aftereffects. Thus adaptive coding is a strategy employed by the neural system throughout the perceptual processing hierarchy and is not only employed at early stages of information processing. Moreover, even high-level, socially important properties such as facial identity (Leopold, O'Toole, Vetter, & Blanz, 2001), attractiveness (Rhodes, Jeffery, Watson, Clifford, & Nakayama, 2003), eye-gaze (Jenkins, Beaver, & Calder, 2006) or expression (Fox & Barton, 2007; Webster et al., 2004) can be isolated using adaptation. Researchers have interpreted the resulting aftereffects to reflect a recalibration of neural processes to continuously updated stimulation (Webster & MacLin, 1999).

While there is substantial evidence for contrastive aftereffects for social information in face perception, auditory aftereffects in the perception of human voices have received surprisingly little scientific attention. Belin, Fecteau, and Bédard (2004) have proposed a model of voice processing in which voices are represented in analogous ways to faces. Just as in Bruce and Young's (1986) model, Belin et al. (2004) proposed that after an initial low-level auditory and structural analysis, vocal information processing is dealt with by three independent modules dedicated to higher-level processing of speech, vocal identity and affect information (though empirically a strict independence of the three systems is not supported either in the face (see e.g. Bestelmeyer et al., 2008) or voice literature (see e.g. Nygaard, Sommers, & Pisoni, 1994)). Although there has been a surge in studies testing proposed models of face perception using adaptation paradigms, Belin et al.'s model remains largely untested. Given analogous findings in the face literature and the claimed but yet scarcely tested resemblance between voice and face perception models, it should be possible to adapt to the information handled by each one of these three systems.

Despite extensive research on adaptation in perceiving linguistic contrasts in speech perception (e.g. Ades, 1976; Sawusch, 1977) few studies have investigated adaptation of non-linguistic attributes of voice. Mullennix, Johnson, TopcuDurgun, and Farnsworth (1995) were the first to investigate voice gender adaptation using a synthetic male–female voice continuum. Although some evidence for adaptation to voice gender was found, the effects were small and unreliable across experiments. More recently, and with more advanced morphing technology between natural male and female voices, Schweinberger et al. (2008) provided evidence for systematic voice gender aftereffects by repeating brief vocal stimuli to induce adaptation. Adaptation to female voices caused androgynous voices to sound more male and vice versa. These effects were still measurable after several minutes and were not simply due to adaptation to pitch. Thus, there is evidence of aftereffects caused by adaptation in the domain of voice processing; however no study has yet investigated other vocal properties, in particular socially relevant aspects of voice perception such as affect.

In two experiments we investigated whether aftereffects can be observed for adaptation to affective vocal expressions such as anger and fear. We created morph continua between angry and fearful vocalisations (Belin, Fillion-Bilodeau, & Gosselin, 2008) because these emotions are divided by a clear category boundary both in the voice (Laukka, 2005) and face literature (e.g. Calder, Young, Perrett, Etcoff, & Rowland, 1996). First, we established the baseline affect classification psychophysical function (i.e. an acoustic to perceptual mapping function) without adaptation. We then examined whether adaptation to fearful vocalisations would shift perception of affectively ambiguous stimuli towards anger and whether exposure to angry vocalisations would shift perception in the opposite direction. In a second experiment we created vocal caricatures of both emotions to examine whether adaptation to these acoustically and affectively exaggerated versions would enhance the adaptation effect. We reasoned that an enhancement would mean adaptation to acoustic properties of the voices, and thus would suggest that the adaptation effect takes place at a relatively low-level of the auditory processing hierarchy. By contrast, no change in the adaptation effects would imply that adaptation is relatively independent of acoustic properties, and occurs at higher-level stages of auditory information processing.

2. Methods

2.1. Participants

Fifteen volunteers took part in Experiment 1 (9 female, mean age = 20.6, standard deviation (SD) = 2.70). Fourteen different volunteers contributed to Experiment 2 (12 female, mean age = 24.3, SD = 5.97). All participants reported normal hearing and were reimbursed £9 for their time. Volunteers belonged to the undergraduate community of the University of Glasgow.

2.2. Stimuli

Recordings were taken from the Montreal Affective Voices (Belin et al., 2008) in which actors were instructed to produce emotional interjections using the vowel/a/. The voices from four identities (two female) expressed anger and fear. Stimuli were normalised in energy (root mean square) and presented in stereo via Beyerdynamic headphones at an intensity of 80 dB SPL(C). In Experiment 1 angry to fearful continua were created separately for each identity, in seven steps that corresponded to 5/95%, 20/80%, 35/65%, 50/50%, 65/35%, 80/20% and 95/5% fear/anger. The original angry (0/100%) and fearful (100/0%) voices were used as adaptors.

For Experiment 2 we re-used the stimuli generated for Experiment 1. Additionally, we created caricatures corresponding to −25/125% (caricatured anger relative to the fear stimulus) and 125/−25% fear/anger (caricatured fear relative to the anger stimulus). The duration of the vocalisations within each continuum was kept constant and ranged between 0.6 and 0.9 s across continua. We used STRAIGHT (Kawahara & Matsui, 2003) for stimulus manipulation (see Supplementary section for more detail) and the Psychtoolbox3 (Brainard, 1997; Pelli, 1997) for stimulus presentation. Both programs are based on MatlabR2007b (Mathworks, Inc).

To verify that the caricatures were affectively enhanced compared to the original voices, we conducted a control experiment (13 participants; 8 female) assessing intensity, arousal and valence of the stimuli using four separate analogue rating scales presented in separate blocks. We measured the affective intensity of the stimuli using rating scales ranging from "extremely angry" to "not angry at all" (same for fearful). Caricatures were rated as significantly more angry and fearful, respectively, than the original versions (paired *t*-tests both at: $t(12) > 2.2$; $p < .05$). Caricatures were rated as more arousing than the original stimuli (paired *t*-tests both at: $t(12) > 2.47$; $p < .05$). The valence ratings did not differ between the two adaptor types and were perceived as equally negative. Original and caricatured voices differed acoustically (see Supplementary section for analyses).

2.3. Procedure

Experiments were conducted in a soundproof booth and consisted of two main parts, a baseline emotion categorisation task without prior adaptation and the adaptation tasks. All tasks required a two-alternative forced choice judgement of whether the voice expressed either anger or fear by means of a button press.

The baseline task consisted of two blocks of trials, one for each gender and was always administered first. The voice of each identity at each of the seven morph steps was repeated six times, leading to 84 trials per gender block. In Experiment 2, the baseline task contained stimuli from nine morph steps (7 stimuli from Experiment 1, plus 2 caricatures) which were repeated four times, providing a total number of 72 trials per block. Within each block voices were presented randomly with an inter-stimulus interval of 2–3 s (randomly jittered with a uniform distribution). Following the baseline task we presented participants with the adaptation blocks.

The trial structure of the adaptation tasks consisted of one adapting voice played four times in succession and followed by an ambiguous morph (test stimulus) after a silent gap of 1 s. The inter-trial interval was 5 s during which the participants judged the expression of the test stimulus. Experiment 1 consisted of four adaptation blocks (2 emotion × 2 gender) and each of the seven test stimuli per identity was repeated six times leading to a total of 84 trials per block. Experiment 2 consisted of eight adaptation blocks (2 emotion × 2 adaptors × 2 gender). Here, each of the seven test stimuli per identity was repeated four times leading to 56 trials per block.

To avoid low-level adaptation to factors such as identity-specific acoustic properties, participants were always tested on a different identity than the one they were adapted to. Half of the participants were adapted to angry and then fearful voices while the remaining participants were presented with the reverse order of blocks in Experiment 1. Additionally, in Experiment 2, original and caricatured adaptor blocks of one emotion type were presented in succession. After the caricatured and original adaptors of one type (e.g. caricatured male fear, caricatured female fear, original male fear, original female fear) had been presented participants were asked to take a 10 min break before the next set of original and caricatured adaptor blocks of the remaining emotion (e.g. caricatured male anger, caricatured female anger, original male anger, original female anger) was administered. The order of adaptor type, emotion and gender was counterbalanced.

3. Results

For Experiment 1, data were averaged as a function of seven morph steps and a psychophysical curve (based on the hyperbolic tangent function) was fitted to the mean data for each adaptor type (baseline condition, adaptation to anger or fear; Fig. 1). The point of subjective equality (PSE), i.e. the centre of symmetry of the psychophysical function, was then computed and is illustrated with a star on all average curves in corresponding colour. Excellent fits were obtained for all three conditions (mean $R^2 = .982$, SD = .016). A one-way repeated measures ANOVA on PSE abscissa values revealed a significant main effect of adaptation to affective voices ($F(2,28) = 21.558$, $p < .001$, $\eta_p^2 = .606$). Exploring the main effects with Bonferroni correction showed that the PSE as a result of adaptation to fear was significantly larger (i.e. more angry) than in the baseline ($p < .004$) and anger adaptation conditions ($p < .001$). During the baseline condition (approximately) the 50% morph was perceived as the most ambiguous. However, after adaptation to fear the 50% morph was more likely perceived as angry and the most ambiguous morph was now shifted to 65%. Conversely, the PSE was significantly smaller (i.e. more fearful) after adaptation to anger compared to the baseline condition ($p < .05$). During anger adaptation the most ambiguous morph was approximately the 35% morph and the 50% morph was now more likely perceived as fearful.

For Experiment 2, data were averaged as a function of morph step and adaptor type (baseline condition, adaptation to anger/fear or caricatured anger/fear) and psychophysical curves were fitted to the mean data (Fig. 2; mean $R^2 = .980$, SD = .023). Prior adaptation shifted the PSE away from the adapting stimulus as in Experiment 1 however this shift was not larger for the caricatured vocalisations compared to the unmanipulated affective voices. One-way repeated measures ANOVA on the PSE abscissa

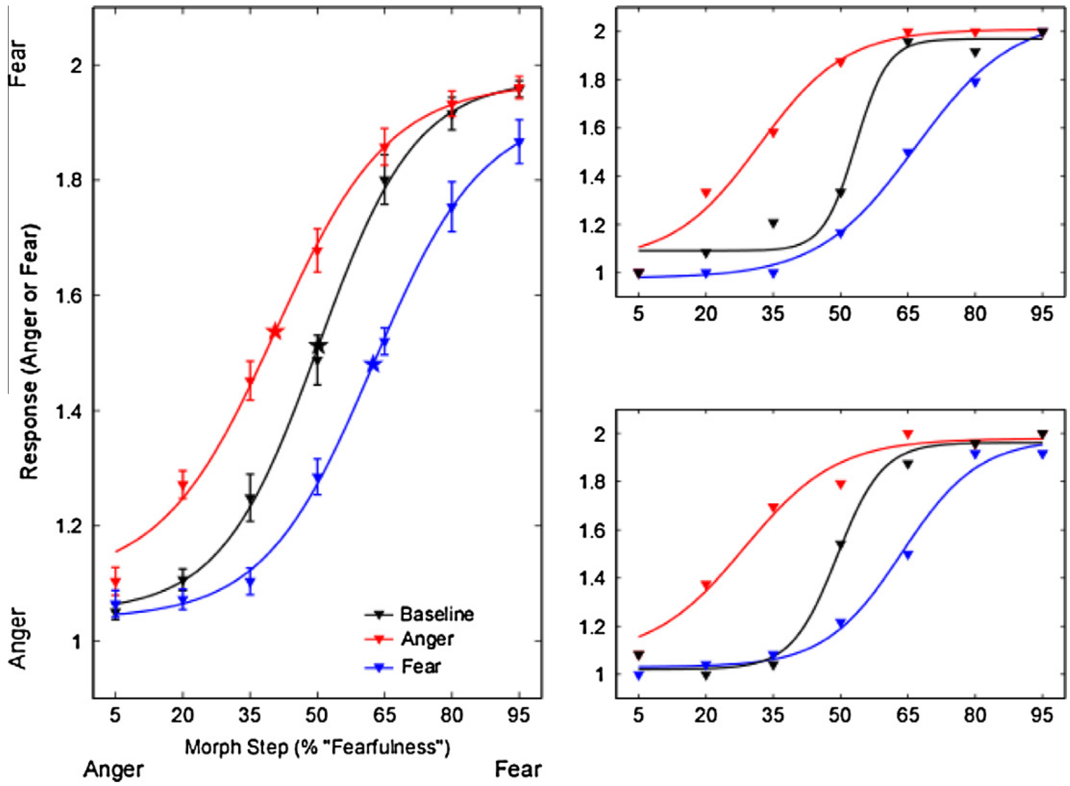

Fig. 1. Psychophysical functions for the three adaptation conditions: baseline (black), adaptation to anger (red) and fear (blue). Left enlarged graph displays the grand average of all participants; the two graphs to the right display individual participants. Stars coloured correspondingly for each condition indicate the PSE. Error bars represent standard error of the mean (SEM).

values revealed the predicted main effect of adaptation to affective voices ($F(4,52) = 70.838$, $p < .001$, $\eta_p^2 = .845$). We decomposed this main effect using post hoc tests with Bonferroni correction which revealed that the PSE as a result of adaptation to original and caricatured fearful voices was significantly larger compared to the PSE in the baseline test (both $p < .001$) and both anger adaptation conditions (both $p < .001$). Similarly, the PSE was significantly smaller in the adaptation to original ($p < .009$) and caricatured angry voices ($p < .001$) compared to the baseline condition. Further, these post hoc tests with Bonferroni correction showed that the caricatured voices did not elicit a stronger adaptation effect than the original voices ($p = 1.000$, Bonferroni corrected for both anger and fear). Even when using paired *t*-tests without correction for multiple comparisons the difference between adaptation to original versus caricatured affect was not significant ($p = .12$ for fear versus caricatured fear and $p = .28$ for anger versus caricatured anger).

It is noteworthy that although we repeated relatively brief stimuli we are confident that the pattern of results is characteristic of adaptation and not priming. Priming would cause the opposite effect (i.e. a facilitation rather than desensitisation). This pattern is also evident in the reaction time (RT) data described in the Supplementary section. Here RT peaks at the location of the PSE.

4. Discussion

In Experiment 1 we demonstrated that adaptation to angry vocalisations made stimuli from a morphed anger–fear continuum sound less angry and more fearful, while adaptation to fearful vocalisations elicited the opposite response. In Experiment 2 we replicated this effect and showed that adaptation to caricatures did not induce stronger aftereffects than adaptation to the original voices. This additional finding suggests that these aftereffects cannot be solely explained by adaptation to acoustical properties of the adaptors but adaptation at high-level information processing stages. For the first time we demonstrate that adaptation to affective sounds can bias perception of affective information and shift the point of subjective equality (PSE) away from the adapting stimulus. Adaptation to affective vocalisations re-sets or re-calibrates auditory representations of affective voices and could thereby influence perception in everyday social situations (Kloth & Schweinberger, 2008; Leopold, Rhodes, Müller, & Jefferey, 2005; Schweinberger et al., 2008).

Research has questioned whether face aftereffects are due to low-level (e.g. adaptation at the retinal level) or high-level adaptation (e.g. adaptation in cerebral areas implicated in face processing). Since adaptation to feature spacing (Little, DeBruine, & Jones, 2005), expression (Bes-

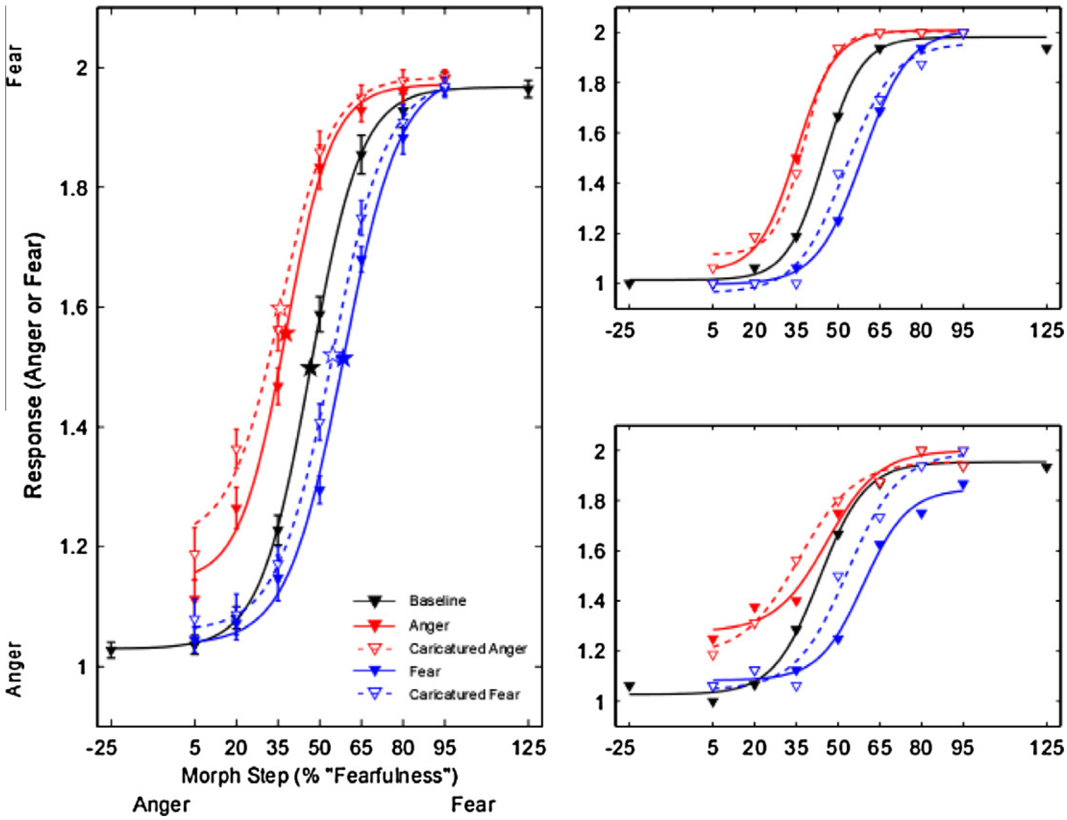

Fig. 2. Psychophysical functions for the five adaptation conditions: baseline (black), adaptation to anger (red), adaptation to caricatured anger (dashed red), fear (blue) and caricatured fear (dashed blue). Left enlarged graph displays the grand average of all participants; the two graphs to the right display individual participants. Stars coloured correspondingly for each condition indicate the PSE. Error bars represent SEM.

telmeyer, Jones, DeBruine, Little, & Welling, 2010; Fox & Barton, 2007) and identity (Leopold, O'Toole, Vetter, & Blanz, 2001), for example, is robust to changes in retinal size and position points towards adaptation involving high-level stages of face perception. Critically, in our experiments participants were tested on a different identity than the one they were adapted to, so that these aftereffects are unlikely to reflect simple adaptation to pitch and timbre and clearly generalize to affective vocal modulations across different individual voices. Of particular importance is that adaptation to caricatured vocal expressions did not elicit stronger adaptation effects than adaptation to original vocal expressions. Taken together our findings imply the existence of high-level representations of vocal affect which can be manipulated using adaptation.

Several potential caveats to our conclusion from Experiment 2 should be noted. First, it is not optimal to predict a null result because it is impossible to determine whether we found no difference in adaptation between original and caricatured voices because none exists or because we lacked statistical power. Second, although participants perceived the caricatures as affectively enhanced compared to the original expressions it is possible that the adaptation effect reached a ceiling. However, the fact that fearful caricatures actually elicited a slightly weaker effect than adaptation to original fearful voices speaks against both points. It is also worth noting that the caricatures were created by exaggerating the original vocalisations relative to one another rather than relative to a neutral stimulus. This may have contributed to perceptual effects of caricaturing that were more marked at one end of the fear–anger continuum than at the other. For example, anger and fear are both usually characterized by a high fundamental frequency (F0) when compared to emotionally neutral stimuli. Consequently, raising the F0 of the fear–caricature probably makes it more expressive of fear when compared to a neutral stimulus, whereas lowering the F0 of the anger–caricature could instead make it express anger less intensively. Although such an effect did not seem to occur in the present case as indicated by the perceptual ratings of the caricatured stimuli, and would not compromise a test of the reliance of the aftereffects on the underlying acoustics, it should be controlled in a comparison of the magnitude of the aftereffect between affective categories, for example by generating stimuli caricatured relative to an affectively neutral stimulus. Additional research needs to address these issues and replicate the results using continua of other affect types. As a first approach, however, the aftereffects reported here could reflect neuronal adaptation phenomena in specific populations of neurons that are tuned to perceptual category and are largely independent of the processing of low-level acoustic properties.

Neural correlates of vocal identity adaptation were identified in the anterior part of the right superior temporal sulcus (Belin & Zatorre, 2003). Adaptation to socially salient information, such as vocal affect, depends on neural substrates as yet unknown. However, likely candidates are the voice-sensitive bilateral areas located in the superior temporal sulcus (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000), which have recently been shown to subserve the processing of several categories of emotional speech (Ethofer, De Ville, Scherer, & Vuilleumier, 2009; Leitman et al., 2010), as well as the orbitofrontal cortex active during evaluating emotional valence of speech intonation (Wildgruber et al., 2004). In addition to these areas, Ethofer et al. (2009) showed that amygdalae, insulae and mediodorsal thalami reacted more strongly to angry compared to neutral prosody.

Adaptation effects to vocal gender (Mullennix et al., 1995; Schweinberger et al., 2008) and affect highlight the fact that high-level adaptation effects are not specific to faces but extend to other modalities. Investigations into so-called 'contingent' aftereffects have revealed interdependent processing of functionally different aspects of faces (e.g. Bestelmeyer et al., 2008, 2010; Little et al., 2005). Together with adaptation experiments on voice gender, affect and identity, future studies on contingent aftereffects will help the building and evaluating of a model of voice perception and further specifying similarities and differences between face and voice perception.

Acknowledgements

We are grateful to Stefan Schweinberger, Mike Burton, Ben C. Jones and Robert Jenkins as well as anonymous reviewers for very helpful comments on a previous version of our manuscript. This research is part of the Social Interactions Project (co-funded by ESRC/MRC; RES-060-25-0010).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.cognition.2010.08.008.

References

Ades, A. E. (1976). New approaches to language mechanisms. In R. J. Wales & E. Walker (Eds.), *Adapting the Property Detectors for Speech Perception* (pp. 55–107). Amsterdam: North-Holland.

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences, 8*, 129–135.

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The montreal affective voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods, 40*, 531–539.

Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport, 14*, 2105–2109.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*, 309–312.

Bestelmeyer, P. E. G., Jones, B. C., DeBruine, L. M., Little, A. C., Perrett, D. I., Schneider, A., et al. (2008). Sex-contingent face aftereffects depend on perceptual category rather than structural encoding. *Cognition, 107*, 353–365.

Bestelmeyer, P. E. G., Jones, B. C., DeBruine, L. M., Little, A. C., & Welling, L. L. M. (2010). Face aftereffects demonstrate interdependent processing of expression and sex and of expression and race. *Visual Cognition, 18*, 255–274.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*, 305–327.

Calder, A. J., Young, A. W., Perrett, D. I., Etcoff, N. L., & Rowland, D. (1996). Categorical perception of morphed facial expressions. *Visual Cognition, 3*, 81–117.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin, 128*, 203–235.

Ethofer, T., De Ville, D. V., Scherer, K., & Vuilleumier, P. (2009). Decoding of emotional information in voice-sensitive cortices. *Current Biology, 19*, 1028–1033.

Ethofer, T., Kreifelts, B., Wiethoff, S., Wolf, J., Grodd, W., Vuilleumier, P., et al. (2009). Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. *Journal of Cognitive Neuroscience, 21*, 1255–1268.

Fox, C. J., & Barton, J. J. S. (2007). What is adapted in face adaptation? The neural representation o expression in the human visual system. *Brain Research, 1127*, 80–89.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron, 24*, 187–203.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*, 223–233.

Jenkins, R., Beaver, J. D., & Calder, A. J. (2006). I thought you were looking at me! Direction-specific aftereffects in gaze perception. *Psychological Science, 17*, 506–514.

Kawahara, H., Matsui, H. (2003). Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation. In: Proceedings IEEE international conference on acoustics, speech, and signal processing, Vol. 1, pp. 256–259.

Kloth, N., & Schweinberger, S. R. (2008). The temporal decay of eye gaze adaptation effects. *Journal of Vision, 8*, 1–11.

Laukka, P. (2005). Categorical perception of vocal expressions. *Emotion, 5*, 277–295.

Leitman, D. I., Wolf, D. H., Ragland, J. D., Laukka, P., Loughead, J., Valdez, J. N., et al. (2010). "It's not what you say, but how you say it": A reciprocal temporo-frontal network for affective prosody. *Frontiers in Human Neuroscience, 4*, 4–19.

Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level after effects. *Nature Neuroscience, 4*, 89–94.

Leopold, D. A., Rhodes, G., Müller, K.-M., & Jefferey, L. (2005). The dynamics of visual adaptation to faces. *Proceedings of the Royal Society, 272*, 897–904.

Little, A. C., DeBruine, L. M., & Jones, B. C. (2005). Sex-contingent face after-effects suggest distinct neural populations code male and female faces. *Proceedings of the Royal Society B-Biological Sciences, 272*, 2283–2287.

Mullennix, J. W., Johnson, K. A., TopcuDurgun, M., & Farnsworth, L. M. (1995). The perceptual representation of voice gender. *Journal of the Acoustical Society of America, 98*, 3080–3095.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5*, 42–46.

Pelli, D. G. (1997). The Video Toolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Rhodes, G., Jeffery, L., Watson, T. L., Clifford, C. W. G., & Nakayama, K. (2003). Fitting the mind to the world: Face adaptation and attractiveness aftereffects. *Psychological Science, 14*, 558–566.

Sawusch, J. R. (1977). Peripheral and central processes in selective adaptation of place of articulation in stop consonants. *Journal of the Acoustical Society of America, 62*, 738–750.

Schweinberger, S. R., Casper, C., Hauthal, N., Kaufmann, J. M., Kawahara, H., Kloth, N., et al. (2008). Auditory adaptation in voice perception. *Current Biology, 18*, 684–688.

Wallbott, H., & Scherer, K. R. (1986). How universal and specific is emotional experience. *Social Science Information, 26*, 763–795.

Watson, T. L., & Clifford, C. W. G. (2003). Pulling faces: An investigation of the face, distortion aftereffect. *Perception, 32*, 1109–1116.

Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature, 428*, 557–561.
Webster, M. A., & MacLin, O. H. (1999). Figural aftereffects in the perception of faces. *Psychonomic Bulletin & Review, 6*, 647–653.
Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W., et al. (2004). Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. *Cerebral Cortex, 14*, 1384–1389.
Winston, J. S., Henson, R. N. A., Fine-Goulden, M. R., & Dolan, R. J. (2004). FMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. *Journal of Neurophysiology, 92*, 1830–1839.
Zhao, L., & Chubb, C. (2001). The size-tuning of the face, distortion after-effect. *Vision Research, 41*, 2979–2994.