Cognition 134 (2015) 245–251

Contents lists available at ScienceDirect

Cognition

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

Brief article

Evidence for a supra-modal representation of emotion from cross-modal adaptation

Annie Pye, Patricia E.G. Bestelmeyer *

School of Psychology, Bangor University, Gwynedd LL57 2AS, UK

ARTICLE INFO

Article history:
Received 31 July 2013
Revised 27 August 2014
Accepted 2 November 2014
Available online 19 November 2014

Keywords:
Supra-modal representation
Cross-modal
Adaptation
Emotion
Voice

ABSTRACT

Successful social interaction hinges on accurate perception of emotional signals. These signals are typically conveyed multi-modally by the face and voice. Previous research has demonstrated uni-modal contrastive aftereffects for emotionally expressive faces *or* voices. Here we were interested in whether these aftereffects transfer across modality as theoretical models predict. We show that adaptation to facial expressions elicits significant auditory aftereffects. Adaptation to angry facial expressions caused ambiguous vocal stimuli drawn from an anger-fear morphed continuum to be perceived as less angry and more fearful relative to adaptation to fearful faces. In a second experiment, we demonstrate that these aftereffects are not dependent on learned face-voice congruence, i.e. adaptation to one facial identity transferred to an unmatched voice identity. Taken together, our findings provide support for a supra-modal representation of emotion and suggest further that identity and emotion may be processed independently from one another, at least at the supra-modal level of the processing hierarchy.

1. Introduction

The face and voice convey important information about an individual such as gender, identity or emotional state. The accurate perception and integration of relevant signals from the face and voice plays an important role in successful social interaction. The 'auditory face' model of voice perception (Belin, Fecteau, & Bedard, 2004), based on the model of face perception by Bruce and Young (1986), suggests that following separate low-level visual and auditory analysis, (lip) speech, affect and identity are processed by dissociable perceptual systems. Furthermore, the model hypothesises the presence of interactions between face and voice at each of the dissociable processing pathways. While Belin et al.'s model was published nearly a decade ago and is widely cited in the literature, only few studies have empirically tested its assumptions. One method of exploring the predictions made by this model is with the use of adaptation paradigms.

Adaptation refers to the process during which continued stimulation results in biased perception towards opposite features of the adapting stimulus (Grill-Spector et al., 1999). The use of adaptation paradigms provides information about what is encompassed within a given representation (Dosenbach et al., 2007) and enables further understanding of the neural underpinnings of our sensory and perceptual experiences through the study of aftereffects (Kohn, 2007). 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).

* Corresponding author at: School of Psychology, Bangor University, Brigantia Building, Penrallt Road, Bangor LL57 2AS, UK. Tel.: +44 (0) 1248 383488; fax: +44 (0) 1248 382599.

E-mail address: p.bestelmeyer@bangor.ac.uk (P.E.G. Bestelmeyer).

http://dx.doi.org/10.1016/j.cognition.2014.11.001
0010-0277/

Adaptation is therefore a useful tool to investigate different levels of processing including "low-level" stimulus properties such as motion, shape or colour (e.g. Wright, 1934) but also "higher-level" properties such as facial features. Webster and MacLin (1999) were the first to show that adaptation to a face with distorted features resulted in a perceptual bias in which the test face was interpreted as being distorted in the opposite direction to that of the adaptor face. Aftereffects have been found for a variety of facial features and have been widely used to investigate the mechanisms that underpin face perception (e.g. Bestelmeyer, Jones, DeBruine, Little, & Welling, 2010; Bestelmeyer et al., 2008; Fox & Barton, 2007; Jenkins, Beaver, & Calder, 2006; Leopold, O'Toole, Vetter, & Blanz, 2001; Leopold, Rhodes, Muller, & Jeffery, 2005; Little, DeBruine, & Jones, 2005; Webster, Kaping, Mizokami, & Duhamel, 2004; Welling et al., 2009). While this research has largely concentrated on evaluating face perception models it has recently been shown that corresponding auditory aftereffects can be elicited for voices (e.g. identity (Zäske, Schweinberger, & Kawahara, 2010), gender (Schweinberger et al., 2008), age (Zäske & Schweinberger, 2011) and affect (Bestelmeyer, Rouger, DeBruine, & Belin, 2010; Skuk & Schweinberger, 2013)).

Few cross-modal adaptation studies have probed the existence of supra-modal, or modality-independent, representations of person-specific information such as identity. While Belin et al. (2004) clearly predict cross-modal interactions between perceptual pathways (i.e. lip speech, identity, affect) empirical studies have yielded inconsistent results. For example, studies concerning cross-modal gender adaptation between face and voice have usually failed to find significant aftereffects (Owlia & Jordan, 2009; Schweinberger et al., 2008; but see Little, Feinberg, DeBruine, & Jones, 2013). While cross-modal studies regarding identity information unanimously report aftereffects (Hills, Elward, & Lewis, 2010; Zäske et al., 2010), cross-modal studies of emotion again report mixed findings. Adapting to sentences spoken in an emotional tone did not distort perception of static facial expressions (Fox & Barton, 2007), yet when using dynamic facial adaptors cross-modal aftereffects were found for emotional prosody (expressed as brief syllables) but only in male listeners (Skuk & Schweinberger, 2013). Experiment 1 explored whether adaptation to facial expressions affects the perception of non-linguistic, vocal expressions.

An on-going debate exists as to whether functionally different aspects of faces are indeed processed independently of one another, as several theoretical models suggest (Belin, Bestelmeyer, Latinus, & Watson, 2011; Belin et al., 2004; Bruce & Young, 1986; Haxby, Hoffman, & Gobbini, 2000; Le Gal & Bruce, 2002; Young, 1998; but see Young & Bruce, 2011 for a revised account). This idea implies that the processing of one dimension (e.g. expression) is not influenced by the processing of a functionally distinct dimension (e.g. identity). While neuropsychological research (e.g. Garrido & et al., 2009; Hailstone, Crutch, Vestergaard, Patterson, & Warren, 2010; Humphreys, Donnelly, & Riddoch, 1993; Winston et al., 2004) as well as some behavioural studies (e.g. Campbell, Brooks, deHaan, & Roberts, 1996; Ellis, Young, & Flude, 1990; Young, McWeeny, Hay, & Ellis, 1986) support the notion of independent pathways in the processing of identity and affect, other behavioural studies (Ganel & Goshen-Gottstein, 2004; Schweinberger, Burton, & Kelly, 1999; Schweinberger & Soukup, 1998) as well as a neuroimaging study (Fox, Moon, Iaria, & Barton, 2009) suggest an *inter*dependence between variant and invariant aspects of the face, especially between identity and expression. Belin et al.'s model (2004) assumes independence of (lip) speech, identity and expression both at the uni-modal and bi-modal level. Experiment 2 addressed whether cross-modal adaptation is contingent upon the adaptor and test stimuli coming from the same, familiarised identity or whether these aftereffects are identity independent.

The present study examined some of the predictions of the 'auditory face' model (Belin et al., 2004). Using dynamic video stimuli, Experiment 1 tested whether adaptation to angry faces would shift perception of affectively ambiguous voice morphs towards fear and whether repeated exposure to fearful faces would shift perception in the opposite direction. Experiment 2 aimed to explore the relationship between identity and emotion at a supra-modal level of representation. If identity and affect are processed by dissociable pathways, it is likely that any cross-modal aftereffect should be robust to changes in the identity between face adaptor and vocal test morphs. Thus, we predicted that adaptation to one individual's face would elicit aftereffects in the perception of a different individual's voice.

2. Methods

2.1. Participants

Student volunteers from Bangor University participated in return for course credit. Participants had normal or corrected-to-normal vision and normal hearing, were Caucasian due to known effects of race on face perception (e.g. Lindsay, Jack, & Christian, 1991) and had no psychiatric or neurological problems. Twenty-five volunteers contributed to Experiment 1 (19 female, mean age = 23.4, standard deviation (SD) = 6.1) and twenty-three participants contributed to Experiment 2 (15 female, mean age = 20.7 SD = 4.09). None of the participants took part in more than one reported study and none were personally familiar with any of the age-matched individuals displayed in the stimulus material.

2.2. Stimuli

A pilot study using static images of emotional faces resulted in no cross-modal aftereffects (see supplementary online material (SOM) for details). We therefore created full high-definition videos of twelve individuals displaying seven expressions (6 basic emotions and neutral) in a professional sound studio with controlled lighting. We recorded videos and high quality audio simultaneously. Volunteers were instructed to use the vowel "ah" for each expression.

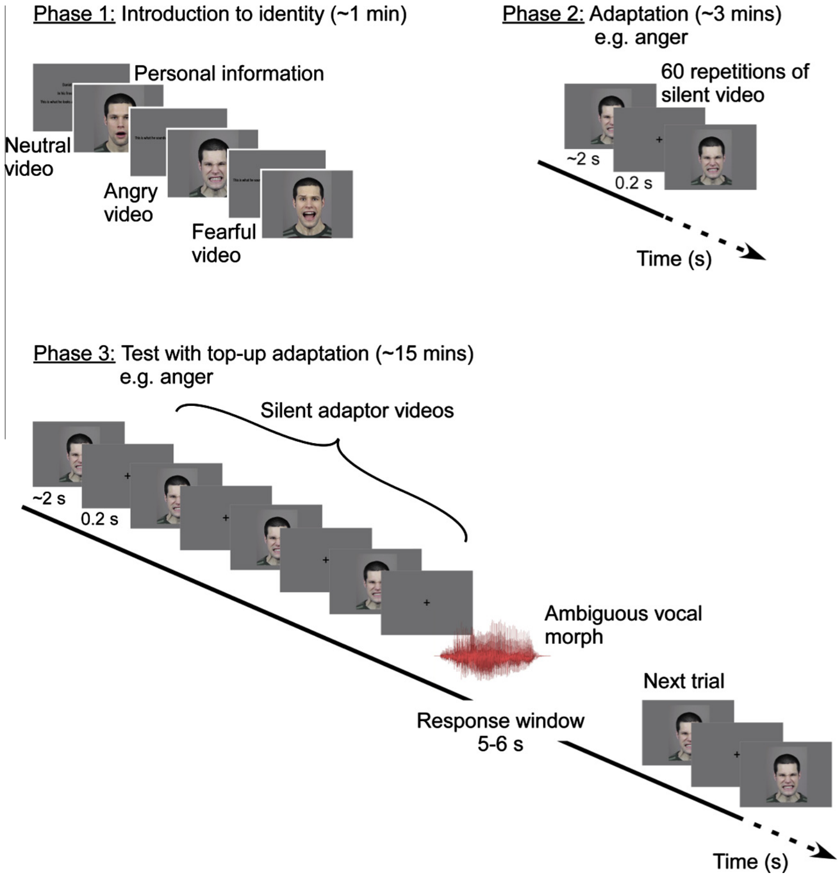

Fig. 1. A schematic diagram of the three phases employed in both experiments. Experiment 2 would show a different facial identity of the same sex during the next trial of Phase 3.

We edited the videos using Adobe Premiere CS6 so that they started and ended on a neutral frame. Videos were sized at 105 mm × 140 mm and positioned centrally in the frame, with a grey background. Audio recordings were edited using Cool Edit Pro 2.0 and normalised in energy (root mean square). Based on the outcome of a stimulus validation study (nine participants; 7-alternative forced-choice task) we selected four identities per gender with the highest recognition rates (>80%) of their uni-modal visual and auditory, fearful and angry expressions.

Two voice continua (1 female) were created between anger and fear in seven morph steps at intervals of 5% fear/95% anger, 20/80%, 35/65%, 50/50%, 65/35%, 80/20% and 95/5% using Tandem-STRAIGHT (Banno et al., 2007; Kawahara et al., 2008; see SOM for details). Psychtoolbox-3 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997) was used for stimulus presentation. Toolboxes were run on MatlabR2012b (The Mathworks, Inc.). Voices were presented via headphones (Beyerdynamic) at 75 dB SPL(C).

For Experiment 1, we employed videos of one male and one female identity. For Experiment 2, we used an additional three facial identities for each gender. Vocal morphs were the same in both experiments. To ensure that the newly generated morphs were perceived categorically, we asked pilot subjects to complete a two-alternative forced choice judgement as to which emotion each morph portrayed. The results replicated the sigmoid-shaped curves obtained in previous research (Bestelmeyer, Rouger et al., 2010; Bestelmeyer, Maurage, Rouger, Latinus, & Belin, 2014; Laukka, 2005) with the 50% morph being perceived as most ambiguous. We also ensured that we obtain comparable uni-modal auditory aftereffects with our voice stimuli and revised adaptation methodology as previously reported in Bestelmeyer, Rouger et al. (2010; see SOM for results).

2.3. Procedure

Both experiments consisted of three phases: an introduction to a particular identity, an adaptation phase in which participants were adapted to the silent facial expression of the introduced identity and an adaptation-test phase in which each test voice was preceded by four repetitions of the same silent video as “top-up adaptation” (Fig. 1). Phases 2–3 were repeated four times (gender 1 anger, gender 2 anger, gender 1 fear, gender 2 fear). The two anger and two fear runs were always administered consecutively but the emotion order was counterbalanced. Gender was counterbalanced across the four runs. Participants had a minimum of 1-minute break between the first and last two runs (based on Zäske et al.’s (2010) finding on durations of cross-modal aftereffects) and a 5-minute break between emotions, i.e. at the half-way point of the experiment.

In Experiment 1, the adaptor face and test voice always belonged to the same person (one identity per gender). In Experiment 2, we used three additional faces per gender. Participants in Experiment 2 were familiarised with only one of the four identities. Thus only one quarter of the trials was identity-congruent while on the remainder of the trials the face adaptor did not identity-match the test voice. Identity-congruent and incongruent trials were randomised.

We introduced the participant to each adaptor-identity using a short brief which contained the name of the identity, their profession, one hobby and bi-modal video clips portraying them as neutral, angry and fearful. The purpose was to encourage face-voice binding by familiarising the participant with the identity and how that identity expresses both emotions. Participants clicked through the brief in their own time and were required to read aloud the information to ensure it was being attended to.

Following the introduction, participants completed an adaptation phase to one of two identities and emotions (angry or fearful depending on the subsequent test with top-up adaptation run). Here a silent video of the identity was repeated sixty times with an inter-stimulus interval (ISI) of 200 ms.

Participants were then required to complete the test with top-up adaptation phase. We emphasised at the start of each run of this phase that the voice they would hear will always be that of the individual in the introduction and adaptation phase. Trials consisted of 4 repetitions of the silent adaptor video followed by the onset of the vocal test morph after 500 ms. The task involved a two-alternative forced choice judgement of whether the voice expressed either anger or fear by means of a button response. Trials were response-terminated unless no response was given within 5–6 s (randomly jittered) in which case, the next trial commenced automatically (excluded missed trials in each experiment: <1.1%). Each run consisted of 56 trials, with each of the seven morphs being rated eight times.

2.4. Statistical analyses

We used bootstrapping to assess statistical significance of the difference between conditions (adaptation to anger or fear) using Matlab2014a with Statistics Toolbox. Each bootstrap sample was derived by randomly sampling from our participants with replacement. Thus a participant and associated data could be selected more than once or not at all according to conventional bootstrap methodology (Wilcox, 2012). Data from both conditions were selected for each sampled participant because of the within-subjects design of the experiments. For each participant, data were then averaged as a function of the seven morph steps and a psychophysical curve (based on the hyperbolic tangent function) was fitted for each condition. We then computed the difference between adaptation conditions. We repeated this process 9999 times which led to a distribution of 10,000 bootstrapped estimates of the fit to the psychophysical curves as well as differences between curves for the two conditions. Lastly, we calculated the 95% confidence intervals ($CI_{95\%}$) for the fitted curve (Figs. 2A and 3A/C) and the differences between two conditions (Figs. 2B and 3B/D/E). A difference between conditions is deemed significant if the mean of the differences and its $CI_{95\%}$ excludes 0 (e.g. Cumming, 2012). The point of subjective equality (PSE), i.e. the centre of symmetry of the psychophysical function, was also computed and is illustrated with a star on all average psychophysical curves.

3. Results

Fig. 2A illustrates the fitted functions for each adaptation condition for Experiment 1, $CI_{95\%}$ and error bars (SEM). Fig. 2B shows the $CI_{95\%}$ of the differences between adaptation conditions and illustrates a significant cross-modal aftereffect which was largest around the 50% morph, the mathematically and perceptually most ambiguous morph.

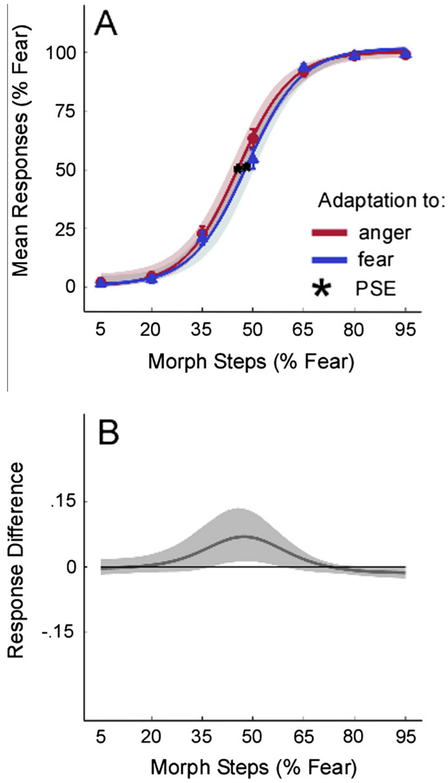

Fig. 2. Results of Experiment 1. (A) Average psychophysical functions for the two adaptation conditions anger (red) and fear (blue). Error bars represent standard error of the mean (SEM) and shaded areas illustrate the bootstrapped 95% confidence intervals ($CI_{95\%}$) of the fitted curve to participants' responses. (B) Mean difference and $CI_{95\%}$ (shaded area) of the differences in identification during the two adaptation conditions for Experiment 1. Significant identification differences between adaptation conditions occurred along the most ambiguous part of the continuum (i.e. where the $CI_{95\%}$ area does not overlap with the $y = 0$ line (black)). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

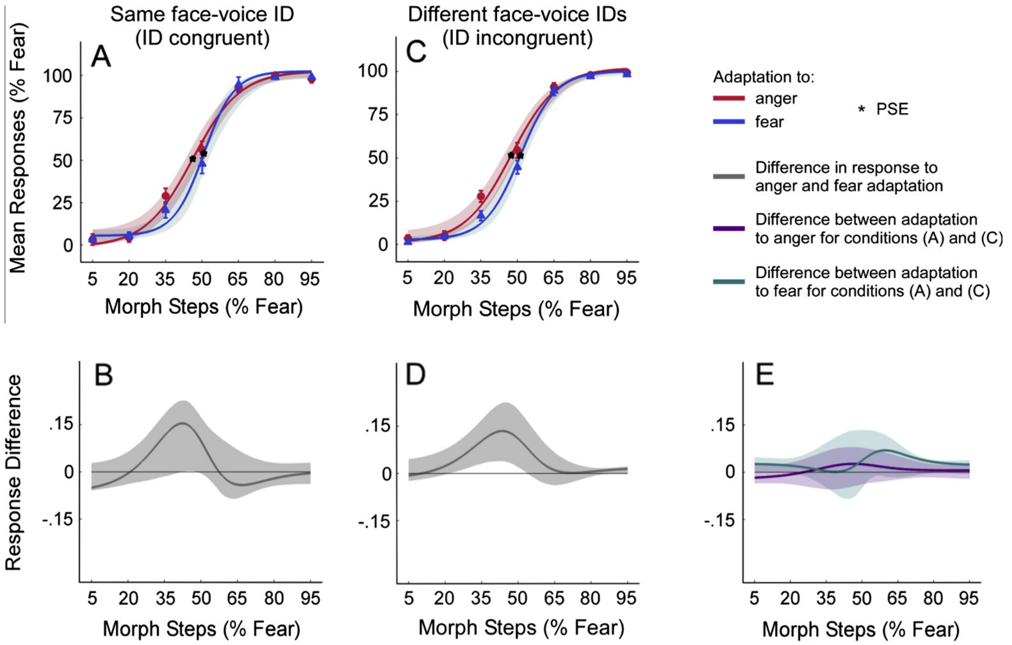

Fig. 3. Results of Experiment 2. (A) Average psychophysical functions for the two adaptation conditions anger (red) and fear (blue) for the face-voice identity (ID) congruent condition and (B) incongruent condition. Error bars represent SEM and shaded areas illustrate the bootstrapped $CI_{95\%}$ of the fitted curve to participants' responses. (C) Mean difference and $CI_{95\%}$ (shaded area) of the differences in identification during the ID congruent condition. (D) Mean difference and $CI_{95\%}$ (shaded area) of the differences in identification during the ID incongruent condition. Significant identification differences between adaptation conditions occurred along the most ambiguous part of the continuum in both identity conditions (i.e. where the $CI_{95\%}$ area does not overlap with the $y = 0$ line (black)). (E) Mean differences and $CI_{95\%}$ (shaded area) of the differences between ID congruent and ID incongruent anger (purple) and fear (turquoise) adaptation conditions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In Experiment 2 participants knew that the voice morphs always belonged to the face of the introduced identity so that on some of the trials the adaptor-test pair was identity-congruent and on others it was incongruent. Responses were therefore averaged separately for congruent (Fig. 3A/B) and incongruent (Fig. 3C/D) adaptor-test identity trials as a function of morph step. We found significant cross-modal adaptation effects for both identity-congruent and incongruent trials which were largest around the 50% morph. To assess if face-voice identity congruence modulated the aftereffect we also calculated the difference in adaptation effect size between anger congruent versus incongruent and fear congruent versus incongruent conditions but found no significant differences (Fig. 3E).

Due to a recent finding of gender-specificity of cross-modal aftereffects (Skuk & Schweinberger, 2013) we pooled the data across the two experiments. When assessing the difference between genders within each adaptation condition for the seven morph steps (using a bootstrap technique adjusted for a between-subjects design) we found no significant gender differences (all $p > .88$).

4. Discussion

Experiment 1 demonstrates that adaptation to emotion in the face can result in perceptual aftereffects in the voice, i.e. adapting to an angry face caused subsequent ambiguous vocal stimuli, taken from an anger-fear continuum, to be perceived as more fearful relative to fear adaptation. Experiment 2 replicated this effect and further suggests that this aftereffect is independent of identity. Taken together, our results provide support for the existence of a high-level, supra-modal representation of emotion that is independent of low-level perceptual analysis. Our finding suggests that despite the existence of separate channels for emotion processing, interaction of these systems causes re-adjustments in perception, a finding with significant implications for everyday social interactions.

The aftereffects we report here are smaller in magnitude than those previously described in uni-modal affect adaptation studies (e.g. Bestelmeyer, Rouger, et al., 2010; Fox & Barton, 2007) including a study in which we employ the same stimuli and methodology as in the present experiments (see SOM for comparison). Previous research has suggested sensory processing to be hierarchical with higher-level neurons theorised to have a higher degree of specificity than lower-level neuronal populations (Xu, Dayan, Lipkin, & Qian, 2008). Therefore it could be that the smaller effects observed from higher-level adaptation studies reflect the reduced number of neurons targeted by the adaptation effect.

The discrepancy between our results and that of previous studies may be related to this point. Fox and Barton (2007) reported significant uni-modal but not cross-modal aftereffects for static facial expressions. In the cross-modal condition they used neutral sentences read with emotional prosody as adaptors and still images of facial expressions as test stimuli. We have maximised adaptation by adapting participants prior to the test phase and again before each test stimulus with *dynamic* stimuli of the same event. Again, using emotional linguistic stimuli, Skuk and Schweinberger (2013) found cross-modal aftereffects but only for male participants. Maximising adaptation, the use of dynamic stimuli and non-linguistic expressions may be key in obtaining consistent cross-modal aftereffects behaviourally.

Our finding of Experiment 2 suggests that the representations of identity and emotion may be processed independently of one another at least at the supra-modal level. Although this finding cannot clarify the relationship between identity and expression at a uni-modal level, it may suggest that supra-modal representations are somewhat broader in what they encompass relative to modality-specific representations. It is also possible that the introductory brief was not sufficient to enable face-voice binding and that a modulation of the aftereffect size by identity is apparent only when using adaptors that are *personally* familiar to participants (Schweinberger, Robertson, & Kaufmann, 2007). Furthermore, cross-modal effects may be so small that a modulation with identity is not detectable with behavioural experiments.

Neuroscientific models implicate the superior temporal sulcus/gyrus (STS/STG) in the perception of emotional cues from the voice (Schirmer & Kotz, 2006) and face (Haxby et al., 2000). While research suggests that the integration of emotional cues from the voice and face involve the STS/STG (Kreifelts, Ethofer, Grodd, Erb, & Wildgruber, 2007) studies that specifically investigated which area(s) in the brain code emotion independent of modality provide mixed results that are not entirely in agreement with aforementioned models (Klasen, Kenworthy, Mathiak, Kircher, & Mathiak, 2011; Peelen, Atkinson, & Vuilleumier, 2010). Discrepancies in results are likely due to differences in paradigms, analysis techniques and stimulus sets. Utilising adaptation paradigms in conjunction with fMRI may assist in clarifying the neural correlates of the supra-modal representation of emotion through minimising activations related to a more general network dealing with the integration of affective information across modalities and task demands.

5. Conclusion

Our experiments show that high-level contrastive aftereffects of emotion are modality and perhaps also identity independent. Our results confirm the notion of a supra-modal representation of emotion and therefore support Belin et al.'s (2004) 'auditory face' model. Future studies are necessary for the development and evaluation of multimodal person perception models and the neural correlates that underpin these processes.

Acknowledgments

Annie Pye is funded by a Ph.D. studentship sponsored by the School of Psychology, Bangor University.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2014.11.001.

References

Banno, H., Hata, H., Morise, M., Takahashi, T., Irino, T., & Kawahara, H. (2007). Implementation of realtime STRAIGHT manipulation system: Report on its first implementation. *Acoustical Science and Technology, 28*, 140–146.

Belin, P., Bestelmeyer, P. E. G., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology, 102*, 711–725.

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

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.

Bestelmeyer, P. E. G., Rouger, J., DeBruine, L. M., & Belin, P. (2010). Auditory adaptation in vocal affect perception. *Cognition, 117*, 217–223.

Bestelmeyer, P. E. G., Maurage, P., Rouger, J., Latinus, M., & Belin, P. (2014). Adaptation to vocal expressions reveals multistep perception of auditory emotion. *Journal of Neuroscience, 34*, 8098–8105.

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.

Campbell, R., Brooks, B., deHaan, E., & Roberts, T. (1996). Dissociating face processing skills: Decisions about lip-read speech, expression, and identity. *Quarterly Journal of Experimental Psychology, Section A, 49*, 295–314.

Cumming, G. (2012). *Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. New York: Routledge.

Dosenbach, N. U. F., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A. T., et al. (2007). Distinct brain networks for adaptive and stable task control in humans. *Proceedings of the National Academy of Sciences of the United States of America, 104*, 11073–11078.

Ellis, A. W., Young, A. W., & Flude, B. M. (1990). Repetition priming and face processing: Priming occurs within the system that responds to the identity of a face. *Quarterly Journal of Experimental Psychology Section A – Human Experimental Psychology, 42*, 495–512.

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

Fox, C. J., Moon, S. Y., Iaria, G., & Barton, J. J. S. (2009). The correlates of subjective perception of identity and expression in the face network: An fMRI adaptation study. *Neuroimage, 44*, 569–580.

Ganel, T., & Goshen-Gottstein, Y. (2004). Effects of familiarity on the perceptual integrality of the identity and expression of faces: The parallel-route hypothesis revisited. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 583–597.

Garrido, L. et al. (2009). Developmental phonagnosia: A selective deficit of vocal identity recognition. *Neuropsychologia, 47*, 123–131.

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.

Hailstone, J. C., Crutch, S. J., Vestergaard, M. D., Patterson, R. D., & Warren, J. D. (2010). Progressive associative phonagnosia: A neuropsychological analysis. *Neuropsychologia, 48*, 1104–1114.

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.

Hills, P. J., Elward, R. L., & Lewis, M. B. (2010). Cross-Modal Face Identity Aftereffects and Their Relation to Priming. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 876–891.

Humphreys, G. W., Donnelly, N., & Riddoch, M. J. (1993). Expression is computed separate from facial identity, and it is computed separately for moving and static faces-Neuropsychological evidence. *Neuropsychologia, 31*, 173–181.

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*, 506–513.

Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., & Banno, H. (2008). Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. In *Proceedings ICASSP 2008* (pp. 3933–3936). Las Vegas.

Klasen, M., Kenworthy, C. A., Mathiak, K. A., Kircher, T. T. J., & Mathiak, K. (2011). Supramodal representation of emotions. *Journal of Neuroscience, 31*, 13635–13643.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception, 36*, 14.

Kohn, A. (2007). Visual adaptation: Physiology, mechanisms, and functional benefits. *Journal of Neurophysiology, 97*, 3155–3164.

Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., & Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: An event-related fMRI study. *Neuroimage, 37*, 1445–1456.

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

Le Gal, P. M., & Bruce, V. (2002). Evaluating the independence of sex and expression in judgments of faces. *Perception & Psychophysics, 64*, 230–243.

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., Muller, K. M., & Jeffery, L. (2005). The dynamics of visual adaptation to faces. *Proceedings of the Royal Society B: Biological Sciences, 272*, 897–904.

Lindsay, D. S., Jack, P. C., & Christian, M. A. (1991). Other-race face perception. *Journal of Applied Psychology, 76*, 587–589.

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.

Little, A. C., Feinberg, D. R., DeBruine, L. M., & Jones, B. C. (2013). Adaptation of faces and voices: Uni-modal, cross-modal, and sex-specific effects. *Psychological Science, 24*, 2279–2305.

Owlia, A., & Jordan, H. (2009). Same- and cross-modal perceptual effects on gender and identity adaptation. *Journal of Vision, 9*, 469.

Peelen, M. V., Atkinson, A. P., & Vuilleumier, P. (2010). Supramodal representations of perceived emotions in the human brain. *Journal of Neuroscience, 30*, 10127–10134.

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

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences, 10*, 24–30.

Schweinberger, S. R., Burton, A. M., & Kelly, S. W. (1999). Asymmetric dependencies in perceiving identity and emotion: Experiments with morphed faces. *Perception & Psychophysics, 61*, 1102–1115.

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.

Schweinberger, S. R., Robertson, D., & Kaufmann, J. M. (2007). Hearing facial identities. *The Quarterly Journal of Experimental Psychology, 60*, 1446–1456.

Schweinberger, S. R., & Soukup, G. R. (1998). Asymmetric relationships among perceptions of facial identity, emotion, and facial speech. *Journal of Experimental Psychology – Human Perception and Performance, 24*, 1748–1765.

Skuk, V. G., & Schweinberger, S. R. (2013). Adaptation aftereffects in vocal emotion perception elicited by expressive faces and voices. *PLoS One, 8*, e81691.

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.

Welling, L. L. M., Jones, B. C., Bestelmeyer, P. E. G., DeBruine, L. M., Little, A. C., & Conway, C. A. (2009). View-contingent aftereffects suggest joint coding of face shape and view. *Perception, 38*, 133–141.

Wilcox, R. R. (2012). *Introduction to robust estimation and hypothesis testing* (3rd ed.). Academic Press.

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.

Wright, W. D. (1934). The measurement and analysis of colour adaptation phenomena. *Proceedings of the Royal Society of London. Series B, 115*, 49–87.

Xu, H., Dayan, P., Lipkin, R. M., & Qian, N. (2008). Adaptation across the cortical hierarchy: Low-level curve adaptation affects high-level facial-expression judgments. *Journal of Neuroscience, 28*, 3374–3383.

Young, A. W. (1998). *Face and mind*. Oxford, UK: Oxford University Press.

Young, A. W., & Bruce, V. (2011). Understanding person perception. *British Journal of Psychology, 102*, 959–974.

Young, A. W., McWeeny, K. H., Hay, D. C., & Ellis, A. W. (1986). Matching familiar and unfamiliar faces on identity and expression. *Psychological Research, 48*, 63–68.

Zäske, R., & Schweinberger, S. R. (2011). You are only as old as you sound: Auditory aftereffects in vocal age perception. *Hearing Research, 282*, 283–288.

Zäske, R., Schweinberger, S. R., & Kawahara, H. (2010). Voice aftereffects of adaptation to speaker identity. *Hearing Research, 268*, 38–45.